

顔検出・顔認識のための統計的手法

栗田 多喜夫

産業技術総合研究所 脳神経情報研究部門

takio-kurita@aist.go.jp

あらまし 顔認識・顔検出を実現するためのプログラムでは、統計的パターン認識手法が重要な役割を演じる。これまでも多くの統計的パターン認識手法が顔認識や顔検出に利用されてきた。本稿では、顔認識や顔検出に利用可能な代表的な統計的パターン認識手法について紹介する。

1 はじめに

顔画像の自動認識は、建物への入退出の監視 [21]、銀行の ATM などでの個人識別 [92]、マンマシンインタフェース [33]、ロボットとの対話 [32]、ビデオ映像の検索のためのキーとしての利用 [98] などの様々な応用があり、盛んに研究されている [96, 19, 3]。実環境下での顔画像の自動認識はコンピュータにとってまだ難しい課題であるが、ある程度制限された環境下では高い認識率が得られることが報告されている [103, 53, 31, 57, 64, 100, 93, 71]。この分野の研究動向については、すでにいくつかの解説論文が書かれている [96, 19, 3, 35, 43] ので、本稿では、顔画像の認識・検出に有効な統計的手法について概説する。

2 顔認識検出のための特徴

画像中の顔の認識や検出では、画像からどのような特徴を抽出するかで性能に大きな差がでる。その意味で、特徴抽出は顔認識や顔検出で非常に重要であり、様々な特徴抽出手法が提案されている。ここでは、最近、Appearance-based 顔認識・検出手法で良く利用されている Gabor 特徴を用いる特徴抽出手法について紹介し、その簡略版と解釈でき、文字認識において実績のある特徴抽出法（4方向面特徴）について概説する。

2.1 人間の初期視覚系を参考にした特徴

現在のビジョンシステムの能力は、人間の視覚と比べると柔軟性・頑健性・適応性などの点でまだまだ劣っている。人間の視覚は、現在のビジョンシステムとはかなり異なった仕組みに基づいて素晴らしい能力を発揮している。このような人間の視覚系で行われている特徴抽出法を模倣すると、実環境下での顔画像認識の

精度が向上すると期待される。

網膜は、眼底にはりついている透明な神経組織であり、外界の視覚情報を受けとり、局所的な情報処理の結果を神経パルス列に符合化して視覚中枢に送り込む働きをしている。網膜からの情報は、中継ぎである外側膝状体 (lateral geniculate nucleus) を経て後頭部にある視覚野へ伝えられる。Field は、風景や顔などの人工物を含まない自然画像の局所的な自己相関を調べ、そのパワースペクトルが空間周波数の 2 乗に反比例することを示した [27]。Atik 等は、神経節細胞の出力のパワースペクトルは、低周波域では、平坦（コンスタント）になることを示した [12]。これは、元の空間で考えると、自己相関を空間的に無相関化していることに対応し、入力情報から空間的な冗長性を取り除いている (whitening) と解釈できる。逆に、高周波域では、whitening によりノイズが増幅されることを防ぐために高周波成分を抑制するような働きがある。

外側膝状体からの情報は、6 層構造をした大脳皮質の視覚第一次野 (V1) の 4C 層へ伝えられる。V1 のニューロンは、受容野により規定される方向を持った直線状のコントラストに対して強く反応する。このような方向選択性を持ったニューロンは単純細胞と呼ばれている。これに対し、光刺激の位置が方位に垂直方向に対象ずれても反応の強さが変化しない複雑型細胞と呼ばれるニューロンも存在する。Olshausen と Field は、いくつかの基底ベクトルの線形結合により入力画像をなるべく近似し、しかもその結合係数なるべくスパースになるような基準で基底ベクトルを求めると、V1 の単純型細胞の特性と似たものが得られることを示した [74]。一方、Bell と Sejnowski は、独立成分分析 (ICA) を用いて Olshausen と Field と同様な結果が得られることを示した [16]。

以上のような結果から、少なくとも、網膜や第一次視覚野などの初期視覚では、入ってきた情報からなる

べく多くの情報を取り込み、しかも取り込んだ情報に含まれる冗長性を取り除くような情報処理を実現するための自己組織化が行なわれていると解釈できる。これは、入力情報を取り込むという機能を実現するための最も自然な動作原理であると考えられる。

著者等は、網膜のガングリオン細胞の受容野特性に類似したコントラストフィルタ [12, 75] と第一次視覚野の単純型細胞の受容野特性に類似した Gabor フィルタ [20] を用いた顔検出・認識手法を提案した [40, 41, 58, 59, 42]。

2.1.1 コントラストフィルタ

Gabor 特徴を抽出する前に、前処理として、入力画像にコントラストフィルタを適用する。ここでは、Atick 等 [12] によって提案されたコントラストフィルタを紹介する。Atick 等は、コントラストフィルタのフーリエ変換の核関数として、

$$K(f) = W(f)L(f) = f \exp\left\{-\left(f/f_0\right)^4\right\}, \quad (1)$$

を用いた。ここで、 f と f_0 は、周波数と遮断周波数を表している。このフィルタは、実際の網膜のガングリオン細胞の空間周波数特性を良く近似していることが知られている [12]。

自然画像のスペクトルは、 $1/f$ に比例することが知られている [27] ので、式 (1) の $W(f)$ 項は、低周波成分の影響を弱め、高周波成分の影響を強めることで、このスペクトルを一定にする効果がある [75]。一方、 $L(f)$ 項は、ノイズ等の高周波数成分の影響を削減する効果がある。この効果を持った実空間でのコントラストフィルタは、式 (1) の $K(f)$ を周波数軸の原点周りで回転させ、更に、フーリエ逆変換することにより得られる。

式 (1) の $K(f)$ と実空間におけるコントラストフィルタを図 1(a),(b) に示す。図 1(b) から、使用するフィルタは、オン中心オフ周辺型の受容野に類似していることが分かる。顔画像への適用例を図 1(c),(d) に示す。

2.1.2 Gabor フィルタ

生体の視覚皮質 (第一次視覚野) には、方位選択性を持つ神経細胞があることが知られている。また、猫の視覚皮質の単純型細胞の受容野特性は、Gabor フィルタでうまく近似されることが知られている [20]。Gabor フィルタは、

$$\psi_{\mathbf{k}}(\mathbf{x}) = \frac{k^2}{\sigma^2} \exp\left(\frac{-k^2 \mathbf{x}^2}{2\sigma^2}\right)$$

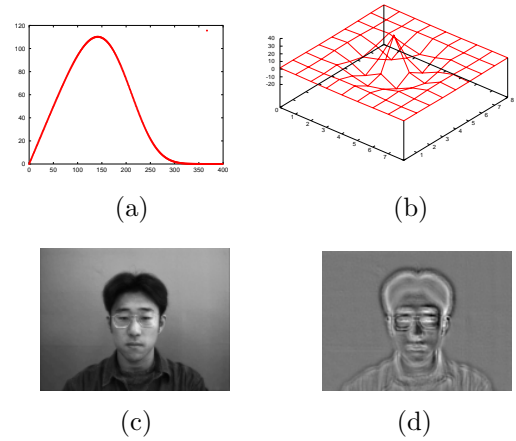


図 1: コントラストフィルタとその適用例

$$\left[\exp(i\mathbf{k}\mathbf{x}) - \exp(-\sigma^2/2)\right], \quad (2)$$

で定義される [64]。式 (2) 中のパラメタは、それぞれ、 $\mathbf{x} = (x, y)^T$, $\mathbf{k} = k_\nu \exp(i\phi) = (k_\nu \cos \phi, k_\nu \sin \phi)^T$, $k_\nu = k_{max}/f^\nu$, $\phi = \pi\mu/8$ である。

Gabor フィルタの出力は、ほとんどの出力が 0 であり、少数個の出力だけが値を持つような表現となることが知られている [20]。つまり、スパースコーディング (sparse coding) になっている。また、自然画像から切り出した局所領域のスパースさを最大にするような制約条件を用いて自己組織的にフィルタを構成したところ、Gabor フィルタに類似したフィルタが得られたことが報告されている [74]。図 2 に顔画像に対する Gabor フィルタの出力の例を示す。

2.2 4 方向面特徴

上述の特徴は人間の視覚系を模倣したものであるが、文字認識では、4 方向面特徴と呼ばれる特徴の有効性が知られている。4 方向面特徴は、各画素のエッジ勾配を 4 つの方向面にわり当てて表現したもので、上述の Gabor 特徴を簡略化したような特徴である。具体的には、Prewitt オペレータで求めた各画素のエッジ勾配を、エッジ方向に基づき、水平、垂直、右斜め、左斜めの 4 方向に分解した 4 枚の画像を作成し、それらの画像をガウシアンフィルタに通して低解像度化した特徴である。4 方向面特徴は解像度を低下させてもエッジ方向情報を保持しており、低解像度化によりエッジの形状変動に対してロバストになる。また、低解像度化することで、特徴量の数が削減できる。

本郷等は、4 方向面特徴を顔認識に応用し、顔認識

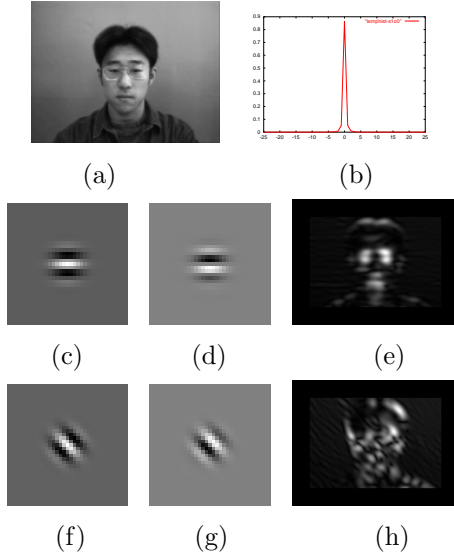


図 2: Gabor フィルタとその出力例

でもこの特徴が有効であることを示した [36, 37]。

3 正面顔認識のための統計手法

顔画像の認識でまず考えられるのは、目・鼻・口などの顔を構成する部品の形状やそれらの配置の個人差に着目し、これらから特徴点を抽出して認識に利用する方法である。しかし、顔画像からこれらの部品を精度よく抽出することはかなり難しい。また、各部品がうまく抽出できたとしても、類似した形状の差を認識に利用することはそれほど容易ではない。そこで、このような顔に特有の知識を用いるのではなく、顔画像そのものをパターンとして扱い、統計的パターン認識手法を適用する方向の研究が活発に行われている。

最も簡単なパターン認識手法は、パターン間のマッチングに基づく方法であるが、画像そのものをパターンとして扱った場合には、パターンの次元が膨大になってしまう。そのため、パターンを情報圧縮した後でマッチングを行う方法がいくつか提案されている。パターンを情報圧縮すれば、入力条件の変動に対して頑健な認識結果が得られると期待できる。Turk 等 [103, 104] が提案した固有顔 (eigenface) による方法では、主成分分析によりパターンを情報圧縮し、顔画像の識別に利用している。手法の単純さと固有顔という名前の付け方の上手さから、顔画像の認識において最も有名な手法のひとつとなっている。このような手法は、Appearance-based 顔認識手法と呼ばれている。

3.1 主成分分析を用いた顔認識

ここでは、 N 枚の顔画像のうちの i 番目の画像を、各画素の値をならべた M 次元のベクトル x_i として表現する。また、 N 枚の画像の平均ベクトルを $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ とし、各画像から平均ベクトルを引いたベクトルを $\tilde{x}_i = x_i - \bar{x}$ で表し、各画像から平均ベクトルを引いた画像の集合を行列 $\tilde{X} = [\tilde{x}_1, \dots, \tilde{x}_N]$ で表す。

画像集合を平均 2 乗誤差の意味で最適に近似する正規直交基底 U は、主成分分析 (KL 展開) を用いて構成することができる。

まず、ある正規直交基底 U が与えられたとき、ある画像 \tilde{x}_i の U の列の張る空間への射影

$$\hat{\tilde{x}}_i = UU^T \tilde{x}_i \quad (3)$$

は、平均 2 乗誤差の意味での \tilde{x}_i の近似になっている [78]。画像集合を平均 2 乗誤差

$$\varepsilon^2(U) = \frac{1}{N} \sum_{i=1}^N |\tilde{x}_i - \hat{\tilde{x}}_i|^2 \quad (4)$$

の意味で最適に近似する正規直交基底 $U = [u_1, \dots, u_L]$ は、 X の共分散行列 $\Sigma_X = \tilde{X}\tilde{X}^T$ の固有値問題

$$\Sigma_X U = U\Lambda, \quad (U^T U = I) \quad (5)$$

の解として求まる [78]。ただし、 Λ は固有値行列である。また、 U としては、固有値の大きさの順番に対応する固有ベクトルを L 個まで取るものとする。こうして求めた固有ベクトル u_i は、固有顔 (eigenface) と呼ばれている [103, 104]。

このとき、ある画像 x に対する主成分スコア (固有空間での表現) は

$$y = U^T \tilde{x} = U^T (x - \bar{x}) \quad (6)$$

のように計算される。 y の各成分は、画像 x を表現するための各固有顔の貢献度を表していると解釈できる。固有ベクトルの次元 L を小さくすることにより、固有顔 U とスコア y を用いて、もとの画像を

$$\hat{\tilde{x}} = UU^T \tilde{x} = Uy \quad (7)$$

のように低次元で近似して表現することが可能となる。

また、2 枚の画像 x_1 と x_2 が与えられた場合、各画像の主成分スコア間の距離は、

$$|y_1 - y_2|^2 = |U^T(\tilde{x}_1 - \tilde{x}_2)|^2 = |\hat{\tilde{x}}_1 - \hat{\tilde{x}}_2|^2 \quad (8)$$

のように、もとの画像の U の列の張る空間への射影 (もとの画像の平均 2 乗近似) 間の距離と同じになる。

したがって、顔画像の識別に画像の主成分スコア間の距離を用いることは、もとの画像間の距離を U の列の張る空間で近似的に計算することに対応する。

また、行列の特異値分解の関係から、

$$\tilde{X}^T \tilde{X} V = V \Lambda, \quad (V^T V = I) \quad (9)$$

のような固有値問題を考えると、固有ベクトルを要素とする行列 U と V との間には、

$$\tilde{X} V = U \Lambda^{\frac{1}{2}}, \quad \tilde{X}^T U = V \Lambda^{\frac{1}{2}} \quad (10)$$

のような関係が成り立つ。従って、画像の大きさ M に比べて画像の枚数 N が小さい場合には V に関する固有値問題を解いて、それから固有顔 U を計算すればよい。一般には、画像の大きさ M は学習に用いる画像の枚数 N よりもかなり大きいので、 V に関する固有値問題を解くことにより、必要な計算量をかなり削減できる。

Kirby 等 [48] は、100 枚の顔画像に対して主成分分析を行い固有空間の次元を 50 次まで取れば 95% 以上の情報が復元できることを示している。また、Turk 等 [103, 104] は、16 人の被験者から撮影した 2500 枚の顔画像のデータベースを用いた認識実験を行い、向きや大きさが一定なら 96% の認識率を得ている。Pentland 等 [85] は、3000 人から取った 7562 枚の顔画像のデータベースを用い、128 枚の代表的な顔画像から主成分分析により構成した 20 次元の固有空間を用いた識別で、95% の認識率を得ている。さらに、目・鼻・口などの部品に対しても固有空間を構成し、それらと顔全体の固有空間を組み合わせると認識により、98% の認識率を得ている [85]。

3.1.1 判別分析を用いた顔認識

主成分分析は、上記のように情報の圧縮という観点では有効な手法であるが、各データがどのクラスに属しているかの情報は利用しておらず、識別のための固有空間を構成する手法としては必ずしもよい手法ではない。一方、判別分析は、同一クラス内のデータはなるべく近くなり、逆にクラス間のデータはなるべく離れるような線形写像を構成する手法であり、固有顔を構成する場合にも有効である [101]。

今、画像の集合 $X = [x_1, \dots, x_N]$ には、 K 個のクラス $C_k = \{x\}$ ($k = 1, \dots, K$) のどのクラスに属しているかの情報が与えられているとする。

判別分析では、画像 x に対するスコア (固有空間での表現) を、

$$y = A^T \tilde{x} \quad (11)$$

により構成する。この時、係数行列 $A = [a_1, \dots, a_L]$ は、判別基準

$$J = \text{tr}(\hat{\Sigma}_W^{-1} \hat{\Sigma}_B) \quad (12)$$

が最大となるように決定する。ここで、 $\hat{\Sigma}_W$ および $\hat{\Sigma}_B$ は、それぞれ、スコアの空間での平均グループ内共分散行列と平均グループ間共分散行列である。この判別基準を最大とする係数行列 A は、固有値問題

$$\Sigma_B A = \Sigma_W A \Lambda, \quad A^T \Sigma_W A = I_L \quad (13)$$

の解として求まる。ここで、 L は判別空間の次元で、 $L \leq \min(K - 1, M)$ である。また、 Σ_B および Σ_W は、それぞれ、クラス間共分散行列およびクラス内共分散行列である。係数行列の各列 a_l は、主成分分析の場合と同様に、固有顔として解釈できる [101]。

判別分析の場合には、2 枚の画像 x_1 と x_2 に対する主成分スコア間の距離 $|y_1 - y_2|^2$ は、多クラスの分布間の平均マハラノビス汎距離 $(\tilde{x}_1 - \tilde{x}_2)^T \Sigma_W^{-1} (\tilde{x}_1 - \tilde{x}_2)$ と密接に関係していることが知られている [78]。つまり、判別分析の場合には、平均クラス内分散の逆 Σ_W^{-1} で重み付けた画像間の距離を近似的に計算していることに対応する。

判別分析を用いた固有顔に関しては、Swets 等 [101] が主成分分析を用いた場合との比較実験を行ない、判別分析を用いた方法の優位性を示している。

4 正面顔検出のための統計手法

与えられた画像中の顔を検出するには、顔と顔以外のものを正確に区別できる識別器が必要である。顔と顔以外の対象の識別では、顔以外の対象は無数に存在し、それらすべてを訓練データとして学習することは不可能である。また、人毎に顔の形や目鼻等の顔の部品も異なっているため、顔のクラスも変動も少なくない。そのため、未学習の顔を顔として識別でき、しかも、あらゆる顔以外の対象を顔以外に識別できるような識別器を比較的小数の訓練サンプルから学習しなければならない。そのためには、汎化性能の高い手法が必要となる。

顔検出手法には、肌色情報を用いた方法や顔の部品に注目する方法などのように顔に関する知識を利用する方法と訓練サンプルとして多くの顔画像と顔以外の対象の画像を用意し、学習により顔検出のための識別器を構成する Example-based 顔検出法がある。

肌色情報は、高速な処理が可能であること、また、画像内の顔の大きさ、回転、多少の向きの変化等に影響を受け難い等の理由から、顔の検出では頻繁に利用

されている [35, 110]。しかし、色恒常性等の難しい問題もあり、明るさが動的に変化する環境下では色情報のみで顔検出することは容易ではない。そのため、明るさの変化の影響を受け難い表色系を利用することにより対処していることが多い [109, 107, 108]。人間が顔の認識する場合には、色情報よりも濃淡情報の方が重要であるとの心理学的知見もあり [47, 18]、形状情報を全用いないで色情報のみで顔を正確に検出することは難しい。むしろ、色情報は画像中からの初期候補を絞り込むために利用するのが良いと考えられる。

心理学の知見によれば、人間は顔写真を見せられた時、目、鼻、口を重点的に見ることが報告されている [112, 6]。こうした心理学の知見に基づき、目、鼻、口を検出して、その位置から顔を検出する手法が開発されている [113]。しかし、画像中から目、鼻、口を検出する問題は、顔を検出する問題と同等の難しさを含んでおり、必ずしもそれらを安定に検出できるとは限らない。

Example-based 顔検出法では、多くの訓練サンプルから未学習のデータに対して誤り無く識別できる識別器を構成するための方法が重要となる。ニューラルネットワークは、そのための代表的な統計手法である。Rowley 等 [93] は、多層パーセプトロンを顔検出のための顔と顔以外の対象の識別に利用した。この手法は、現在、最も有名な顔検出法のひとつと見なされている。ニューラルネットに関しては、すでに多くの解説等もあるので、ここでは、顔検出に利用可能で、最近注目されている統計手法として、サポートベクターマシンとそれに関連してロジスティック回帰を紹介する。

4.1 サポートベクターマシン

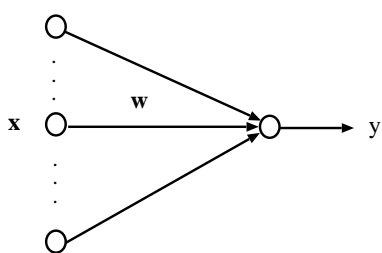


図 3: 線形しきい素子

サポートベクターマシンは、ニューロンのモデルとして最も単純な線形しきい素子を用いて、2クラスのパターン識別器を構成する手法である。訓練サンプル集合から、「マージン最大化」という基準で線形しきい

素子のパラメータを学習する。線形しきい素子は、図 3 に示すようなニューロンを単純化したモデルで、入力特徴ベクトルに対し、識別関数（線形識別関数）

$$y = \text{sign}(w^T x - h) \quad (14)$$

により 2 値の出力値を計算する。ここで、 w はシナプス荷重に対応するパラメータであり、 h はしきい値である。また、関数 $\text{sign}(u)$ は、 $u > 0$ のとき 1 をとり、 $u \leq 0$ のとき -1 をとる符号関数である。このモデルは、入力ベクトルとシナプス荷重の内積がしきい値を超えれば 1 を出力し、超えなければ -1 を出力する。これは、幾何学的には、識別平面により、入力特徴空間を 2 つに分けることに相当する。今、2 つのクラスを C_1, C_2 とし、各クラスのラベルを 1 と -1 に数値化しておくとする。また、訓練サンプル集合として、 N 個の特徴ベクトル x_1, \dots, x_N と、それぞれのサンプルに対する正解のクラスラベル t_1, \dots, t_N が与えられているとする。また、この訓練サンプル集合は、線形分離可能であるとする。すなわち、線形しきい素子のパラメータをうまく調整することで、訓練サンプル集合を誤りなく分けることができると仮定する。

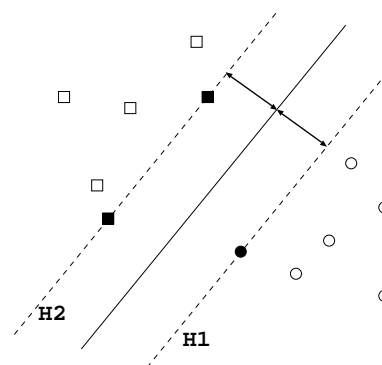


図 4: 線形しきい素子の分離超平面とマージン (がクラス 1 のサンプルで、 がクラス -1 のサンプルを示す。 と はサポートベクターを示す。)

訓練サンプル集合が線形分離可能であるとしても、一般には、訓練サンプル集合を誤りなく分けるパラメータは一意には決まらない。サポートベクターマシンでは、訓練サンプルをすれすれに通るのではなく、なるべく余裕をもって分けるような識別平面が求められる。具体的には、最も近い訓練サンプルとの余裕をマージンと呼ばれる量で測り、マージンが最大となるような識別平面を求める。もし、訓練サンプル集合が線形分離可能なら、

$$t_i(w^T x_i - h) \geq 1, \quad i = 1, \dots, N \quad (15)$$

を満たすようなパラメータが存在する。これは、H1: $w^T x - h = 1$ と H2: $w^T x - h = -1$ の2枚の超平面で訓練サンプルが完全に分離されており、2枚の超平面の間にはサンプルがひとつも存在しないことを示している。このとき、識別平面とこれらの超平面との距離(マージンの大きさ)は、 $\frac{1}{\|w\|}$ となる。したがって、マージンを最大とするパラメータ w と h を求める問題は、結局、制約条件

$$t_i(w^T x_i - h) \geq 1, \quad (i = 1, \dots, N) \quad (16)$$

の下で、目的関数

$$L(w) = \frac{1}{2} \|w\|^2 \quad (17)$$

を最小とするパラメータを求める問題と等価になる。この最適化問題は、数理計画法の分野で2次計画問題として知られており、さまざまな数値計算法が提案されている。ここでは、双対問題に帰着して解く方法を紹介する。まず、Lagrange 乗数 $\alpha_i (\geq 0)$, $i = 1, \dots, N$ を導入し、目的関数を

$$L(w, h, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i \{t_i(w^T x_i - h) - 1\} \quad (18)$$

と書き換える。パラメータ w および h に関する偏微分から停留点では、

$$w = \sum_{i=1}^N \alpha_i t_i x_i \quad (19)$$

$$0 = \sum_{i=1}^N \alpha_i t_i \quad (20)$$

という関係が成り立つ。これらを上の目的関数の式に代入すると、制約条件、

$$\sum_{i=1}^N \alpha_i t_i = 0 \quad (21)$$

$$0 \leq \alpha_i, \quad i = 1, \dots, N \quad (22)$$

の下で、目的関数

$$L_D(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j t_i t_j x_i^T x_j \quad (23)$$

を最大とする双対問題が得られる。これは、Lagrange 乗数 $\alpha_i (\geq 0)$, $i = 1, \dots, N$ に関する最適化問題となる。その解で α_i^* が0でない、すなわち、 $\alpha_i^* > 0$ となる訓練サンプル x_i は、先の2つの超平面 $w^T x - h = 1$ か $w^T x - h = -1$ のどちらかにのっている。このことから、 α_i^* が0でない訓練サンプル x_i のことを「サ

ポートベクター」と呼んでいる。これが、サポートベクターマシンの名前の由来である。直感的に理解できるように、一般には、サポートベクターは、もとの訓練サンプル数に比べてかなり少ない。つまり、沢山の訓練サンプルの中から少数のサポートベクターを選び出し、それらのみを用いて線形しきい素子のパラメータが決定されることになる。

実際、双対問題の最適解 $\alpha_i^* (i \geq 0)$ 、および停留点での条件式から、最適なパラメータ w^* は、

$$w^* = \sum_{i \in S} \alpha_i^* t_i x_i \quad (24)$$

となる。ここで、 S はサポートベクターに対応する添え字の集合である。また、最適なしきい値 h^* は、2つの超平面 $w^T x - h = 1$ か $w^T x - h = -1$ のどちらかにのっているという関係を利用して求めることができる。すなわち、任意のサポートベクター $x_s, s \in S$ から

$$h^* = w^{*T} x_s - t_s \quad (25)$$

により求まる。

また、最適な識別関数を双対問題の最適解 $\alpha_i^* (i \geq 0)$ を用いて表現すると

$$\begin{aligned} y &= \text{sign}(w^{*T} x - h^*) \\ &= \text{sign}\left(\sum_{i \in S} \alpha_i^* t_i x_i^T x - h^*\right) \end{aligned} \quad (26)$$

となる。すなわち、 $\alpha_i^* = 0$ となる多くの訓練サンプルを無視し、 $\alpha_i^* > 0$ となる識別平面に近い少数の訓練サンプルのみを用いて識別関数が構成される。ここで、重要な点は、「マージン最大化」という基準から自動的に識別平面付近の少数の訓練サンプルのみが選択されたことであり、その結果として、未学習データに対してもある程度良い識別性能が維持できていると解釈できる。すなわち、サポートベクターマシンは、マージン最大化という基準を用いて、訓練サンプルを撰択することで、モデルの自由度を抑制するようなモデル選択が行われていると解釈できる。

4.1.1 ソフトマージン

上述のサポートベクターマシンは、訓練サンプルが線形分離可能な場合についての議論であるが、パターン認識の実問題で線形分離可能な場合は稀である。したがって、実際的な課題にサポートベクターマシンを使うには、さらなる工夫が必要である。まず考えられるのは、多少の識別誤りは許すように制約を緩める方法である。これは、「ソフトマージン」と呼ばれている。

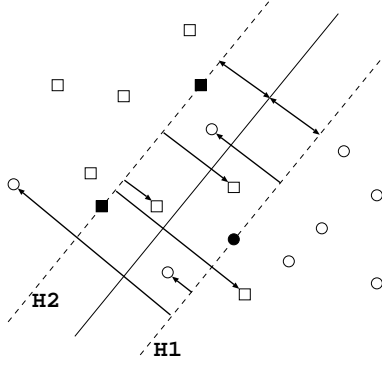


図 5: ソフトマージン (がクラス 1 のサンプルで、がクラス-1 のサンプルを示す。 と はサポートベクターを示す。)

ソフトマージン法では、マージン $\frac{1}{\|w\|}$ を最大としながら、図 5 に示すように、幾つかのサンプルが超平面 H1 あるいは H2 を越えて反対側に入ってしまうことを許す。反対側にどれくらい入り込んだかの距離を、パラメータ $\xi_i (\geq 0)$ を用いて、 $\frac{\xi_i}{\|w\|}$ と表すとすると、その和

$$\sum_{i=1}^N \frac{\xi_i}{\|w\|} \quad (27)$$

はなるべく小さいことが望ましい。これらの条件から最適な識別面を求める問題は、制約条件

$$\xi_i \geq 0, \quad t_i(w^T x_i - h) \geq 1 - \xi_i, \quad (i = 1, \dots, N) \quad (28)$$

の下で、目的関数

$$L(w, \xi) = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^N \xi_i \quad (29)$$

を最小とするパラメータを求める問題に帰着される。ここで、あらたに導入したパラメータ γ は、第 1 項のマージンの大きさと第 2 項のはみ出しの程度とのバランスを決める定数である。

この最適化問題の解法は、基本的には線形分離可能な場合と同様にふたつの制約条件に対して、Lagrange 乗数 α_i 、および、 ν_i を導入し、目的関数を

$$\begin{aligned} L(w, h, \alpha, \nu) = & \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^N \xi_i \\ & - \sum_{i=1}^N \alpha_i \{t_i(w^T x_i - h) - (1 - \xi_i)\} \\ & - \sum_{i=1}^N \nu_i \xi_i \end{aligned} \quad (30)$$

と書き換える。パラメータ w 、 h 、 ν_i に関する偏微分を 0 とする停留点では、

$$w = \sum_{i=1}^N \alpha_i t_i x_i \quad (31)$$

$$0 = \sum_{i=1}^N \alpha_i t_i \quad (32)$$

$$\alpha_i = \gamma - \nu_i \quad (33)$$

という関係が成り立つ。これらを目的関数の式に代入すると、制約条件

$$\sum_{i=1}^N \alpha_i t_i = 0 \quad (34)$$

$$0 \leq \alpha_i \leq \gamma, \quad i = 1, \dots, N \quad (35)$$

の下で、目的関数

$$L_D(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j t_i t_j x_i^T x_j \quad (36)$$

を最大とする双対問題が得られる。線形分離可能な場合には、最適解 α_i^* の値により、平面 H1 および H2 上の訓練サンプル (サポートベクター) とそれ以外のサンプルに分類されたが、ソフトマージンの場合には、さらに、H1 および H2 をはさんで反対側にはみ出すサンプルが存在する。それらは、同様に、最適解 α_i^* の値により区別することができる。具体的には、 $\alpha_i^* = 0$ なら、平面 H1 あるいは H2 の外側に存在し、学習された識別器によって正しく識別される。また、 $0 < \alpha_i^* < \gamma$ の場合には、対応するサンプルは、ちょうど平面 H1 あるいは H2 の上に存在するサポートベクターとなり、これも正しく識別される。 $\alpha_i^* = \gamma$ の場合には、対応するサンプルはサポートベクターとなるが、 $\xi_i \neq 0$ となり、平面 H1 あるいは H2 の内側に存在することになる。

4.2 ロジスティック回帰

サポートベクターマシンは、線形しきい素子を用いた識別器であるが、Rosenblatt が提案したパーセプトロンも、同様に、線形しきい素子を用いて、訓練サンプルから学習する識別機械である。サポートベクターマシンと同様に、単純パーセプトロンでは、入力 $x = (x_1, \dots, x_M)^T$ に対する出力 y を

$$\begin{aligned} y &= f(\eta) \\ \eta &= w^T x - h = \tilde{w}^T \tilde{x} \end{aligned} \quad (37)$$

のように計算する。ここで、 w_i は、 i 番目の入力から出力への結合荷重であり、 h はバイアスである。

以下では、簡単のために、これらをまとめて、 $\tilde{w} = (h, w_1, \dots, w_M)^T$ のように表すものとする。また、入力特徴ベクトルに定数項を加えたベクトルを $\tilde{x} = (-1, x_1, \dots, x_M)^T$ と表す。出力ユニットの入出力関数 f は、Rosenblatt のオリジナルなモデルではしきい関数

$$f(\eta) = \text{sign}(\eta) = \begin{cases} 1 & \text{if } \eta \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (38)$$

が用いられた。この他の入出力関数としてロジスティック関数

$$f(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)} \quad (39)$$

がよく使われる。多変量データ解析の用語を用いれば、出力ユニットの入出力関数がロジスティック関数の単純パーセプトロンは、ロジスティック回帰モデルである。ロジスティック回帰では、普通、最尤法によりパラメータが推定される。

今、訓練サンプル集合を $\{(x_i, u_i) | i = 1, \dots, N\}$ とする。ここでは、教師信号 u_i は、0 か 1 の 2 値で与えられるものとする。

入力 x を与えたときの出力 y を、入力 x のもとで教師信号 u が 1 である確率の推定値と考え、訓練サンプル集合に対するネットワークの尤度は、

$$L = \prod_{i=1}^N y_i^{u_i} (1 - y_i)^{(1-u_i)} \quad (40)$$

で与えられる。従って、その対数 (対数尤度) は、

$$\begin{aligned} l &= \sum_{i=1}^N \{u_i \log y_i + (1 - u_i) \log(1 - y_i)\} \\ &= \sum_{i=1}^N \left\{ u_i \log \left\{ \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right\} \right. \\ &\quad \left. + (1 - u_i) \log \left\{ \frac{1}{1 + \exp(\eta_i)} \right\} \right\} \\ &= \sum_{i=1}^N \{u_i \eta_i - \log\{1 + \exp(\eta_i)\}\} \end{aligned} \quad (41)$$

となる。これを最大とするパラメータがネットワークの最尤推定値である。

まずは、最急降下法によりパラメータ \tilde{w} を逐次更新することで最適なパラメータを求める方法について考えてみよう。対数尤度のパラメータ w_j に関する偏微分は、

$$\frac{\partial l}{\partial w_j} = \sum_{i=1}^N (u_i - y_i) x_{ij} = \sum_{i=1}^N \delta_i x_{ij} \quad (42)$$

のようになる。ここで $\delta_i = (u_i - y_i)$ である。一方、対数尤度のパラメータ h に関する偏微分は、

$$\frac{\partial l}{\partial h} = \sum_{i=1}^N (u_i - y_i)(-1) = \sum_{i=1}^N \delta_i(-1) \quad (43)$$

となる。したがって、パラメータの更新式は、

$$w_j \leftarrow w_j + \alpha \left(\sum_{i=1}^N \delta_i x_{ij} \right) \quad (44)$$

$$h \leftarrow h + \alpha \left(\sum_{i=1}^N \delta_i(-1) \right) \quad (45)$$

のようになる。

最尤推定においては、Fisher 情報行列が重要な役割を演じる。一般に、データ y がパラメータ $\theta_1, \dots, \theta_M$ をもつ密度関数 $f(y, \theta_1, \dots, \theta_M)$ をもつ分布に従うとき、

$$F_{ij} = -E \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(y, \theta_1, \dots, \theta_M) \right) \quad (46)$$

を Fisher 情報量と呼び、行列 $F = [F_{ij}]$ を Fisher 情報行列という。Fisher 情報量は不変推定量の分散と密接に関係している。

ここでは、ロジスティック回帰の Fisher 情報量を具体的に計算する。そのためには、式 (41) の対数尤度の 2 次微分を計算する必要がある。対数尤度の 2 次微分は、

$$\frac{\partial^2 l}{\partial \tilde{w}_k \partial \tilde{w}_j} = - \sum_{i=1}^N \omega_i \tilde{x}_{ik} \tilde{x}_{ij} \quad (47)$$

となる。ただし、 $\omega_i = y_i(1 - y_i)$ である。1 次微分と 2 次微分をまとめて行列表現すると

$$\nabla l = \sum_{i=1}^N \delta_i \tilde{x}_i = X^T \delta, \quad (48)$$

$$\nabla^2 l = - \sum_{i=1}^N \omega_i \tilde{x}_i \tilde{x}_i^T = -X^T W X$$

となる。ただし、 $X^T = [\tilde{x}_1, \dots, \tilde{x}_N]$ 、 $W = \text{diag}(\omega_1, \dots, \omega_N)$ および $\delta = (\delta_1, \dots, \delta_N)^T$ である。

これらを用いて、パラメータ \tilde{w} に対する Fisher 情報行列、すなわち、Hessian 行列の期待値のマイナスは、

$$F = -E(\nabla^2 l) = X^T W X \quad (49)$$

となる。これは、入力ベクトル $\{\tilde{x}_i\}$ の ω_i で重み付けた相関行列である。そのときの重み ω_i は、図 6 に示すような 2 次関数で、ニューロンの出力が確定している (0 あるいは 1 に近い) 場合には小さくなり、出力が不確定な (0.5 に近い) 場合には大きくなる。従っ

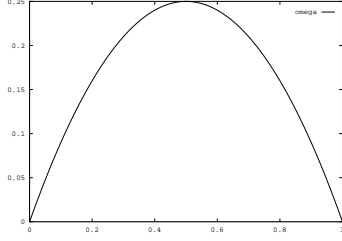


図 6: 重み ω_p

て、Fisher 情報行列は、主に、出力が不確定な入力ベクトルの相関行列であると考えられることができる。

対数尤度 (41) を最大とするようなパラメータを求めるためには、非線形最適化法を用いる必要がある。ロジスティック回帰では、このために Fisher のスコアリングアルゴリズムが使われる [68]。これは、一種のニュートン法で、Hessian 行列のかわりに Fisher 情報行列を用いる。ニューロン 1 個のみからなるネットワークの場合、Fisher 情報行列と Hessian 行列は単に符合が異なるだけなので、Fisher のスコアリングアルゴリズムはニュートン法そのものとなる。

今、現時点でのパラメータの推定値を w とし、それを修正ベクトル $\delta\tilde{w}$ により、

$$\tilde{w}^* = \tilde{w} + \delta\tilde{w} \quad (50)$$

のように更新するものとする。修正ベクトル $\delta\tilde{w}$ は、線形方程式

$$F\delta\tilde{w} = \nabla l \quad (51)$$

を解くことにより求められる。パラメータの更新式 (50) に左から F を掛けると、

$$F\tilde{w}^* = F\tilde{w} + F\delta\tilde{w} = F\tilde{w} + \nabla l \quad (52)$$

となる。今、 $F\tilde{w}$ は、

$$F\tilde{w} = X^T W \eta \quad (53)$$

となる。ただし、 $\eta = (\eta_1, \dots, \eta_N)^T$ である。従って、新しい推定値 \tilde{w}^* は、

$$\begin{aligned} \tilde{w}^* &= F^{-1}(F\tilde{w} + \nabla l) \\ &= (X^T W X)^{-1}(X^T W \eta + X^T \delta) \\ &= (X^T W X)^{-1} X^T W (\eta + W^{-1} \delta) \end{aligned} \quad (54)$$

により求めることができる。ただし、 $\delta = (\delta_1, \dots, \delta_N)^T$ である。この式は、入力データ X から目的変数 $\eta + W^{-1} \delta$ への重み付き最小 2 乗法の正規方程式とみなす

ことができる。従って、最尤推定値を求めるには、ある初期値からはじめて、この重み付き最小 2 乗法を繰り返せばよいことになる。

上記のアルゴリズムは、繰り返しアルゴリズムであるためパラメータの初期値が必要である。これは、例えば、以下のような簡単な方法で推定することが可能である。今、結合重みがすべて 0、つまり、 $w = 0$ とする。このとき、 $W = \frac{1}{4}I$ 、 $\eta = 0$ および $\delta = u - \frac{1}{2}\mathbf{1}$ である。従って、これらを (54) の計算式に代入すると、初期パラメータの推定値 \tilde{w}_0 は、

$$\tilde{w}_0 = 4(X^T X)^{-1} X^T (u - \frac{1}{2}\mathbf{1}) \quad (55)$$

となる。これは、初期パラメータを入力から教師信号 $t - \frac{1}{2}\mathbf{1}$ への線形回帰により求めることに対応している。

4.2.1 ロジスティック回帰のための汎化性向上の工夫

(1) 変数選択法

入力特徴ベクトル x の中には予測モデルにとって有用な特徴のみでなく、不要な特徴が含まれていることがある。例えば、極端な場合として、予測に全無関係な特徴が含まれているとすると、その特徴は未学習サンプルの予測には有効に働かないで、逆に予測の邪魔をすることに成りかねない。また、訓練サンプルの数に比べて入力特徴の数が多い場合には、予測モデルのパラメータを一意に決めることすらできなくなってしまう。このような場合には、特徴の中から予測に有効な特徴の部分集合を選び出して予測モデルを構築することが必要となる。このような与えられた特徴の中から予測に有効な特徴の部分集合を選び出して予測モデルを構築する手法は、変数選択法と呼ばれている。

変数選択のためには、すべての特徴の部分集合に対して、予測性能を評価する必要がある。しかし、部分集合の数は、特徴の数が増えると指数関数的に増大する。したがって、特徴の数が多い場合には、すべての部分集合に対して評価することは現実的では無い。そのため、比較的良好な特徴の部分集合を探る手法が提案されている。単純な方法としては、Forward stepwise selection あるいは、Backward stepwise selection と呼ばれる手法がある。Forward stepwise selection は、最初、特徴 1 個のみのモデルからはじめて、特徴を 1 個ずつ追加して行くことで、最も良い特徴の組を選び出す。逆に、Backward stepwise selection は、全ての特徴を含むモデルから特徴を 1 個ずつ取り除いて行くことで、最も良い特徴の組を選び出す。これらの他にも遺伝的アルゴリズムを用いて特徴の組を選択すること

なども可能である。

変数選択を行うためには、特徴の部分集合に対して学習が終わった予測モデルの予測性能を評価できなければならない。先の訓練サンプルに対する 2 乗誤差基準は、特徴の数を増やせば増やすほど小さくなるので、この基準で特徴の部分集合を選択することはできない。

予測モデルの汎化性能は、学習に用いない未知のデータに対する予測性能であるので、訓練サンプル以外に汎化性能を評価するためのデータを比較的容易に集めることができる場合には、訓練サンプル以外のサンプルに対する予測性能を評価することも可能である。つまり、汎化性能を評価するためのサンプルを用意し、そのサンプルに対する予測性能が最大となるような特徴の部分集合を選択することが可能である。この方法は、最も簡単で、最も直接的な方法であり、訓練サンプル以外にデータを集めることが可能な場合には、まず試みてみるべき方法である。

訓練サンプルを集めることが難しく、訓練サンプルが少ない場合には、訓練サンプル以外の評価用データを用意することが難しい。このような場合には、訓練サンプルのみから予測性能を評価しなければならない。かなり多くの計算量が必要であるが、計算パワーさえあれば、比較的簡単に予測性能を評価できる方法に、resampling 手法がある。leave-one-out 法は、その中でも最も単純な手法である。leave-one-out 法では、 N 個のサンプルが与えられた場合、それを $N - 1$ 個の訓練サンプルと 1 個の評価用サンプルとに分割し、 $N - 1$ 個の訓練サンプルを用いた学習結果で 1 個の評価用サンプルを評価する。このような分割の仕方は N 通りあるので、その全てに対する評価結果の平均を計算し、それを予測性能の評価値として利用する。その他、もう少し洗練された手法として、jackknife 法 [70, 99] や bootstrap 法 [23, 24, 25] 等の resampling 手法もある。resampling 手法は、コンピュータの計算パワーを最大限に利用することで、予測性能を評価する手法であり、現在のようにコンピュータの性能が急激に向上し、コンピュータの計算パワーが至る所で有り余っているような状況では、もっともっと利用しても良い手法であると考えられる。

訓練サンプルに対する 2 乗誤差基準の代わりに、予測性能を評価するための訓練サンプルのみから計算できる評価基準も提案されている。重回帰分析では、F 統計量を用いる方法もあるが、その他にも、赤池の AIC (An Information Theoretical Criterion) [2, 95] や Rissanen の MDL (Minimum Description Length) [90, 91] などの情報量基準も有名である。このような方法は、学習

は一回のみでよく、比較的簡便な評価が可能となる。重回帰分析を用いたパーセプトロンの結合係数の学習は最尤推定とみなすことができるので、学習されたパラメータを使って計算した対数尤度 (最大対数尤度) から AIC や MDL などの情報量基準を計算することにより、予測モデルの予測性能を比較することが可能となる。

(2) Weight Decay

変数選択法では、説明変数の組を選択することで、未学習のデータに対する予測性能の良いモデルを構築しようとするが、この変数選択のプロセスは、変数を選ぶ、選ばないというように離散的である。それに対して、もう少し連続的にモデルを制限する方法として、Shrinkage 法と呼ばれる方法がある。その代表的なものが、対数尤度最大化基準にパラメータが大きくなりすぎないようにペナルティ (正則化項) を追加する Weight Decay 法である。つまり、Weight Decay 法では、目的関数を、

$$\begin{aligned} Q(\tilde{w}) &= -l + \lambda \sum_{j=1}^M w_j^2 \\ &= \sum_{i=1}^N \{\log\{1 + \exp(\eta_i)\} - u_i \eta_i\} \\ &\quad + \lambda \sum_{j=1}^M w_j^2 \end{aligned} \quad (56)$$

のように変更する。これを最小化するパラメータを求めるために、 $Q(\tilde{w})$ のパラメータ w_j に関する偏微分を計算してみると、

$$\begin{aligned} \frac{\partial Q}{\partial w_j} &= -\frac{\partial l}{\partial w_j} + 2\lambda w_j \\ &= -\sum_{i=1}^N (u_i - y_i) x_{ij} + 2\lambda w_j \end{aligned} \quad (57)$$

となる。また、 $Q(\tilde{w})$ のパラメータ h に関する偏微分は、

$$\begin{aligned} \frac{\partial Q}{\partial h} &= -\frac{\partial l}{\partial h} \\ &= -\sum_{i=1}^N (u_i - y_i) (-1) \end{aligned} \quad (58)$$

となる。したがって、Weight Decay でのパラメータの更新式は、

$$w_j \Leftarrow w_j + \alpha \left(\sum_{i=1}^N (u_i - y_i) x_{ij} \right) - 2\alpha \lambda w_j \quad (59)$$

$$h \Leftarrow h + \alpha \left(\sum_{i=1}^N (u_i - y_i) (-1) \right) \quad (60)$$

となる。ここで、 w_j の更新式の第 2 項は、 w_j の絶対値を小さくする方向に作用する。つまり、予測に不必要な無駄なパラメータを 0 にするような効果がある。

5 非線形手法への拡張

これまで紹介した主成分分析、判別分析、サポートベクターマシン、ロジスティック回帰は、全て線形あるいは準線形の手法である。したがって、本質的に非線形で複雑な識別課題に対しては、必ずしも良い性能の識別器を構成できるとは限らない。本質的に非線形な問題に対応するための方法として、特徴ベクトルを非線形変換して、その空間で線形の識別を行う「カーネルトリック」と呼ばれている方法が知られている。この方法を用いることでサポートベクターマシンの性能が飛躍的に向上した。それがサポートベクターマシンを有名にした大きな要因である。

一般に、線形分離可能性はサンプル数が大きくなればなるほど難しくなり、逆に、特徴空間ベクトルの次元が大きくなるほど易くなる。例えば、特徴ベクトルの次元が訓練サンプルの数よりも大きいなら、どんなラベル付けに対しても線形分離可能である。しかし、高次元への写像を行うと、次元の増加に伴い汎化能力が落ちてしまう。また、難しい問題を線形分離可能にするためには、訓練サンプルと同程度の大きな次元に写像しなければならないので、結果的に膨大な計算量が必要となってしまふ。

今、元の特徴ベクトル x を非線形の写像 $\phi(x)$ によって変換し、その空間で線形識別を行うことを考えてみよう。例えば、写像 ϕ として、入力特徴を 2 次の多項式に変換する写像を用いるとすると、写像した先で線形識別を行うことは、元の空間で 2 次の識別関数を構成することに対応する。一般には、こうした非線形の写像によって変換した特徴空間の次元は非常に大きくなりがちである。しかし、主成分分析、判別分析、サポートベクターマシン等では、幸いにも、目的関数や識別関数が入力パターンの内積のみに依存した形になっており、内積が計算できれば最適な識別関数を構成することができる。つまり、もし非線形に写像した空間での二つの要素 $\phi(x_1)$ と $\phi(x_2)$ の内積が

$$\phi(x_1)^T \phi(x_2) = K(x_1, x_2) \quad (61)$$

のように、入力特徴 x_1 と x_2 のみから計算できるなら、非線形写像によって変換された特徴空間での特徴 $\phi(x_1)$ や $\phi(x_2)$ を陽に計算する代わりに、 $K(x_1, x_2)$ から最適な非線形写像を構成できる。ここで、このような K

のことをカーネルと呼んでいる。このように高次元に写像しながら、実際には写像された空間での特徴の計算を避けて、カーネルの計算のみで最適な識別関数を構成するテクニックのことを「カーネルトリック」と呼んでいる。

実用的には、 K は計算が容易なものが望ましい。例えば、多項式カーネル

$$K(x_1, x_2) = (1 + x_1^T x_2)^p \quad (62)$$

Gauss カーネル

$$K(x_1, x_2) = \exp\left(\frac{-\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (63)$$

シグモイドカーネル

$$K(x_1, x_2) = \tanh(ax_1^T x_2 - b) \quad (64)$$

などが使われている。

5.1 カーネル判別分析

カーネル判別分析では、画像 x に対するスコア (固有空間での表現) を N 個の訓練サンプルに対応するカーネルベースの線形結合として、

$$y = \sum_{i=1}^N a_i K(w_i, x), \quad (65)$$

により構成する。この時、 K はカーネル関数であり、 a_i は、 i 番目の訓練サンプルに対応するカーネルベースに対する係数ベクトルである。

今、 $k(x) = (K(x_1, x), \dots, K(x_N, x))^T$ を N 個の訓練サンプルに対応するカーネルベースをならべたベクトルとすると、上式 (65) は

$$y = A^T k(x) \quad (66)$$

のように書ける。ここで、 $A^T = [a_1, \dots, a_N]$ は、結合係数ベクトルをならべた行列である。

線形の場合と同様に、判別基準

$$J = \text{tr}(\hat{\Sigma}_W^{-1} \hat{\Sigma}_B) \quad (67)$$

を考え、これを最大とする係数行列を求めることを考える。ここで、 $\hat{\Sigma}_W$ および $\hat{\Sigma}_B$ は、それぞれ、スコアの空間での平均グループ内共分散行列と平均グループ間共分散行列である。この判別基準を最大とする係数行列 A は、固有値問題

$$\Sigma_B^{(K)} A = \Sigma_W^{(K)} A \Lambda \quad (A^T \Sigma_W^{(K)} A = I), \quad (68)$$

の解として求まる。ただし、行列 $\Sigma_W^{(K)}$ および $\Sigma_B^{(K)}$ は、カーネルベーストクトルにより定義された平均クラス内分散共分散行列および平均クラス間分散共分散行列である。

以上の定式化では、 N 個の訓練サンプルから $N \times L$ 次元の行列の要素を推定しなければならないので、必ずしも安定な推定結果が得られるとは考えにくい。これを安定化させるための最も簡単な方法として、平均クラス内分散共分散行列に

$$\tilde{\Sigma}_W^{(K)} = \Sigma_W^{(K)} + \alpha I \quad (69)$$

のように、対角要素に小さな値を加える方法（正則化法）が知られている。この方法で、数値計算が安定化される。また、これは、各カーネルベースに独立なノイズを加えることと等価な効果を持つ。

Liu 等は、顔認識実験によりカーネル判別分析と線形判別分析およびカーネル主成分分析とを比較し、カーネル判別分析の優位性を示した [65]。また、Yang 等は、カーネル判別分析、カーネル主成分分析、独立性分析、サポートベクターマシンを比較し、カーネル法の優位性を示している [111]。栗田等は、判別基準を顔と顔以外の対象の識別のために変更することで、カーネル判別分析を顔検出に応用した [62]。

5.2 カーネルサポートベクターマシン

サポートベクターマシンの式 (23) や式 (36) の目的関数 L_D は、

$$\begin{aligned} L_D(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j t_i t_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \\ &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j t_i t_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (70) \end{aligned}$$

のように内積をカーネルで置き換えた形に書ける。また、式 (26) から最適な識別関数は、

$$\begin{aligned} y &= \text{sign}(\mathbf{w}^{*T} \phi(\mathbf{x}) - h^*) \\ &= \text{sign}\left(\sum_{i \in S} \alpha_i^* t_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}) - h^*\right) \\ &= \text{sign}\left(\sum_{i \in S} \alpha_i^* t_i K(\mathbf{x}_i, \mathbf{x}) - h^*\right) \quad (71) \end{aligned}$$

のようにサポートベクターマシンの内積をカーネルで置き換えた形に書ける。ここで、この式にシグモイドカーネルを代入すると、いわゆる3層の多層パーセプトロンと同じ構造となる。また、Gauss カーネルを代入すると、Radial Basis Function (RBF) ネットワーク

と同じ構造になり、構造的には従来のニューラルネットワークと同じになる。しかし、カーネルトリックを用いて非線形に拡張したサポートベクターマシンでは、中間層から出力層への結合荷重のみが学習により決定され、前段の入力層から中間層への結合荷重は固定で、訓練データから機械的に求められる。また、中間層のユニット数が非常に大きく、訓練サンプル数と同じになる。つまり、カーネルトリックを用いて非線形に拡張したサポートベクターマシンでは、入力層から出力層への結合荷重を適応的に学習により求めない代わりにあらかじめ中間層に非常に多くのユニットを用意することで複雑な非線形写像を構成しようとする。

Osuna 等は、サポートベクターマシンを顔検出に応用し、その有効性を示した [76]。

6 向きの変化への対応

上述の顔認識・顔検出手法では、顔は正面を向いていると仮定していた。しかしながら、ロボットとの対話や監視などの多くの実時間での応用では、正面顔だけでなく向きの異なる顔画像の認識が必要となる。3次元対象の見えは向きに依存してかなり変化するため、向きに依らない3次元対象の認識は必ずしも容易では無い。例えば、形の上ではある人の横顔はその人の正面顔よりも他の人の横顔に似ている。一方、人間の視覚系は、このような向きによって大きく変化する3次元対象を比較的容易に認識する能力を持っている。その仕組みを模倣することで、機械による向きに依らない認識が実現できると考えられる。

3次元対象を認識するように訓練されたサル（inferotemporal cortex）野のニューロン活動を記録すると、3次元対象の向きに選択的に反応するニューロンが見つかることが報告されている [84]。これらの向きに選択的に反応するニューロンは、向きの変化に対してかなり系統的な反応を示すことが確認されている。顔の識別タスクに対しても、同様に、顔の向きに選択的に反応するニューロンがIT野で見つかり、それらのニューロンの方向選択性はかなり広く、かなり広い角度に対して反応することが知られている [86, 34]。Perrett 等 [86] は、正面、上下左右の5つの典型的な向きに対する選択性を調べ、40°位の範囲で反応することを示した。

これまでに、こうした対象の向きに対する選択性を示す生体の視覚系を説明する工学的モデルもいくつか提案されている。例えば、Poggio 等 [87] は RBF (Radial Basis Functions) ネットワークを用いて、比較的小数の

典型的な見えの補間により、3次元対象の任意の見えを生成できることを示した。また、Bartlett等[15]は、入力パターンの時間的な関係を捉えることにより、向きに依存しない顔の表現がどのように形成されるかを示すネットワークモデルを提案した。さらに、Ando等[5]は、複数の恒等写像を実現する砂時計型のニューラルネットワーク(autoencoder)を統合したネットワークを用いて、向きに依存しない顔の表現の学習を試みている。これらの手法は、顔の向きに依らない表現の自己組織化にはむいているが、顔画像の識別課題では、一般に学習データには誰の顔であるかのラベルが付いており、認識率を上げるためにはそれらの情報も利用した識別器を構成すべきである。

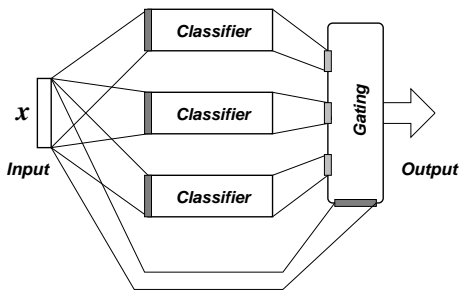


図 7: 向きに依存しない顔認識のための識別器

ここでは、著者等が提案した向きに依存する複数の識別器を統合して向きに依存しない顔画像の認識を実現する方法[60, 61]を紹介する。図7に、複数の識別器を統合して向きに依存しない顔認識を実現するためのネットワークの構造を示す。このネットワークでは、gating ネットワークにより、与えられた入力画像の向きに応じて複数の識別器の中から適切な識別器を選択することで、入力画像の向きに依らない顔画像の認識を実現する。各識別器には、multinomial logit model [68]を用いた簡単な識別器(図8)を採用した。一般化線形モデル(generalized linear model)と呼ばれる統計モデルの一つである multinomial logit model を用いて、顔の向きに依存した識別器を構成する。Multinomial logit model は、多クラスの問題のための最も簡単なニューラルネットワークモデルのひとつである。また、gating ネットワークには、恒等写像を実現する砂時計型のニューラルネットワーク(autoencoder)を用いた(図9)。gating ネットワークの中間層のニューロン間に競合を入れることで、入力画像の顔の向きに応じた撰択が自己組織的に学習可能となる。

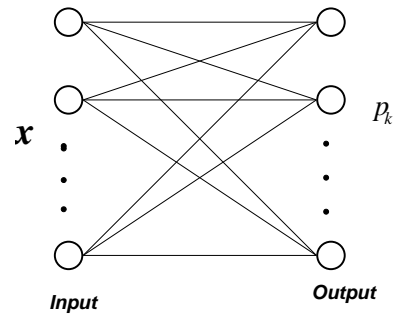


図 8: 向きに依存した識別器

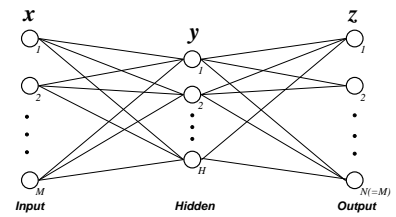


図 9: gating ネットワーク

6.1 gating ネットワーク

図7に示すような複数の識別器を統合して向きに依存しない顔認識を実現するためのネットワークでは、各識別器は入力画像が誰の顔であるかの結果を推定し、それらの推定値を、gating ネットワークの出力に応じて、どの識別器の結果を重要視するかを反映させた形で統合してネットワーク全体の推定結果とする。このネットワークが有効に働くためには、gating ネットワークが入力画像の顔の向きに応じて正しく反応するように学習される必要がある。ここでは、図9のような恒等写像を実現する砂時計型のニューラルネットワーク(autoencoder)を用いる。中間層のニューロン間に競合を入れることで向きに撰択的に反応するような自己組織化が生じる。

今、 $\mathbf{x} = (x_1, \dots, x_N)^T \in R^N$, $\mathbf{g} = (g_1, \dots, g_H)^T \in R^H$, and $\mathbf{z} = (z_1, \dots, z_N)^T \in R^N$ を、それぞれ、入力特徴ベクトル、中間層のニューロンの出力、出力層のニューロンの出力とする。この時、中間層の h 番目のニューロンの出力 g_h は、他のニューロンとの“softmax”により、

$$g_h = \frac{\exp(\gamma_h)}{1 + \sum_{j=1}^{H-1} \exp(\gamma_j)} \quad h = 1, \dots, H-1,$$

$$g_H = \frac{1}{1 + \sum_{j=1}^{H-1} \exp(\gamma_j)} \quad (72)$$

のように計算される。ここで、 $\gamma_h = \mathbf{v}_h^T \mathbf{x}$ である。ここで、 \mathbf{v}_h は、入力層から中間層の h 番目のニューロンへの結合荷重である。同様に、出力層の n 番目のニューロンの出力 z_n は、 $z_n = \mathbf{w}_n^T \mathbf{g}$ である。ここで、 \mathbf{w}_n は、中間層から出力層の n 番目のニューロンへの結合荷重である。

このネットワークの学習は、入力 \mathbf{x} と出力 \mathbf{z} との二乗誤差

$$E = \frac{1}{2} \sum_{n=1}^N (x_n - z_n)^2 \quad (73)$$

が最小となるようなパラメータが求められる。

6.2 向きに依存した識別器

各識別器には、図 8 に示すような multinomial logit model [68] を用いる。

ここでは、 $\{C_1, \dots, C_K\}$ の K 人の顔の識別問題を考えよう。

学習における教師信号は、正解のクラス C_j に対応する要素 t_j のみが 1 で、それ以外の要素が全て 0 の 2 値ベクトル $\mathbf{t} = (t_1, \dots, t_K)$ $T \in \{0, 1\}^K$ のように表されるとする。

Multinomial logit model では、識別器の k 番目の出力は、入力ベクトル \mathbf{x} とパラメータベクトル \mathbf{a} との線形結合 $\eta_k = \mathbf{a}_k^T \mathbf{x}$ の "softmax" として、

$$p_k = \frac{\exp(\eta_k)}{1 + \sum_{j=1}^{K-1} \exp(\eta_j)} \quad k = 1, \dots, K-1,$$

$$p_K = \frac{1}{1 + \sum_{j=1}^{K-1} \exp(\eta_j)}, \quad (74)$$

のように計算される。このとき、パラメータ $A = \{\mathbf{a}_1, \dots, \mathbf{a}_{K-1}\}$ は、入力層から出力層への結合荷重とみなすことができる。

このニューラルネットワークの確率モデルは、

$$P(\mathbf{t}|\mathbf{x}; A) = \prod_{k=1}^K p_k^{t_k}. \quad (75)$$

となる。両辺の対数を取ると、multinomial logit model の対数尤度

$$l = \log P(\mathbf{t}|\mathbf{x}; A) \quad (76)$$

$$= \sum_{k=1}^{K-1} t_k \eta_k - \log \left\{ 1 + \sum_{m=1}^{K-1} \exp(\eta_m) \right\}. \quad (77)$$

となる。これは、ネットワークの出力と教師信号との cross-entropy とみなすことができる。

最急降下法に基づく学習アルゴリズムは、対数尤度の勾配から求めることができ、

$$\Delta \mathbf{a}_k = \alpha \frac{\partial l}{\partial \mathbf{a}_k} = \alpha (t_k - p_k) \mathbf{x}, \quad (78)$$

となる。ここで、 α は学習係数である。

6.2.1 Mixture of Classifiers

複数の向きに依存した識別器を gating network を用いて統合することにより向きに依存しない顔画像の認識を実現する。各識別器には同一の特徴ベクトルが入力され、識別結果が出力される。gating network も識別器と同じ特徴ベクトルを入力として受け取り、その入力に対してどの識別器が適当かを選択する重みを出力する。ネットワーク全体の出力は、各ネットワークを gating network の出力で重み付けたベクトルとして計算される。すなわち、gating network は識別器を選択するスイッチのような役割を持つ。

ここでは、 H 個の識別器を統合する場合について考える。そして、 h 番目の識別器の結合加重を $A^{(h)} = \{\mathbf{a}_1^{(h)}, \dots, \mathbf{a}_{K-1}^{(h)}\}$ 、その識別器の k 番目の出力を $p_k^{(h)}$ とする。この時、 h 番目の識別器の確率モデルは、

$$P^{(h)}(\mathbf{t}|\mathbf{x}; A^{(h)}) = \prod_{k=1}^K p_k^{(h)t_k} \quad (79)$$

で与えられる。

ネットワーク全体の出力は、各識別器の確率 $P^{(h)}$ を gating ネットワークの中間層の出力値 g_h で重みつけて、

$$P(\mathbf{t}|\mathbf{x}) = \sum_{h=1}^H g_h P^{(h)}(\mathbf{t}|\mathbf{x}; A^{(h)}) \quad (80)$$

のように計算される。この対数をとると、ネットワーク全体の対数尤度は、

$$l_1 = \log \left[\sum_{h=1}^H g_h P^{(h)}(\mathbf{t}|\mathbf{x}; A^{(h)}) \right] \quad (81)$$

となる。

一方、gating ネットワークの学習には、入力と出力との二乗誤差を最小とする基準

$$l_2 = -\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - z_n)^2, \quad (82)$$

が必要である。

これらをまとめると、全ネットワークの学習のための評価それらをまとめると、全ネットワークの学習のための評価基準として、

$$L = l_1 + \lambda l_2, \quad (83)$$

が得られる。ここで、 λ は、各識別器の学習と gating ネットワークの学習のバランスをとるためのパラメータである。

この評価基準を最大とするパラメータを最急降下法で求めるとすると、 $\mathbf{a}_k^{(h)}$ に対する学習アルゴリズムは、

$$\Delta \mathbf{a}_k^{(h)} = \alpha \frac{\partial L}{\partial \mathbf{a}_k^{(h)}} = \alpha s_h (t_k - p_k^{(h)}) \mathbf{x}, \quad (84)$$

となる。ここで、 α 学習係数である。同様に、 v_h および w_m に対する学習アルゴリズムは、

$$\Delta v_h = \alpha \{ (s_h - g_h) \mathbf{x} + \lambda \sum_{j=1}^N (x_j - z_j) (w_{hj} - z_j) y_h \mathbf{x} \} \quad (85)$$

$$\Delta w_m = \alpha \lambda (x_m - z_m) \mathbf{g} \quad (86)$$

となる。

ここで、 s_h は、

$$s_h = \frac{g_h P_h(t|\mathbf{x}; A)}{\sum_{j=1}^H g_j P_j(t|\mathbf{x}; A)} \quad (87)$$

のように定義され、入力 \mathbf{x} が与えられたときに h 番目の gate が開く事後確率とみなすことができる。

著者等は、このネットワークを 25 方向から撮影した 10 人の顔の識別問題に適用し、向きに関する情報を教示しなくても向きに依存した識別器が学習により自己組織的に獲得され、どんな向きの顔に対しても安定に認識できるようになることを示した [60, 61]。

7 部分的な隠れへの対応

一般的な環境で画像を撮影すると、背景等の認識対象以外のものが画像中に含まれていたり、認識対象の一部が覆い隠されてしまうことがある。特に、顔検出や顔認識では、サングラス等によって顔の一部が覆い隠されることがある。そのような隠れや認識対象以外のものを自動的に除去する機構が実現できれば、顔認識や顔検出の性能が向上すると期待できる。

Kohonen は、線形の自己連想メモリを用いて、一部分が隠れた画像から元の画像を想起することがある程

度可能であることを示した [50]。主成分分析や恒等写像を学習する階層型のニューラルネットを用いて、与えられたデータの次元を圧縮したデータ表現を中間層に自己組織化させることで、自己連想メモリを実現することも可能であり [13]、そのような方法を用いても一部分が隠れた画像から元の画像を想起することがある程度可能であると考えられる。

人間の視覚系では、視覚経路に沿った順方向の情報の流れだけでなく、逆方向の情報の流れがあることが知られている [83]。Okajima は、これを模倣することで、認識対象を背景から分離することができることを示した [73]。

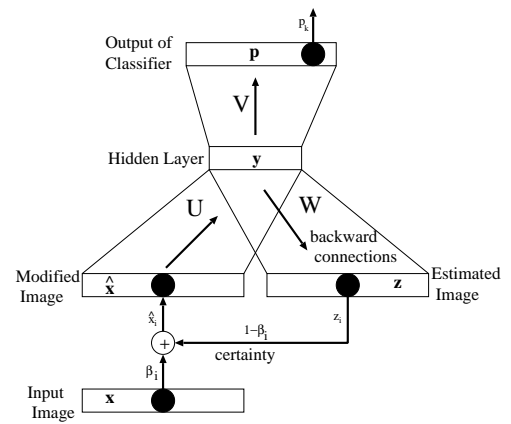


図 10: 画像の隠れ部分を例外として取り除く機構をもったニューラルネット識別器

ここでは、著者等が提案した画像の隠れ部分を例外として取り除く機構をもったニューラルネット識別器 [63] を紹介する。図 10 に、その構造を示す。このニューラルネットワーク識別器は、下段の画像の隠れや認識対象以外の部分を自動的に例外として取り除き、想起した情報で取り除いた部分を埋める機能を実現するための自己連想メモリ [102] とパターン識別器とを統合した構造をしている。

7.1 自己連想メモリ

まず、画像の隠れを自動的に例外として取り除き、取り除いた部分に想起した情報を埋めることで元の画像を想起する自己連想メモリの実現方法について考察する。ここでは、自己連想メモリのモデルとして最も簡単な中間層に線形のニューロンを持つ 3 層のパーセプトロンを利用することを考える。

7.1.1 自己連想メモリを実現する3層パーセプトロン

入力層と出力層に M 個の素子を持ち、中間層に $H (< M)$ 個の線形のニューロンを持つ3層のパーセプトロン (図 9) に、入力と出力がなるべく同じになるように学習させると、中間層には主成分分析と等価な特徴が自己組織化されることが知られている [13].

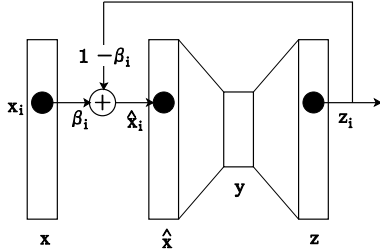


図 11: MLP for auto-associative memory.

今、このネットワークに学習させる訓練データ集合を $\{\mathbf{x}_j = (x_{j1}, \dots, x_{jM})^T \in \mathbf{R}^M\}_{j=1}^N$ とする。また、入力ベクトル \mathbf{x}_j に対する中間層のニューロンの出力をならべたベクトルを $\mathbf{y}_j = (y_{j1}, \dots, y_{jH})^T \in \mathbf{R}^H$ 、ネットワークの出力ベクトルを $\mathbf{z}_j = (z_{j1}, \dots, z_{jN})^T \in \mathbf{R}^M$ とする。さらに、入力から中間層への結合荷重を $U = [u_1, \dots, u_H]$ 、中間層から出力層への結合荷重を $W = [w_1, \dots, w_M]$ とする。このとき、中間層のニューロンの出力および出力層のニューロンの出力は、それぞれ、

$$\begin{aligned} y_{jh} &= \mathbf{u}_h^T \mathbf{x}_j & (h = 1, \dots, H) \\ z_{jm} &= \mathbf{w}_m^T \mathbf{y}_j & (m = 1, \dots, M) \end{aligned} \quad (88)$$

のように計算される。

このネットワークの学修では、入力と出力がなるべく同じになるように、すなわち、二乗誤差

$$\varepsilon^2 = \frac{1}{2} \sum_{j=1}^N \varepsilon_j^2 = \frac{1}{2} \sum_{j=1}^N \|\mathbf{x}_j - \mathbf{z}_j\|^2 \quad (89)$$

が最小となるようなパラメータ (結合荷重) が求められる。

これを、最尤推定の観点から解釈するために、入力ベクトルと出力ベクトルとの誤差 ε_j^2 が平均 0、分散 σ に従うと仮定すると、その対数尤度は、

$$-\frac{J}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^J \varepsilon_j^2 \quad (90)$$

となる。これを最大化することは、第 2 項の $\sum_{j=1}^J \varepsilon_j^2$ を最小化することと同値になる。そこで、

$$l_1 = -\frac{1}{2} \sum_{j=1}^N \varepsilon_j^2 \quad (91)$$

とおき、これを最大化する結合荷重を最急降下法で求めることを考える。そのために、 l_1 の結合荷重に関する偏微分を求めると、

$$\begin{aligned} \frac{\partial l_1}{\partial w_{mh}} &= \sum_{p=1}^P (x_{pm} - z_{pm}) y_{ph} \\ \frac{\partial l_1}{\partial u_{hn}} &= \sum_{p=1}^P \sum_{m=1}^M (x_{pm} - z_{pm}) w_{mh} x_{pn} \end{aligned}$$

となる。従って、学習則は、

$$\begin{aligned} \Delta w_{mh} &= \alpha \sum_{j=1}^N (x_{jm} - z_{jm}) y_{jh} \\ \Delta u_{hn} &= \alpha \sum_{j=1}^N \sum_{m=1}^M (x_{jm} - z_{jm}) w_{mh} x_{jn} \end{aligned} \quad (92)$$

となる。ここで、 α は学習係数である。このような学習により、 N 次元の入力画像を情報をなるべく保持したままで H 次元に圧縮し、圧縮した情報から画像を再構成するような恒等写像が構成できる。このネットワークに学習した画像と同じ画像を入力すると、その画像に近い画像を出力することができるようになる。つまり、このネットワークにより自己連想メモリが実現できる。

7.2 隠れ部分を含む画像からの元画像の想起

上述の恒等写像を学習する多層パーセプトロンに訓練データ集合として画像集合を学習させた後、入力としてある程度ノイズを含んだ画像を与えると、ノイズをある程度除去した近似画像が想起できる。しかし、画像の一部が他のものに置き換わっているような場合には、その入力に対する出力画像のすべての画素が、置き換わった部分の影響を受けてしまう。部分的に例外値を含む画像からなるべく元の画像に近い画像を想起できるようにするために、本論文では、画像の画素毎に入力と想起された出力との差に基づいた「確からしさ」を定義し、それを基にして入力画像の各画素を修正することを繰り返すことで元の画像を想起する方法を考える。図 11 に例外を含む画像から元画像を想起するネットワークの概要を示す。また、想起のための具体的な繰り返し手順を以下に示す。

STEP 0: 繰り返し回数を表す変数 t を $t = 0$ と初期化し、自己連想メモリの入力 $\hat{\mathbf{x}}(0)$ に入力画像そのもの \mathbf{x} を代入する。

STEP 1: 自己連想メモリに $\hat{\mathbf{x}}(t)$ を入力し、出力 $\mathbf{z}(t)$ を想起させる。

STEP 2: 入力画像 x と想起された画像 $z(t)$ との成分毎の二乗誤差 $\varepsilon_i^2(t) = (x_i - z_i(t))^2$ を求め、各成分の「確からしさ」 $\beta_i(t)$ を次式により計算する。

$$\beta_i(t) = \exp\left(-\frac{\varepsilon_i^2(t)}{2\sigma^2(t)}\right), \quad (93)$$

ここで、 $\sigma(t)$ は、 $\varepsilon_i(t)$ の標準偏差の推定値であり、ロバスト統計の手法 [44] を用いて、

$$\sigma(t) = 1.4826 \left(1 + \frac{5}{N-1}\right) \text{med}_i \sqrt{\varepsilon_i^2(t)} \quad (94)$$

のように計算する。なお、 $\text{med}(x)$ は、 x の中央値 (メディアン) を表す。

STEP 3: 自己連想メモリに対する新しい入力 $\hat{x}(t+1)$ を、各画素の「確からしさ」を利用して、

$$\hat{x}_i(t+1) = \beta_i(t)x_i + (1 - \beta_i(t))z_i(t) \quad (95)$$

のように計算する。繰り返し回数を $t \leftarrow t+1$ のように更新し、STEP 1 から同様の処理を繰り返す。

ここで、STEP 2 の処理は、栗田 [56] が部分的に例外値を含む画像間のマッチングのために提案したロバストテンプレートマッチングで例外値を検出する手法を「確からしさ」を求めるために拡張したものである。この手順を適当な回数繰り返すことで、例外値を無視して元の画像を想起することが可能となる。この時、式 (95) は、確からしさの高い (誤差の小さい) 画素では入力画像の情報をそのまま信頼して利用し、確からしさの低い (誤差の大きい) 画素では、入力データを無視し、代わりに自己連想メモリによって想起された推定値を信頼して利用することを意味している。

7.2.1 部分的に例外を含んだ画像の想起実験

提案した想起手法の有効性を確認するために顔画像を用いた想起の実験を行った。訓練データとしては、ARFace Database[67] の 31 名を撮影した濃淡顔画像データを用いた。図 12 (a) に顔画像の例を示す。これらの画像のサイズは、 $18 \times 25 = 450$ 画素である。したがって、入力層および出力層のニューロン数を 450 とし、中間層のニューロン数を 17 とした。

評価実験としては、まず、図 12 (b) に示すような例外値を含むような画像からの想起実験を行った。繰り返し処理を行わないで自己連想メモリのみで想起した画像を図 12 (c) に示す。また、紹介した方法を用いて想起した画像を図 12 (d) に示す。

図 13 に元の画像と想起された画像との二乗誤差の関係を示す。このグラフで横軸は、全画像中の隠れの



図 12: 部分的な隠れからの想起

割合を示す。3枚のグラフは、自己連想メモリの間層のニューロンの数を $H = 9$, $H = 17$, および $H = 21$ とした場合の結果である。各グラフ中の最も上の曲線は、繰り返しの無い単純な自己連想メモリで想起した場合で、最も下の曲線が上述の手法を用いて想起した場合の2乗誤差を示す。また、真中の曲線は、繰り返し回数 1 回のみでの想起の結果である。これらの結果から、ここで示した手法により、約 30% 程度の隠れが含まれていても元の画像がほぼ想起できることがわかる。

7.3 隠れにロバストな識別器

顔画像の検出や認識では、サングラスやマスク等で顔の一部が覆い隠されることがある。このような部分的に隠れを含む画像に対しても誤り無く認識できることが望ましい。上述の隠れ部分を含む画像から元の画像を想起できる自己連想メモリの機能を基本モジュールとして組み込むことで、部分的に隠れを含む画像に対しても安定に認識可能な識別器を構成することが狩野である。

上述の 3 層のパーセプトロンに訓練画像データ集合を学習させると、その中間層のニューロンは、主成分分析と等価な表現を持つことになる。すなわち、訓練画像の情報をなるべく保存するような低次元の表現を獲得している。そこで、図 10 に示すように、中間層のニューロンの出力を特徴ベクトルとみなし、その特徴を用いて識別器を構成する。

識別器としては、他クラスの識別問題のための最も

簡単なニューラルネットワークモデルのひとつである multinomial logit model を用いる。ここで、 K 個のクラス $\{C_1, \dots, C_K\}$ を識別する場合を考える。また、中間層には H 個の線形ニューロンがあり、教師信号としては、正解のクラス C_k に対応する要素 t_k のみが 1 で、それ以外の要素がすべて 0 の 2 値ベクトル $t = (t_1, \dots, t_K)^T \in \{0, 1\}^K$ を用いるとする。

Multinomial logit model では、識別器の k 番目の素子の出力 p_{jk} は、対応する結合加重 $v_k = \{v_{k1}, \dots, v_{kH}\}$ と中間層のニューロンの出力ベクトル y_j との線形結合 $\eta_{jk} = v_k^T y_j, (k = 1, \dots, K-1)$ の "softmax" として、

$$p_{jk} = \frac{\exp(\eta_{jk})}{1 + \sum_{i=1}^{K-1} \exp(\eta_{ji})} \quad (96)$$

$$p_{jK} = \frac{1}{1 + \sum_{i=1}^{K-1} \exp(\eta_{ji})} \quad (97)$$

のように計算される。

今、訓練サンプルを $\{(x_j, t_j)\}_{j=1}^N$ とすると、この訓練サンプルに対する対数尤度は、

$$l_2 = \sum_{j=1}^N \sum_{k=1}^{K-1} t_{jk} \eta_{jk} - \sum_{j=1}^N \log \left(1 + \sum_{i=1}^{K-1} \exp(\eta_{ji}) \right) \quad (98)$$

となる。

この基準と前述の自己連想メモリの学習のための基準を統合して、全ネットワークの評価基準を

$$l = l_2 + \lambda l_1, \quad (99)$$

のように定義する。ここで、 l_1 と l_2 との間の影響を調整するパラメータである。この評価関数を最大化するような結合荷重の更新式は、

$$\begin{aligned} \Delta v_{kh} &= \alpha \sum_{j=1}^N (t_{jk} - p_{jk}) y_{jh} \\ \Delta w_{mh} &= \alpha \lambda \sum_{j=1}^N (x_{jm} - z_{jm}) y_{jh} \\ \Delta u_{hn} &= \alpha \sum_{j=1}^N \sum_{k=1}^{K-1} (t_{jk} - p_{jk}) v_{kh} x_{jn} \\ &\quad + \alpha \lambda \sum_{j=1}^N \sum_{m=1}^M (x_{jm} - z_{jm}) w_{mh} x_{jn} \end{aligned} \quad (100)$$

となる。

7.3.1 顔認識実験

先の実験と同じ顔画像を用いて、提案した識別器の性能を評価するための実験を行なった。ここでも、ニュー

ラルネットワークの入力層、および、出力層のニューロンの数は 450 個とし、中間層のニューロン数を 17 個とした。このネットワークに 31 名の顔を識別する課題を学習させた。

図 14 に、隠れの割合と認識率の関係を示す。上の局線は隠れにロバストな識別器を用いた場合で、下の局線は隠れに対する工夫をしない識別器を用いた場合の結果である。これらの結果から、明かに、上述の手法により隠れに対してロバストになっていることがわかる。また、サングラスをした画像を用いた認識実験も行った。その結果、上述の手法では、90.32% の認識率が得られたが、工夫をしない識別器の場合には、識別率は、77.41% であった。

参考文献

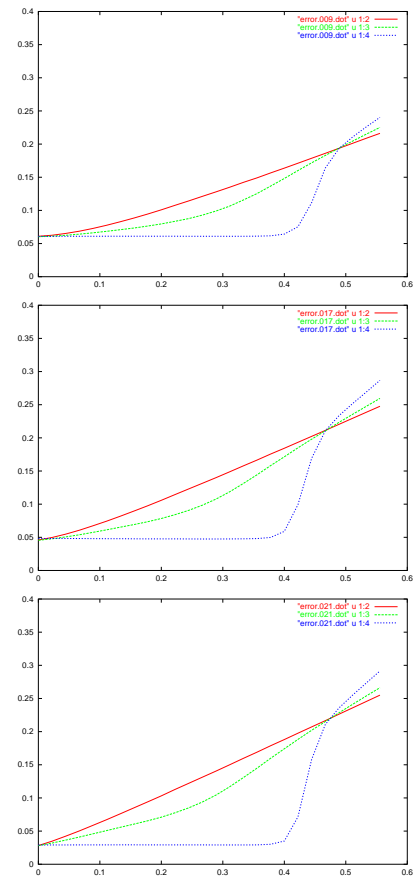
- [1] S.Akaho, "Translation, scale and rotation invariant features based on higher-order autocorrelations," *Bulletin of the Electrotechnical Laboratory*, Vol.57, No.10, pp.973-981 (1993).
- [2] H.Akaike, "A new look at the statistical model identification," *IEEE Trans. on Automatic Control*, vol.AC-19, No.6, pp.716-723 (1974).
- [3] 赤松 "コンピュータによる顔の認識の研究動向," 電子情報通信学会誌, Vol.80, No.3, pp.257-266 (1997).
- [4] 赤松, 佐々木, 深町, 末永, "濃淡画像のマッチングによるロバストな正面顔の識別法—フーリエスペクトルの KL 展開の応用—," 電子情報通信学会論文誌, Vol.J76-D-II, No.7, pp.1363-1373 (1993).
- [5] H.Ando, S.Suzuki, and T.Fujita, "Unsupervised visual learning of three-dimensional objects using a modular network architecture," *Neural Networks*, Vol.12, pp.1037-1051 (1999).
- [6] 安西, 芋阪, 前田, 彦坂, 注意と意識, 岩波書店 (1994).
- [7] Y.Ariki and Y.Sugiyama and N.Ishikawa, "Face indexing on video data - Extraction, recognition, tracking and modeling -," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp.62-69 (1998).
- [8] 麻生, 栗田, 大津, "正準相関分析および判別分析の非線形の定式化による解釈について," 行動計量学, Vol.14, No.2, pp.1-9 (1987).
- [9] H. Asoh & N. Otsu, "Nonlinear data analysis and multilayer perceptrons," *Proc. of International Joint Conference on Neural Networks*, Vol.II, pp.411-415 (1989).
- [10] H. Asoh & N. Otsu, "An approximation of nonlinear discriminant analysis by multilayer neural networks," *Proc. of International Joint Conference on Neural Networks*, Vol.III, pp.211-216 (1990).
- [11] H.Asoh, N.Vlassis, Y.Motomura, F.Asano, I.Hara, S.Hayamizu, K.Itou, T.Kuria, T.Matsui, R.Bunschoten, and B.Krose, "Jijio-2: An office robot that communicates and learns," *IEEE Intelligent Systems*, Sept./Oct., pp.46-55 (2001).

- [12] J.J. Atick and A.N. Redlich, "What does the retina know about natural scenes?" *Neural Computation*, Vol.4, pp.196-210 (1992).
- [13] P. Baldi and K. Hornik, "Neural networks and principal component analysis," *Neural Networks*, vol.2, pp.53-58 (1989).
- [14] D.H. Ballard and C.M. Brown, *Computer Vision*, Prentice-Hall (1982).
- [15] M.S. Bartlett, and T.J. Sejnowski, "Learning viewpoint-invariant face representations from visual experience in an attractor network, Network," *Comput. Neural Syst.*, Vol.9 pp.399-417 (1998).
- [16] A. Bell and T. Sejnowski, "The independent components of natural scenes are edge filters," *Vision Research*, Vol.37, No.23, pp.3327-3338 (1997).
- [17] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press (1995).
- [18] V. Bruce, P.J.B. Hancock, and A.M. Burton, "Comparisons between human and computer recognition of faces," *Proc. third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 408-413 (1998).
- [19] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc. of the IEEE*, Vol.83, NO.5, pp.705-740 (1995).
- [20] J. Daugman, "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields," *Trans. on Biomedical Engineering*, Vol.36, No.1, pp. 107-114 (1989).
- [21] M. Doi, K. Sato, and K. Chihara, "A robust face identification against lighting fluctuation for lock control," *Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp.42-47 (1998).
- [22] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, second edition, John Wiley & Sons (2001).
- [23] B. Efron, "Bootstrap methods: another look at the jackknife," *The Annals of Statistics*, Vol.7, No.1, pp.1-26 (1979).
- [24] B. Efron, "Estimating the error rate of a prediction rule: improvements in cross-validation," *Journal of American Statistical Association*, Vol.78, pp.316-331 (1983).
- [25] B. Efron, "The bootstrap method for assessing statistical accuracy," *Behaviormetrika*, Vol.17, pp.1-35 (1985).
- [26] T. Ezzat and T. Poggio, "Facial analysis and synthesis using image-based models," *Proc. of Inter. Conf. on Automatic Face and Gesture Recognition*, pp.116-121 (1996).
- [27] D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of Optical Soc. America*, A, Vol.4, No.12, pp.2379-2394 (1987).
- [28] R.A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, Vol.7-II, pp.179-188 (1936).
- [29] K. Fukunaga & S. Ando, "The optimum nonlinear features for a scatter criterion in discriminant analysis," *IEEE Trans. on Information Theory*, Vol.IT-23, No.4, pp.453-459 (1977).
- [30] P. Gallinari, S. Thiria, & F.S. Fogelman, "Multilayer Perceptrons and data analysis," *Proceeding of IEEE International Conference on Neural Networks*, Vol.I, pp.391-399 (1988).
- [31] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma, and N. Otsu, "Face recognition system using local autocorrelations and multiscale integration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.18, No.10, pp.1024-1028 (1996).
- [32] I. Hara, T. Matui, H. Asoh, T. Kurita, M. Tanaka, K. Hotta, T. Mishima, and A. Zelinsky, "Communicative functions to support human robot cooperation," *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.683-688 (1999).
- [33] O. Hasegawa, K. Itou, T. Kurita, S. Hayamizu, K. Tanaka, K. Yamamoto, N. Otsu, "Active agent oriented multimodal interface system," *Proc. IJCAI-95*, pp.82-87 (1995).
- [34] M. Hasselmo, E. Rolls, G. Baylis, and V. Nalwa, "Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey," *Exp. Brain Res.*, Vol.75, pp.417-429 (1989).
- [35] E. Hjelmas and B.K. Low, "Face detection: A survey," *Computer Vision and Image Understanding*, Vol.83, pp.236-274 (2001).
- [36] H. Hongo, M. Hoya, M. Yasumoto, Y. Niwa, and K. Yamamoto, "Focus of attention for face and hand gesture recognition using multiple cameras," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp.156-161 (2000).
- [37] 本郷, 安本, 丹羽, 山本, "適応的判別空間生成による顔認識手法の提案," 信学技報, PRMU2001-121 (2001).
- [38] K. Hotta, T. Kurita, and T. Mishima, "Scale invariant face detection method using higher-order local autocorrelation features extracted from log-polar image," *Proc. of the third IEEE International Conference on Automatic Face and Gesture Recognition*, pp.70-75 (1998).
- [39] K. Hotta, T. Kurita, and T. Mishima, "Scale invariant face recognition method using spectral features extracted from log-polar image," *Proc. of SPIE, Applications of Digital Image Processing XXII*, pp.33-43 (1999).
- [40] K. Hotta, T. Kurita, S. Umeyama, and T. Mishima, "Face matching through information theoretical attention points and its applications to face detection and classification," *Proc. of Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 34-39 (2000).
- [41] 堀田, 三島, 栗田, 梅山, "Gabor 特徴の情報量による重みづけマッチングを用いた顔認識," 信学技報, HIP2000-5, pp. 31-38 (2000).
- [42] 堀田, 三島, 栗田, "未知の画像に対する識別率を用いた顔検出のための特徴点の順序付け," 電子情報通信学会論文誌, Vol.J84-D-II, No.8, pp.1781-1789 (2001).

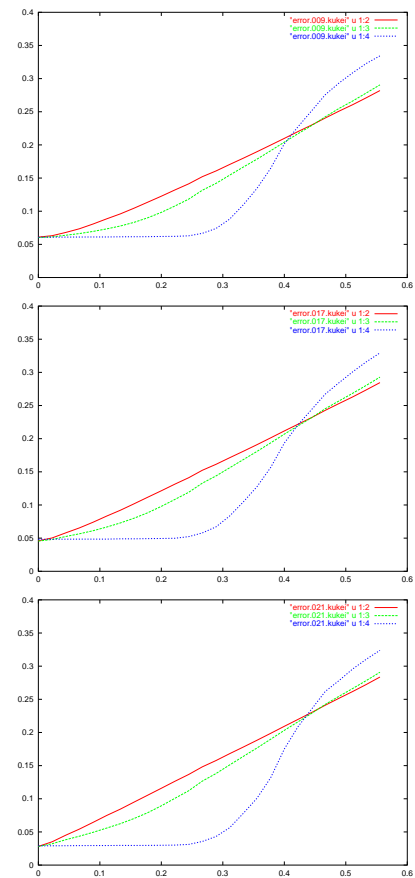
- [43] M.-H.Hsuan, D.J.Kriegman, N.Ahuja, "Detecting faces in images: a survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, AVol.24, No.1, pp.34-58 (2002).
- [44] P. J. Huber, *Robust Statistics*, John Wiley & Sons (1981).
- [45] R.A.Jacobs, M.I.Jordan, S.J.Nowlan, and G.E.Hinton, "Adaptive mixtures of local experts," *Neural Computation*, Vol.3, pp.79-87 (1991).
- [46] M.I.Jordan, and R.A.Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Neural Computation*, Vol.6, pp.181-214 (1994).
- [47] R.Kemp, G.Pike, P.White, and A.Musselman, "Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues," *Perception*, Vol.25, pp. 37-52 (1996).
- [48] M.Kirby and L.Sirovich, "Application of the Karhunen-Loève procedure for the characterization of human faces," *Trans. on Pattern Analysis and Machine Intelligence*, Vol.12, No.1, pp.103-108 (1990).
- [49] J.Kittler, M.Hatef, R.P.W.Duin, and J.Matas, "On combining classifiers," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.20, No.3, pp.223-239 (1998).
- [50] T. Kohonen, *Self-Organization and Associative Memory*, Third Edition, Springer-Verlag, Berlin (1989).
- [51] M.Kreuts, B.Volpel and H.Jansen, "Scale-invariant image recognition based on higher-order autocorrelation features," *Pattern Recognition*, Vol.29, No.1, pp.19-26 (1996).
- [52] 栗田, 麻生, 大津, "数量化手法の背後確率による解釈について," *電子情報通信学会論文誌*, Vol.J70-D, No.4, pp.760-769 (1987).
- [53] T.Kurita, N.Otsu, and T.Sato, "A face recognition method using higher order local autocorrelation and multivariate analysis," *Proc. of Int. Conf. on Pattern Recognition*, Aug. 30 - Sep. 3, The Hague, Vol.II, pp.213-216 (1992).
- [54] T. Kurita, H. Asoh, & N. Otsu, "Nonlinear discriminant features constructed by using outputs of multi-layer perceptron," *Proc. of International Symposium on Speech, Image Processing and Neural Networks*, Vol.2, pp.417-420 (1994).
- [55] 栗田, 堀田, 三島, "Log-Polar 画像の高次局所自己相関特徴を用いた大きさ・回転に不変な 2 次元図形及び顔画像の認識," *電子情報通信学会技術報告*, Vol.PRMU96-212 (1997).
- [56] 栗田, "ロバーストテンプレートマッチングとそのカット変わりの検出への応用" *電子情報通信学会総合大会*, D-12-61, pp.268, 1997.
- [57] T.Kurita, K.Hotta, and T.Mishima, "Scale and rotation invariant recognition method using higher-order local autocorrelation features of log-polar image," *Proc. of the third Asian Conference on Computer Vision*, Vol.II, pp.89-96 (1998).
- [58] 栗田, 堀田, 三島, "Cross Validation を用いた顔検出のための特徴点選択," *信学技報*, PRMU2000-82, pp.51-58 (2000).
- [59] T.Kurita, K.Hotta, and T.Mishima, "Feature Ordering by Cross Validation for Face Detection," *Proc. of IAPR Workshop on Machine Vision Applications*, pp. 211-214 (2000).
- [60] T.Kurita, H.Shimai, T.Mishima, and T.Takahashi, "Self-organization of viewpoint dependent face representation by the self-supervised learning and viewpoint independent face recognition by the mixture of classifiers," *Proc. of IAPR Workshop on Machine Vision Applications*, pp. 319-322 (2000).
- [61] T.Kurita, and T.Takahashi, "Viewpoint independent face recognition by competition of the viewpoint dependent classifiers," *Neurocomputing*, (in Press).
- [62] T.Kurita and T.Taguchi, "A modification of kernel-based Fisher discriminant analysis for face detection," *Proc. of Int'l Conf. on Automatic Face and Gesture Recognition*, pp.300-305 (2002).
- [63] T.Kurita, T.Takahashi, and Y.Ikeda, "A Neural Network Classifier for Occluded Images," *Proc. of International Conference on Pattern Recognition (ICPR2002)*, Vol.III, pp.45-48 (2002).
- [64] M.Lades, J.C.Vorbruggen, J.Buhmann, J.Lange, C.v.d.Malsburg, R.P.Wurtz, and W.Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *Trans. on Computer*, Vol.42, No.3, pp. 300-311 (1993).
- [65] Q.Liu, R.Huang, H.Lu, and S.Ma, "Face recognition using kernel based Fisher discriminant analysis," *Proc. of Int'l Conf. on Automatic Face and Gesture Recognition*, pp.197-201 (2002).
- [66] D. Lowe & A.R. Webb, "Optimized feature extraction and the Bayes decision in feed-forward classifier networks," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.13, No.4, pp.355-364 (1991).
- [67] A.M. Martinez and R. Benavente, "The AR Face Database," *CVC Technical Report*, No.24 (1998).
- [68] P.McCullagh, and J.A.Nelder FRS, "Generalized Linear Models," Chapman and Hall (1989).
- [69] J.A.Mclaughlin and J.Raviv, "Nth-order autocorrelations in pattern recognition," *Information and Control*, Vol.12, pp.121-142 (1968).
- [70] R.G.Miller, "The jackknife -a review," *Biometrika*, Vol.61, No.1, pp.1-15 (1974).
- [71] B.Moghaddam and A.Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp. 696-710 (1997).
- [72] H.Murase and Shree K.Nayar, "Visual learning and recognition of 3-D objects from appearance," *Inter. Journal of Computer Vision*, Vol.14, pp.5-24 (1995).
- [73] K.Okajima, "A recurrent system incorporating characteristics of the visual system: a model for the function of backward neural connections in the visual system," *Biological Cybernetics*, Vol.65, pp.234-241 (1991).
- [74] B.Olshausen and D.Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, Vol.381, pp.607-609 (1996).

- [75] B.A.Olshausen and D.J.Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," *Vision Research*, Vol.37, pp. 3311-3325 (1997).
- [76] E.Osuna, R.Freund, and F.Girosi, "Training support vector machines:an application to face detection," *Proc. of IEEE Computer Society conference on Computer Vision and Pattern Recognition CVPR'97*, pp.130-136 (1997).
- [77] N. Otsu, "Nonlinear discriminant analysis as a natural extension of the linear case," *Behavior Metrica*, Vol.2, pp.45-59 (1975).
- [78] 大津, "パターン認識における特徴抽出に関する数理的研究," 電子技術総合研究所研究報告, Vol.818 (1981).
- [79] N. Otsu, "Optimal linear and nonlinear solutions for least-square discriminant feature extraction," *Proceedings of the 6th International Conference on Pattern Recognition*, pp.557-560 (1982).
- [80] N. Otsu, T. Kurita, & H. Asoh, "A unified study of multivariate data analysis methods by nonlinear formulations and underlying probabilistic structures," in E. Diday, C. Hayashi, & N. Ohsumi (eds.) *Recent Developments in Clustering and Data Analysis*, Academic Press (1988).
- [81] N.Otsu, and T.Kurita, "A new scheme for practical, flexible and intelligent vision systems," *Proc. IAPR Workshop on Computer Vision*, pp.431-435 (1988).
- [82] 大津, 栗田, 関田, パターン認識—理論と応用—, 朝倉書店 (1986).
- [83] S.E.Palmer, *Vision Science: Photons to Phenomenology*, The MIT Press (1999).
- [84] J.Pauls, E.Bricolo, and N.Logothetis, "View invariant representations in monkey temporal cortex: position, scale, and rotational invariance," in: S.K.Nayer and T.Poggio ed. *Early Visual Learning*, Oxford (1996).
- [85] A.Pentland, B.Moghaddam, and T.Starner, "View-based and modular eigenspaces for face recognition," *Proc. of CVPR'94*, pp.84-91 (1994).
- [86] D.Perrett, A.Mistlin, and A.Chitty, "Visual neurons responsive to faces," *Trends Neurosci.*, Vol.10, pp.358-364 (1989).
- [87] T.Poggio, and S.Edelman, "A network that learns to recognize three-dimensional objects," *Nature*, Vol.343, pp.263-266 (1990).
- [88] B.D.Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press (1996).
- [89] M.D. Richard, & R.P. Lippmann, "Neural network classifiers estimate Bayesian *a posteriori* probabilities," *Neural Computation*, Vol.3, No.4, pp.461-483 (1991).
- [90] J.Rissanen, "A universal prior for integers and estimation by minimum description length," *The Annals of Statistics*, Vol.11, NO.2, pp.416-431 (1983).
- [91] J.Rissanen, "Stochastic complexity and modeling," *The Annals of Statistics*, Vol.14, No.3, pp.1080-1100 (1986).
- [92] S.A.Rizvi, P.J.Phillips, and H.Moon, "The FERET verification testing protocol for face recognition algorithm," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp.48-53 (1998).
- [93] H.A.Rowley, S.Baluja, and T.Kanade, "Neural network-based face detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.20, No.1, pp. 23-38 (1998).
- [94] D.W. Ruck, S.K. Rogers, M. Kaabrisky, M.E. Oxley, & B.W. Suter, "The multilayer Perceptron as an approximation to a Bayes optimal discriminant function," *IEEE Trans. on Neural Networks*, Vol.1, No.4, pp.296-298 (1990).
- [95] 坂本, 石黒, 北川, 情報量統計学, 共立出版 (1983).
- [96] A.Samal and P.A.Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, Vol.25, No.1, pp.65-77 (1992).
- [97] V.David Sanchez A., Neural network based pattern recognition, in S. K Pal and A. Pal ed. *Pattern Recognition: from classical to modern approaches*, (World Scientific, 2001), 281-300.
- [98] S.Satoh, and T.Kanade, "Name-it: Association of face and name in video," *Technical Report CMU-CS-96-205* (1996).
- [99] M.Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of Royal Statistical Society*, Vol.B36, pp.111-147 (1974).
- [100] K.Sung and T.Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.20, No.1, pp. 39-51 (1998).
- [101] D.L.Swets and J.Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.18, No.8, pp.831-836 (1996).
- [102] T.Takahashi, Y.Ikeda, T.Mishima, T.Kurita, "Robust data restoration by using autoassociative MLP," *Proc. of Annual Conference of Japanese Neural Network Society*, pp.163-164 (2001).
- [103] M.Turk and A.Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86 (1991).
- [104] M.A.Turk and A.P.Pentland, "Face recognition using eigenfaces," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.586-591 (1991).
- [105] H.D.Wactlar, T.Kanade, M.A.Smith and S.M.Stevens, "Intelligent access to digital video: The informedia project," *IEEE Computer*, Vol.29, No.5, pp.46-52 (1996).
- [106] A.R. Webb & D. Lowe, "The optimized internal representation of multilayer classifier networks performs nonlinear discriminant analysis," *Neural Networks*, Vol.3, pp.367-375 (1990).
- [107] H.Wu, Q.Chen, and M.Yachida, "Face detection from color images using a fuzzy pattern matching method," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.21, No.6, pp. 557-563 (1999).

- [108] J.-C. Terrillon, M.David, and S.Akamatsu, "Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments," *Proc. third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 112–117 (1998).
- [109] G.Xu and T.Sugimoto, "Rits eye: A software-based system for realtime face detection and tracking using pan-tilt-zoom controllable camera," *Proc. 14th International Conference on Pattern Recognition*, pp. 1194–1197, Brisbane, Australia, 1998
- [110] M.-H.Yang, D.Kriegman, and N.Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.1, pp. 34–58 (2002).
- [111] M.-H.Yang, "Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods," *Proc. of Int'l Conf. on Automatic Face and Gesture Recognition*, pp.215-220 (2002).
- [112] A.L.Yarbus, *Eye Movement and Vision*, Plenum Press (1967).
- [113] K.C.Yow and R.Cipolla, "Feature-based human face detection," *Image and Vision Computing*, Vol.15, No.9, pp 713–735 (1997).

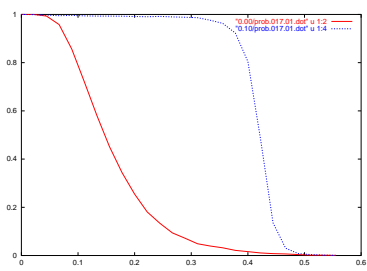


(a) 画素毎の隠れの場合: $H = 9, H = 17, H = 21$

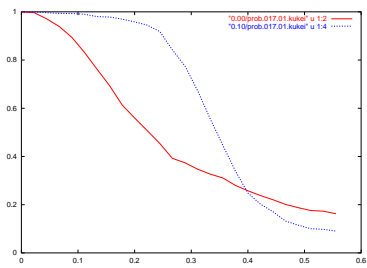


(b) 四角形の隠れの場合: $H = 9, H = 17, H = 21$

図 13: 元画像と想起された画像との二乗誤差



(a) 画素毎の隠れ



(b) 矩形の隠れ

図 14: 隠れの割合と認識率の関係