

## 7 Ia 型超新星の極大等級の変数選択

3章のパワースペクトル推定、4章の電波干渉計の画像再構成、5章のドップラートモグラフィは全て推定したい情報のスパース性を「手段」として利用することで問題を解いた応用例でした。つまり、本来は情報が足りず解けるはずのない問題を、元信号にスパース性があると仮定して解を決定したものです。一方で、 $\mathbf{y} = \mathbf{A}\mathbf{x}$  の係数ベクトル  $\mathbf{x}$  をスパースにすることを「目的」に、1次ノルム最小化を使う応用もあります。このような応用は数理的には同じ形でも、前3つの応用例とは概念的に全く異なるアプローチとなります。ここでは統計学や機械学習の分野で研究が進んでいる「変数選択」というアプローチを Ia 型超新星の極大等級に応用した例を紹介します。

Ia 型超新星の極大等級が遠方銀河の距離指標となり、観測的宇宙論を支える1つの柱となっていることはよく知られています。ただし、Ia 型超新星の極大等級が観測できれば即座に距離が推定できるわけではありません。まず、観測される等級は母銀河・天の川銀河の両方で星間ダストによる減光を受けている可能性があります。実際、観測される等級と色はよく相関し、星間赤化の影響だと考えられています。色で補正された等級は、さらに超新星の減光速度と相関があることもよく知られています。[1] Bバンドの絶対等級  $M_B$  を目的変数に、「色」( $c$ )と「減光速度」(もしくは光度曲線の幅,  $x_1$ )を説明変数にした線形モデル:  $M_B = M_{B,0} + a_1c + a_2x_1$  から係数  $a_1, a_2$  を求め、それで絶対等級を補正することによって、ようやく距離指標として使うことができます。しかし、色と減光速度で補正された絶対等級はまだ有意にばらついており、別の変数を加えることでより精度の高い距離指標が得られる可能性があります。そこで、 $M_B = M_{B,0} + a_1c + a_2x_1 + a_3X$  において、未知の説明変数  $X$  を探す研究がこれまでたくさん報告されてきました。候補としては、吸収線の等価幅、深さ、幅、速度、などで、さらには2つの吸収線の等価幅比、深さの比、そして最近ではスペクトルデータを使い、任意のフラックス比の全てを説明変数の候補とし、有効な変数を探す研究も報告されています。[2] 図1は典型的な Ia 型超新星のスペクトルです。波長 350 nm から 850 nm までのスペクトルを 134 分割してビンごとに平均したものが丸印です。先行研究で説明変数の候補として使われている「任意のフラックス比」とは、この 134 点の任意の組み合わせについてその比をとったもので、 $134 \times 133 = 17822$  個の候補を考えています。

さて、そのような先行研究のモデルは全て線形モデル

$$M_B \simeq M_{B,0} + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_Lx_L, \quad (1)$$

であり、 $M_B$  のサンプル数は数十~100。スペクトルの任意のフラックス比を説明変数候補にすると、変数の数は  $> 10000$  になります。しかし、本当に必要な変数はそのうちのごくわずかなはずです。我々は 10,000 個の変数を用いた超新星極大等級のモデルにはあまり興味がありません。極大等級を決めている数個の重要な変数を知ることが目的です(ここで、数個が具体的にいくつなのか未知であることも重要です)。したがって、係数ベクトル  $\beta$  をスパースにして解を推定するアプローチが有効です。この問題は今まで紹介してきた1次ノルム最小化の応用例と数理的には同じ形、

$$\hat{\beta}_\lambda = \arg \min_{\beta} \{ \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_1 \}, \quad (2)$$

と書くことができます。

Uemura, et al. 2015 [3] では、パークレーチームの Ia 型超新星データベースから 78 サンプルを使い、スペクトルデータを含む 300 個弱の説明変数候補に対して、このような解析を行っています。λは前章で紹介した交差検証によって決めています。詳細は論文をご覧ください。

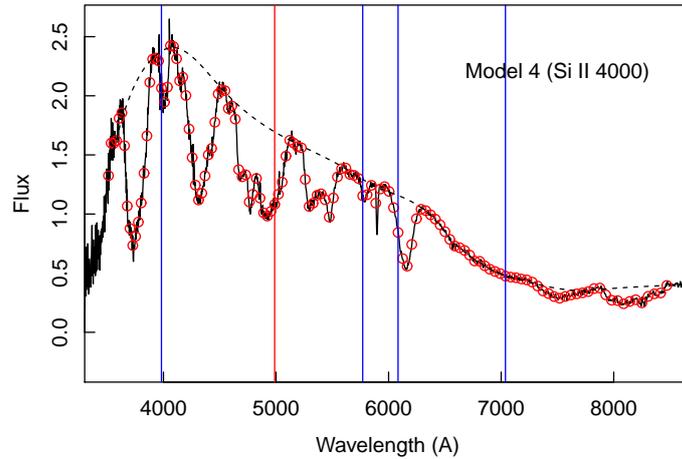


図 1: Ia 型超新星の減光速度の説明変数として選択されたフラックス。典型的なサンプル SN 2006et のスペクトルも示してある。Si II 4000, 5972, 6355 が選択されていることは従来の研究結果と一致する。[3]

結論としては、1) 絶対等級の説明変数としては色と減光速度の組み合わせが最適で、それ以外の変数はモデルを改善しない (= 300 弱の説明変数の係数の内、非ゼロ要素が2つのみ)、2) 減光速度が Si II 吸収線の強度と相関する等、従来の解析結果を再現 (図 1)、となりました。

このようにデータから変数を選択する問題は「変数選択」と呼ばれ、統計学や機械学習の分野で盛んに研究されています。興味のあるパラメータに関わりそうな要因が多数ある場合、これまではサンプル数を越えないようにあらかじめ変数を主観的に選ぶことがあったかもしれませんが、しかし、そのような問題は 1 次ノルム最小化を使うと事情が全く変わります。サンプル数よりも変数の数が多い場合でも、係数ベクトルがスパースなら解は決まるので、事前に変数を主観で絞らなくてもデータ自身が変数を選んでくれます。このようなデータ駆動型のアプローチによって、これまでは不可能と思われていた問題も扱えるようになるかもしれません。

## 参考文献

- [1] Phillips, M. M.: The absolute magnitudes of Type IA supernovae, *Astrophys. J. Letters*, 413, L105 (1993)
- [2] Silverman, J. M., Ganeshalingam, M., Li, W., Filippenko, A. V.: Berkeley Supernova Ia Program - III. Spectra near maximum brightness improve the accuracy of derived distances to Type Ia supernovae, *Mon. Not. Royal. Astron. Soc.*, 425, 1889 (2012)
- [3] Uemura, M., Kawabata, K. S., Ikeda, S., Maeda, K.: Variable Selection for Modeling the Absolute Magnitude at Maximum of Type Ia Supernovae, *Publ. Astrn. Soc. Japan*, 67, 559 (2015)