

加速器実験と機械学習

1 機械学習入門

1950年代からデータに基づいて、そのデータが持つ法則、分類などを機械に自分で発見させるやり方として**機械学習**が発展してきました。学習方法には複数あり、教師あり学習、教師なし学習、強化学習などがありますが、ここでは後述する加速器実験で主に用いられる「教師あり学習」に焦点を当てます。

1.1 サポートベクターマシン

教師あり学習は、テストデータとして $(x_1, y_1), \dots, (x_n, y_n)$ が与えられ、未知の x から y を過学習しないよう推測する**汎化**を行うことが最終的な目標です。過学習とは、テストデータで成功するが、他のデータで予測精度が下がることを指します。今回は、簡単に y を -1 または 1 に限り、

$$y = \text{sign}(-b + \sum_{i=1}^n w_i x_i)$$

なる、 w_1, \dots, w_n, b を決定することにしましょう。^{*1}分類を行うための境界線引きを**サポートベクトル**を利用した**ソフトマージン**をとります。

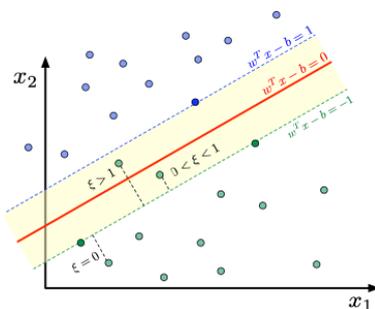


図 1: ソフトマージン ξ をとった線形サポートベクターマシン (引用: K.Crammer et al.[1])

境界線を引くには、その境界の近く

にある分類に困るデータが重要になるため、それらをサポートベクトルと呼びます。さらに過学習を防ぐべく、境界線を引く際に誤判定を許し、かつサポートベクトルのストレスを通らないように間隔を保った境界線を引く方式がソフトマージンです。他にも境界を非線形分類にも応用できる方法に**カーネルトリック**と呼ばれる手法があります。

2 素粒子実験への応用

2.1 素粒子実験の概要

身の回りの物質は原子の集合体であり、その原子は原子核と電子からできています。さらに、原子核は陽子と中性子に分類され、それらは**クォーク**と呼ばれる**素粒子**から構成されます。素粒子自体やそれらの集まった原子核の性質を明らかにする実験が行われています。クォーク同士は 10^{-13} cm 程度の狭い領域に**強い力**と呼ばれる相互作用により、複合粒子として通常現れます。素粒子自体の性質を明らかにするには、高いエネルギーの粒子を衝突させ複合した粒子を「破壊」して出た破片を観測する方法があります。

2.2 なぜ機械学習が?

加速器実験では、1秒間に約数千万、数億回の衝突(事象)が起こります。ここから、衝突後の粒子の種類、運動量、エネルギーの情報を注目している事象のみに限定して取得したいと考えます。全てのデータを保存するには膨大な量であるので、**トリガ**という仕組みでデータの取捨選択を行います。

簡単にいうと、「カメラのシャッターをいつ切るか」を決めています。より精密で鮮明な写真を取ろうとする程、「シャッターを切るタイミング」、「得られた情報からどんな粒子が飛んできたかの分類」、「データの圧縮」といった処理の実装が求められています。

2.3 応用例: 粒子識別

加速器実験にて、散乱した荷電粒子の飛跡を磁場の中で曲げ、その曲がり具合から粒子の種類を識別します。

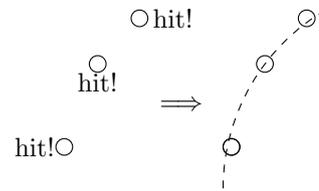


図 2: 検出器の情報と軌跡復元の様子 (検出器には位置分解能があるため、円で示しています。)

データとしては、「どの位置にいる検出器がどのタイミングで信号を受信したか」の点の情報が得られており、粒子の種類をラベルづけした -1 または 1 の数値を割り当てることが目標です。この計算は「組み合わせ最適化問題」であり、データサイズが大きくなると飛跡再構成が困難となります。テストデータには、検出器の環境を再現し、(モンテカルロ)シミュレーションを行えるソフト (Geant4 等) があり、機械学習、量子コンピュータとの関連などが研究されています。

参考文献

- [1] K.Cranmer, U.Seljak, & K. Terao, 'Machine Learning' particle data group. (2021)

^{*1} ここでは、入力特徴空間のテストデータが線形分離可能であることを仮定しています。