

線形回帰モデル

線形回帰モデルとは説明変数の線形変換により目的変数を回帰する手法である。

教師有学習の1つであり、目的変数と予測値との誤差が最も小さくなるように、最小二乗法により重み w を求める。

■回帰

説明変数に演算を施すことにより目的変数を導くこと。この時、目的変数を導く関数 f を回帰モデルと呼ぶ。

■線形変換

ベクトル $x=(x_1, x_2, \dots, x_n)$ の線形変換とは、 x の各成分の一次結合により新たな変数を得ることである。すなわち

$$y = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$$

により、 x ベクトルの線形変換 y を得ることができる。

上記関数をグラフ化すると、超平面（※ x が単成分の場合は直線）で表すことができる。

線形変換は、重みベクトル w と x との内積と一致する。

非線形回帰モデル

非線形回帰モデルとは説明変数の非線形変換により目的変数を回帰する手法である。

非線形変換は、説明変数 x_j を種々の関数で変換した $\phi_j(x_j)$ の線形結合を施すことで行われる（規定展開法）。

$$y = w_1\phi_1(x_1) + w_2\phi_2(x_2) + \dots + w_n\phi_n(x_n) + b$$

教師有学習の1つであり、規定展開法の場合、最尤法により重み w を求めることができる。

ロジスティック回帰モデル

ロジスティック回帰モデルとは説明変数に対し、線形変換→シグモイド関数という2段階の変換をかけることで $(0, 1)$ の範囲で目的変数を回帰する手法である。

$$\text{シグモイド関数 } \sigma(x) = \{1 + \exp(-ax)\}^{-1}$$

ロジスティック回帰モデルでは確率値を予測できるよう、シグモイド関数は全実数範囲に対して $(0, 1)$ の範囲の出力を行う関数となっている。

主成分分析

主成分分析は複数変数のもつ情報をなるべく損なうことなく、別のより少数の変数に変換する教師なし学習手法である。機械学習においては説明変数が多すぎると計算量が増えるだけでなく、Variance増加による汎化性能（予測精度）の低下が問題となる。そのため、主成分分析等の方法により説明変数の数を減らすことが、汎化性能向上にとってしばしば有効である。

アルゴリズム

アルゴリズムとは、説明変数から目的変数を求めるための演算手法である。説明変数に種々のアルゴリズムを施すことで、目的変数を得ることができる。アルゴリズム自体は、線形変換、規定展開法、ロジスティック回帰などSVMなど様々あるが、どのアルゴリズムで目的変数を予測しようとするかは各エンジニアが任意に定める。機械学習ではそれぞれのアルゴリズムにおいて、予測精度が最も高くなるようなパラメータをモデルが学習する。

一般的には、1つの課題に対して種々のアルゴリズムで予測をし、最も精度がよいものを採択する、または各予測の平均値や最頻値を採択する（アンサンブル学習）ことが多い。

サポートベクターマシン

サポートベクターマシン（SVM）とは、教師有2値分類モデルの1つであり、教師データを分割する超平面を想定するモデルである。数学的には、各クラスの最近傍のデータとの距離（マージン）を最大化するような超平面を求めることで学習を行う。全ての教師データに対して誤分類を一切許さないハードマージン、誤分類を許容するソフトマージンがある。また、線形分離では解けない問題に対しては、変数を非線形変換することで、変換後の変数で線形分類（変換前の元の変数で非線形分類）が可能となることがある。この時に使われる計算上の手法がカーネルトリックである。