

# 機械学習レポート

伊原 康行

## 第1章 線形回帰モデル

- 線形構造を内在する現象に対して、ある入力ベクトルから出力（連続したスカラー値）を予測する問題を扱う。
- 線形回帰では、入力ベクトルと多次元パラメータの線型結合を出力するモデルを求める。
- モデルのパラメータは、正解データから学習する教師付き学習により求める。具体的には、学習データの平均二乗誤差を最小とする様な最小二乗法により求める。

## 第2章 非線形回帰モデル

- 非線形現象を内在する現象において、入力ベクトルから出力（連続したスカラー値）を予測する問題を扱う。
- 非線形回帰では、基底関数と多次元パラメータの線型結合を出力するモデルを求める。ただし、過学習の問題が起こりやすいため、正則化項を課して関数を最小化する正則化法により問題を回避する。

## 第3章 ロジスティック回帰モデル

- ある入力ベクトルから出力（クラス、離散値）を予測するクラス分類問題を扱う。2クラス分類の場合、入力ベクトルと多次元パラメータとの線型結合をシグモイド関数に入力し、クラス1である確率を出力するモデルを求める。
- 学習データに対する対数尤度関数を最大化する様な多次元パラメータを、（確率的）勾配法により求める。

## 第4章 主成分分析

- 情報の量を分散の大きさとして捉え、学習データの分散が最大になる方向への線形変換を求める手法。正解データを用いない教師なし学習である。
- 具体的には、変換後の分散が最大となる射影軸を求めるが、解が無限個になるのを防ぐため、射影軸のノルムに制約を入れる。これにより固有値問題に帰着させて解くことができる。

## 第5章 アルゴリズム

- k近傍法...入力したテストデータの近傍k個の学習データからクラスを識別する教師あり学習。k個のクラスラベルの中で最も多いラベルを割り当てるため、kの値により分類の状況は変わってくる。
- k平均法...与えられた多数のデータをk個のクラスタに分類するクラスタリング手法（教師なし学習の1つ）。クラスタ中心の計算→クラスタ中心との距離による各データへのクラスタの割り当て、を収束するまで繰り返す。