

非線形回帰モデル

基底展開法

- 線形回帰モデルにおいて、 x の代わりに $\phi(x)$ を用いる
- $\phi(x)$ は基底関数。多項式関数 x^j やガウス型基底関数など。
- x を $\phi(x)$ に変えても、 w については線形のまま。 w の算出方法も線形回帰と同様に行える。

ガウス型基底関数

$$\phi(x) = \exp(-(x - u_j)^2 / 2h_j)$$

$2h_j$ の部分で広がりをコントロール

未学習と過学習

- 未学習 (underfitting)
 - 学習データについて十分小さな誤差が得られていない状態
 - 対策：より表現力の高いモデルを使用する
- 過学習 (overfitting)
 - 学習データについては誤差が小さいものの、検証データについて誤差が大きい状態
 - 対策1: 学習データの数を増やす
 - 対策2: 不要な基底関数 (変数) を削除して表現力を抑止
 - だが、この方法はかなり難しい
 - 対策3: 正則化法を利用して表現力を抑止
 - 係数の w をコントロール

正則化法

縮小推定

- L2ノルム: Ridge推定量。
 - 円の中から w を選ぶ
 - パラメータを0に近づけるよう推定

スパース推定

- L1ノルム: Lasso推定量。
 - 四角形の中から w を選ぶ。
 - いくつかのパラメータを正確に0に推定。

ホールドアウト

- 有限データを学習用とテスト用の2つに分割
- 先頭から20%は検証用、残りは学習用と決めてしまう手法
- 手元にデータが大量にある場合ならよいが、少量しかない場合、外れ値が検証データに入ってしまう恐れがある
- 1つのモデルに対して1回しか検証しない

クロスバリデーション（交差検証）

- データを学習用と検証用に分割
- 例：5分割した場合、それぞれ違う1/5データが検証データとなり、5回検証が行われる。
 - 各検証での精度の平均をcv値とし、最もcv値がよいものを採用する