前段講義

機械学習モデリングプロセス

1. 問題設定 ：機械学習を使わない方針を十分検討する。
2. データ選定 ：データの精度が重要。GIGO（GarbageInGarbageOut）（※ごみをいれるとばごみがでる）
3. データの前処理 ：ほとんどがこの作業
4. 機械学習モデルの選定 ：モデル構築は1〜2割
5. モデルの学習（パラメータ推定）
6. モデルノの評価

線形回帰モデル

線形とは・・・比例

n次元空間における超平面の方程式 y=a0 + a1*x1 + a2*x2 + ・・・・an-1*xn-1

= a0 + Σ (ai * xi)

教師あり学習

入力とm次元パラメータの線形結合を出力するモデル。

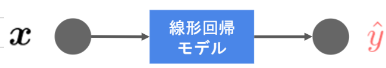
パラメータ $\boldsymbol{w} = (w_1, w_2, \cdots, w_m)^T \in \mathbb{R}^m$

線形結合 $\hat{y} = \boldsymbol{w}^T\boldsymbol{x} + w_0 = \sum_{j=1}^{m} w_j x_j + w_0$

回帰問題 ：ある入力から出力（連続値）を予測する問題。

線形回帰 ：直線で予測

非線形回帰 ：曲線で予測



※慣例的に予測値にはハットをつける。

回帰で扱うデータ

入力 ：説明変数、及び特徴量

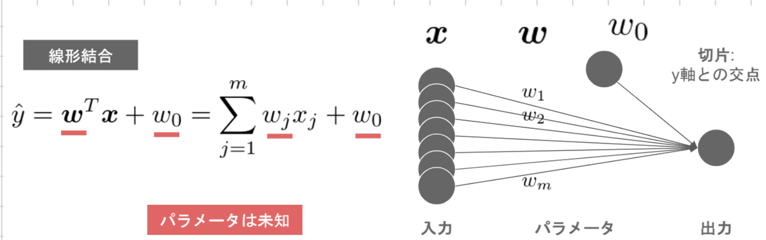
出力 ：目的変数

線形結合 ・・・入力とパラメータの内積

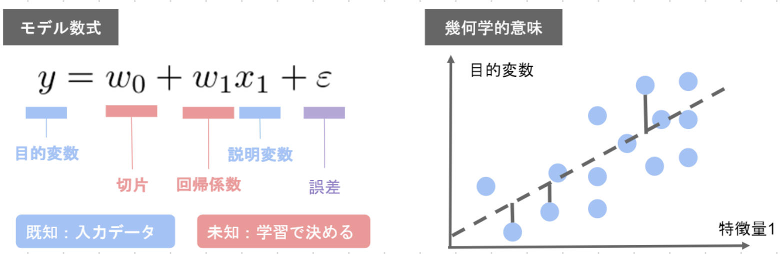
入力ベクトルと未知のパラメータの各要素を掛け算し足し合わせたもの

入力ベクトルの線形結合に加え、切片も足し合わせる。

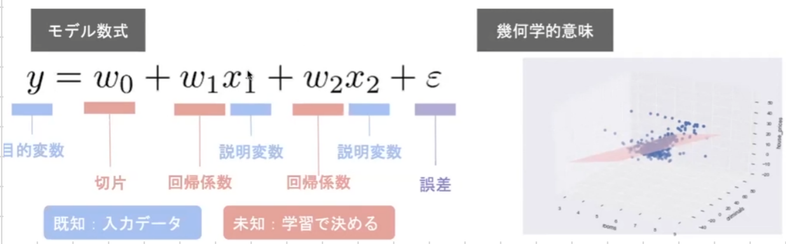
（入力ベクトルが多次元でも）出力は一次元（スカラ）となる。



■説明変数が1次元の場合：単回帰モデル



■説明変数が多次元の場合：重回帰モデル



データの分割とモデルの汎化性能測定

データの分割

学習用データ

検証用データ

汎化性能(Generalization)を測定するためにデータを分割する。

線形回帰モデルのパラメータは最小二乗法で推定

■平均二乗誤差 MSE(MeanSquareError)

マイナス値を除去するために2乗。

デメリット:外れ値があると汎化に影響が出やすい。

Huber損失 :ロバスト回帰で使われる損失関数の一つ。二乗誤差損失よりも外れ値に敏感ではない。

Huber損失関数の定義は以下の通り[1]。

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \le \delta, \\ \delta(|a| - \frac{1}{2}\delta), & \text{otherwise.} \end{cases}$$

Tukey損失:

■最小二乗法

学習データの平均二乗誤差を最小とするパラメータを探索

学習データの平均二乗誤差を最小化は、その勾配が0になる点を求めればよい

$$\mathrm{MSE}_{\mathrm{train}} = \frac{1}{n_{\mathrm{train}}} \sum_{i=1}^{n_{\mathrm{train}}} (\hat{y}_i^{(\mathrm{train})} - y_i^{(\mathrm{train})})^2$$

※参考書:MatrixCookBook

→行列計算参考書

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w} \in \mathbb{R}^{m+1}} \mathrm{MSE}_{\mathrm{train}}$$

MSEを最小にするようなw(m次元)

$$\frac{\partial}{\partial \boldsymbol{w}} \mathrm{MSE}_{\mathrm{train}} = 0$$

MSEをwに関して微分したものが0
となるwの点を求める

$$\Rightarrow \frac{\partial}{\partial \boldsymbol{w}} \left\{ \frac{1}{n_{\mathrm{train}}} \sum_{i=1}^{n_{\mathrm{train}}} (\hat{y}_i^{(\mathrm{train})} - y_i^{(\mathrm{train})})^2 \right\} = 0$$

平均二乗誤差(残差平方和)を微分

$$\Rightarrow \frac{1}{n_{\mathrm{train}}} \frac{\partial}{\partial \boldsymbol{w}} \left\{ (X^{(\mathrm{train})}\boldsymbol{w} - \boldsymbol{y}^{(\mathrm{train})})^T (X^{(\mathrm{train})}\boldsymbol{w} - \boldsymbol{y}^{(\mathrm{train})}) \right\}$$

行列表現

$$\Rightarrow \frac{1}{n_{\mathrm{train}}} \frac{\partial}{\partial \boldsymbol{w}} \left\{ \boldsymbol{w}^T X^{(\mathrm{train})T} X^{(\mathrm{train})} \boldsymbol{w} - 2\boldsymbol{w}^T X^{(\mathrm{train})T} \boldsymbol{y}^{(\mathrm{train})} + \boldsymbol{y}^{(\mathrm{train})T} \boldsymbol{y}^{(\mathrm{train})} \right\}$$

{} 展開

$$\Rightarrow 2X^{(\mathrm{train})T} X^{(\mathrm{train})} \boldsymbol{w} - 2X^{(\mathrm{train})T} \boldsymbol{y}^{(\mathrm{train})} = 0$$

行列微分計算

$$\Rightarrow \hat{\boldsymbol{w}} = (X^{(\mathrm{train})T} X^{(\mathrm{train})})^{-1} X^{(\mathrm{train})T} \boldsymbol{y}^{(\mathrm{train})}$$

回帰係数

回帰係数

$$\hat{\boldsymbol{w}} = (X^{(\mathrm{train})T} X^{(\mathrm{train})})^{-1} X^{(\mathrm{train})T} \boldsymbol{y}^{(\mathrm{train})}$$

予測値

$$\hat{\boldsymbol{y}} = \underset{n_{new} \times m+1}{X} (X^{(train)T} X^{(train)})^{-1} X^{(train)T} \boldsymbol{y}^{(train)}$$