二乗誤差と最小二乗法

データの総数を N とすると、二乗誤差は、次のような数式で表される

ref: 線形代数の半歩先 p123〜127

$$J(\boldsymbol{w}) = \sum_{n=1}^{N} (y_n - f(\boldsymbol{x}_n))^2$$

n 番目の実際の出力 y_n と、n 番目の入力を使ったときのモデルの出力 $f(\boldsymbol{x}_n)$ との差を見ている

符号を正にするために二乗し、それをすべてのデータについて合計したものが二乗誤差である

この誤差関数 $J(\boldsymbol{w})$ を最小にするパラメータを探すことが目標となる

モデルの式を整理する

まずは、モデルの式を整理する

$$f(\boldsymbol{x}) = w_0 + \sum_{d=1}^{D} w_d x_d$$

右辺はベクトルの内積で書けそうだが、w_0 が余分なので、$x_0 = 1$ と定義して、次のように書き換える

$$f(\boldsymbol{x}) = \sum_{d=0}^{D} w_d x_d$$

そして、次のようなベクトルを導入する

$$\boldsymbol{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{bmatrix}, \quad \boldsymbol{x}' = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}, \quad \boldsymbol{x}'_n = \begin{bmatrix} 1 \\ x_{n,1} \\ x_{n,2} \\ \vdots \\ x_{n,D} \end{bmatrix}$$

すると、先ほどのモデルの式は、次のように $\boldsymbol{w}$ と $\boldsymbol{x}'$ の内積で表せる

$$f(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{x}'$$

N 個分のデータをまとめる

N 個分のデータをまとめた出力 $\boldsymbol{y}$ と入力 X を、それぞれ次のように書く

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,D} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,D} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & x_{N,2} & \cdots & x_{N,D} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{x}'_1)^\top \\ (\boldsymbol{x}'_2)^\top \\ \vdots \\ (\boldsymbol{x}'_N)^\top \end{bmatrix}$$

X は $N \times (D+1)$ 行列で、定数項の分だけ列が一つ増えている

この定数項の列を含まず、データだけを並べたものはデータ行列と呼ばれる

ただし、ここでは定数項の列を含めた X もデータ行列と呼ぶことにする

誤差関数をベクトルと行列で表す

ここまでの記号を使って、誤差関数 $J(\boldsymbol{w})$ を書き直す

まずは n 番目のデータにのみ注目すると、実際の値とモデルの差は、

$$\begin{aligned} y_n - f(\boldsymbol{x}_n) &= y_n - \boldsymbol{w}^\top \boldsymbol{x}'_n \\ &= y_n - (\boldsymbol{x}'_n)^\top \boldsymbol{w} \end{aligned}$$

内積の順番を変える

ベクトルと行列を使うと、N 個のデータに対しては次のように書ける

$$\boldsymbol{z} = \begin{bmatrix} y_1 - (\boldsymbol{x}'_1)^\top \boldsymbol{w} \\ y_2 - (\boldsymbol{x}'_2)^\top \boldsymbol{w} \\ \vdots \\ y_N - (\boldsymbol{x}'_N)^\top \boldsymbol{w} \end{bmatrix} = \boldsymbol{y} - X\boldsymbol{w}$$

この二乗をとった形は、$\boldsymbol{z}$ 自身との内積で書き表せる

$$J(\boldsymbol{w}) = \boldsymbol{z}^\top \boldsymbol{z} = (\boldsymbol{y} - X\boldsymbol{w})^\top (\boldsymbol{y} - X\boldsymbol{w})$$

ベクトルの微分で最小化問題を解く

誤差関数を最小にする $\boldsymbol{w}$ を求めるには、

$$\frac{\partial J(\boldsymbol{w})}{\partial \boldsymbol{w}} = 0$$

を解けばよい

[Todo 1: ref: 線形代数の半歩先 p125〜127]

. . . .

Zebra Notes

Type	Number
todo	1