* 線形重回帰 (multiple linear regression : MLR).

・ データセット $\mathcal{D} = \{(x_1, y_1), \cdots, (x_N, y_N)\}$: given.

$$x_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,D} \end{pmatrix} \in \mathbb{R}^D, \quad y_i \in \mathbb{R} \quad (i = 1, \cdots, N) \quad \text{とする.}$$

(サンプル数 N コ, 特徴量 D コ.)

・各特徴量について, "$a := b$" は "a を b と定義する" の意味.

$$\text{標本平均}: \quad \bar{x}_j := \frac{1}{N} \sum_{i=1}^{N} x_{ij} \qquad (j = 1, \cdots, D)$$

$$\text{標本分散}: \quad \sigma_j^2 := \frac{1}{N} \sum_{i=1}^{N} (x_{ij} - \bar{x}_j) \qquad (j = 1, \cdots, D)$$

とし, 標準化 (standardize)

"$a \leftarrow b$" は "a に値 b を代入する" の意味.

$$x_{ij} \leftarrow \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

を行うことで, データセットの各サンプルに対し平均 0, 分散 1 が成立しているとしても良い.

・ 以下では, 扱うサンプルは全て標準化されているとする.

・ 計画行列 (design matrix)

N×D 行列全体の集合

$$X := \begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix} = \left.\overbrace{\begin{pmatrix} x_{1,1} & \cdots & x_{1,D} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,D} \end{pmatrix}}^{N}\right\} D \in M_{N,D}(\mathbb{R})$$

$$\mathbf{y} := \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \in \mathbb{R}^N$$ とする.

- 線形重回帰モデル：各サンプル $(\mathbf{x}_i, y_i) \in \mathcal{D}$ に対して,

$$y_i \simeq \sum_{j=1}^{D} b_j x_{ij} \qquad (i = 1, \cdots, N)$$

と線形の近似ができると仮定するモデル.

$$\mathbf{b} := \begin{pmatrix} b_1 \\ \vdots \\ b_D \end{pmatrix} \in \mathbb{R}^D$$ とすると, 上の近似式は

$y_i \simeq \mathbf{x}_i^\top \mathbf{b} \quad (i = 1, \cdots, N)$ とかける.

全サンプルについてまとめて書くと, $\mathbf{y} \simeq X\mathbf{b}$ と表すことができる

$\boxed{\text{やりたいこと}}$ 近似 $\mathbf{y} \simeq X\mathbf{b}$ がなるべく「良くなる」ような係数 $\hat{\mathbf{b}}$ を決定したい.

→ y の値が未知のデータ $\mathbf{x}_* \in \mathbb{R}^d$ に対しても, $\hat{y} = \mathbf{x}_*^\top \hat{\mathbf{b}}$ により y の値を推定できてすごい!

Q. 近似が「良い」とはどういうこと?

A. $\mathbf{y} \simeq X\mathbf{b}$ の近似で表現できなかった誤差項

$\boldsymbol{\varepsilon} := \mathbf{y} - X\mathbf{b} \in \mathbb{R}^N$ (残差ベクトル) の大きさが小さいということ.

ある性質をみたしていれば
どのようなものでも良い.
詳しくは線形代数で.

ε の「大きさ」とは ε のノルムのこと. ここでは次のように定義する:

$$\|\varepsilon\| := \sqrt{\varepsilon^T \varepsilon} \quad (\text{内積から定まるノルム}).$$

・先述の やりたいこと を数学的に書き直してみる:

(*) $\|y - Xb\|$ を最小化するような X の係数 $\hat{b}$ を決定したい.

この問題 (*) を

with respect to. (~に関して)

$$\min \|y - Xb\| \quad \text{w.r.t.} \ b$$

とか

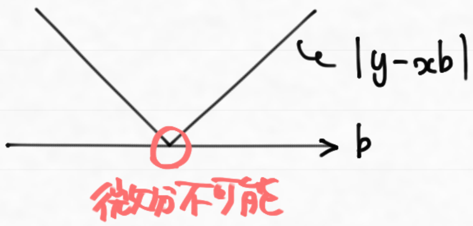
“$\mathop{\mathrm{argmin}}_x f(x)$” は “$f(x)$ を最小にする x” を表す

$$\hat{b} = \mathop{\mathrm{argmin}}_b \|y - Xb\|$$

などと書く.

・ところで, $\|y - Xb\| = \sqrt{(y - Xb)^T (y - Xb)}$ の最小化はめんどう.

(イメージ: 右のような関数の最小化.
最小点で微分できないので
とてもつらい.)



→ $\|y - Xb\| \geq 0$ なので,

全ての x で $f(x) \geq 0$ なら, どんな a, b に対しても
$f(a) \geq f(b) \Leftrightarrow f(a)^2 \geq f(b)^2$.

$$\mathop{\mathrm{argmin}}_b \|y - Xb\| = \mathop{\mathrm{argmin}}_b \|y - Xb\|^2$$

となっている! 結局, 残差のノルムの2乗を最小化すればよい.
(最小二乗法)

- $\hat{\mathbb{b}} = \underset{\mathbb{b}}{\operatorname{argmin}} \|\mathbb{y} - X\mathbb{b}\|^2$ を考える．

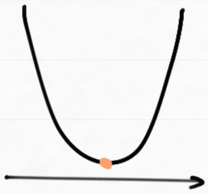
最小化するべき関数（目的関数）$\|\mathbb{y} - X\mathbb{b}\|^2$ は，

$\mathbb{b}$ について 2次関数．

→ 唯一の極小値で最小となるような

とてもよい最小化問題になっている．

（イメージ



本当はこれの多次元版．すごそう）

専門用語では「凸最適化問題」というが気にしないでよい

極小値を与える $\mathbb{b}$ をみつければよい！

→ $\dfrac{\partial}{\partial \mathbb{b}} \|\mathbb{y} - X\mathbb{b}\|^2 = \mathbb{0}$ をみたす $\mathbb{b}$ が求める $\hat{\mathbb{b}}$．

（Remark：

$$\frac{\partial}{\partial \mathbb{b}} := \begin{pmatrix} \frac{\partial}{\partial b_1} \\ \vdots \\ \frac{\partial}{\partial b_N} \end{pmatrix}$$ と定義される．$$\frac{\partial}{\partial \mathbb{b}} f(\mathbb{b}) = \begin{pmatrix} \frac{\partial}{\partial b_1} f(\mathbb{b}) \\ \vdots \\ \frac{\partial}{\partial b_N} f(\mathbb{b}) \end{pmatrix}.$$ ）

ベクトル微分の公式を使うと，

$$\frac{\partial}{\partial \mathbb{b}} \|\mathbb{y} - X\mathbb{b}\|^2 = -2X^T(\mathbb{y} - X\mathbb{b})$$

X^T は一般に正方行列でない．「左から $(X^T)^{-1}$ をかけて $\mathbb{y} - X\mathbb{b} = \mathbb{0}$」とかいうヤバいことはやってはいけない．

となるので，$\hat{\mathbb{b}}$ は

$$-2X^T(\mathbb{y} - X\hat{\mathbb{b}}) = \mathbb{0} \iff X^T(\mathbb{y} - X\hat{\mathbb{b}}) = \mathbb{0}$$

をみたす．

$X^T(y - X\hat{b}) = 0$ の左辺を展開して整理すると.

$$X^T X \hat{b} = X^T y \quad (正規方程式)$$

を得る.

- $X^T X$ は $D \times D$ 正方行列.

rank $X = D$ なら
rank $X^T X = D$ となり
$(X^T X)^{-1}$ が存在.

$X^T X$ に逆行列が存在するなら,

$$\hat{b} = (X^T X)^{-1} X^T y$$

と求まる！ めでたいねえ