2. Inputs : Variable

Outputs :

X : vector

$$\underline{\underline{X}} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}_{3 \times 1}$$

Task$_1$: relate

Learner: algo.

Classification

Regression

$\underline{\underline{Y}}$: quantitative

$\underline{\underline{G}}$:

10 = N "observations" i^{th} observed input : x_i is p-dimensional

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_{10}^T \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_{10} \end{bmatrix}$$

10x1

N x p matrix.

10x3

$$(X, Y) \longrightarrow \text{function } \text{"}f\text{"} \; X = \hat{Y} \approx Y$$

3. Linear model

K-nearest neighbor

$$f(X) = \hat{Y} \approx Y$$

p dim

1 dim vector

$$X^T \underline{\underline{\beta}} = [X_1 \; X_2 \; X_3] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

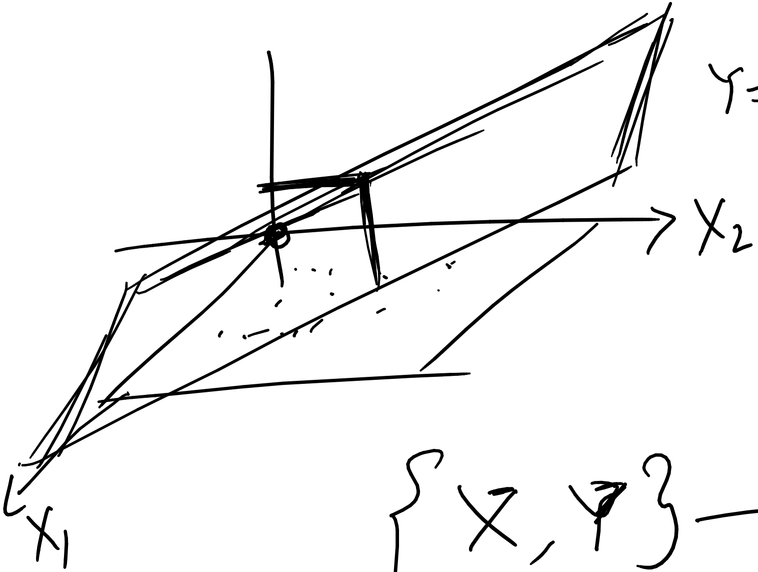
eg: squared loss $(X^T\beta - Y)^2$ | $\beta^T X = X^T \beta$

$\underline{\underline{X, Y}}$ $\quad \underbrace{\frac{1}{10}\sum_{i=1}^{10}(x_i^T\beta - y_i)^2}$ lower is better.

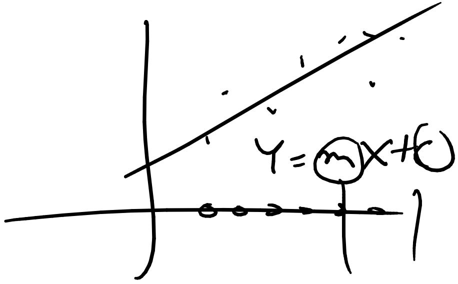
$|X^T\beta - Y|$

$\hat{\beta}$

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T Y}$$



$$Y = \hat{\beta}^T X$$



$$\underline{\{X, Y\}} \longrightarrow p \text{ numbers}$$

$$\frac{d}{d\beta}\left((10-3\beta)^2 + (5-4\beta)^2\right) = 0$$

$$\hat{\beta} =$$

Knn

X, Y

$N \times p \quad N \times 1$

$N \times (p+1)$

given new input x

$$\hat{y} = \frac{1}{K} \sum_{\substack{x_i \text{ which are in the} \\ \text{neighborhood of } x}} y_i$$

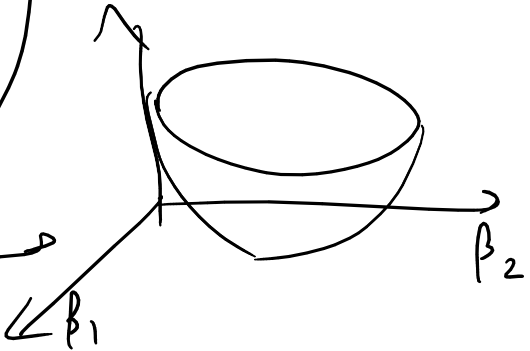
$= \text{Knn}(x)$

$$\sum_{i=1}^{N}(y_i - x_i^T\beta)^2 = (Y - X\beta)^T(Y - X\beta)$$

$$\begin{bmatrix} y_1 - x_1^T\beta & y_2 - x_2^T\beta_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 - x_1^T\beta \\ y_2 - x_2^T\beta \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} - \begin{bmatrix} x_1^T \\ x_2^T \end{bmatrix}\beta$$



$$(Y - X\beta)^T (Y - X\beta) = Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X \beta$$

$$\frac{d}{d\beta}$$ p-dim

(1) $$\frac{d}{d\beta}(a^T\beta) = a_{p\times 1}$$

$$\frac{d}{d\beta_1}(a_1\beta_1 + a_2\beta_2) = a_1$$

$$\frac{d}{d\beta_2}(a_1\beta_1 + a_2\beta_2) = a_2$$

② $\frac{d}{d\beta}\left(\beta^T \underbrace{C}_{p\times p} \beta\right) = \underbrace{\left(C^T + C\right)}_{p\times p} \underbrace{\beta}_{p\times 1}$

$a^T x = x^T a$

$$\frac{d}{d\beta}\left(0 - (Y^T \cancel{X})^T - (Y^T \cancel{X})^T + 2X^T X\beta\right) = 0$$

$$\cancel{2}\underline{\underline{(X^T X)\beta}} = \cancel{2}\underbrace{(Y^T X)^T}_{} = X^T Y$$