① $\frac{d}{dx}\left(\frac{1}{1+x}\right) = \frac{-1}{(1+x)^2}$

$\frac{d}{dx}\Big($

$(1+x)^n$

$n(1+x)^{n-1}$

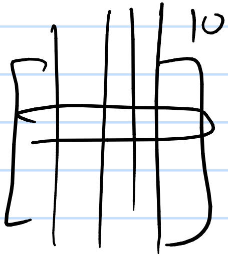
② e^{x}

$\underbrace{\log(y)}$

$\frac{d}{dx} e^x = e^x$

(3) $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$ $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $A = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}$

$\det(A) = ad - bc = 0$

(4) $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \longrightarrow \begin{bmatrix} a & c \\ b & d \end{bmatrix}$ $X^t =$ 

(5) $\left.\begin{aligned} x+y+z&=1 \\ 2x+2y+2z&=2 \end{aligned}\right\}$ $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}$

6. Bernoulli → $X = \begin{cases} 1 & \frac{1}{2} \text{ prob} \\ 0 & \frac{1}{2} \end{cases}$

Binomial

$\Omega = \{ \text{Heads}, \text{Tails} \}$

$X : \Omega \to \{a, b\}$, $a = 1$, $b = 0$

$P(X = 1)$

$= P(\text{get heads})$

$= \frac{1}{2}$

$Y = \#$ of "heads" in "n" trials $\boxed{n = 10}$

$Y \sim Bin(p, n)$

$\Omega = \left\{ \begin{matrix} HH \cdots\cdots H \\ H - \cdots\cdots T \\ \end{matrix} \right\}$

7. $\underset{a_1}{(-3},\ \underset{a_2}{-2},\ \underset{a_3}{0},\ \underset{a_4}{2},\ \underset{a_5}{3)}$ = set

empirical mean = 0
M

Var( )

$$= \underbrace{\frac{1}{5}\sum_{i=1}^{5}(a_i - M)^2}$$ //

2. Inputs and outputs. "variables"

e^{x} f

$\boxed{\text{Supervised learning}}$

predictors
Independent variables
features

response
dependent variables
labels
targets.

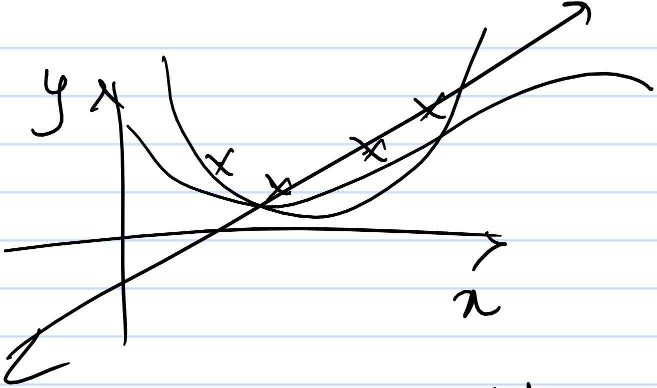
quantitative

regression → numbers

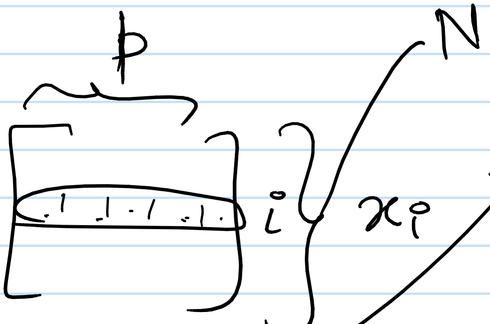
qualitative

classification → qualitative

1. Pattern mining
2. Clustering
3. Ranking
4. Density estimation



X matrix $\longrightarrow$



$$\begin{bmatrix} 3 & 2 & 1 \\ 1 & 5 & 0.1 \\ +3 & \frac{1}{2} & 0 \end{bmatrix}$$

3×3

x_1

X random variable

[a b c]

$$\begin{bmatrix} a & b & c \end{bmatrix} = X^T$$

$X_1 \quad X_2 \quad X_3$

X_j

$$X = \begin{bmatrix} 10 & 4 \\ 2 & -3 \\ 1 & 1.1 \\ 5 & 2 \end{bmatrix}$$

$N = 4$

$p = 2$

$$Y = \begin{bmatrix} 5 \\ 1 \\ 4 \\ 3 \end{bmatrix}$$

$$x_4 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

$N = 4$

$p = 1$

$X \qquad Y$

$$\underbrace{f(X)}_{\hat{Y}} \approx Y$$

$f(10) \qquad Y$

$e^x \qquad e^4$

$\hat{Y} = e^{10}$

e^{x+1}

Two Learners   model   function

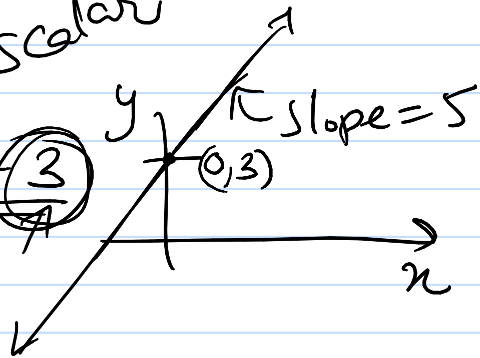
$5x+3$

1. linear regression
2. K-nearest neighbor

linear function

scalar

$5x + 3$ → affine



matrix is a representation for a linear function.

$$f\left(\begin{bmatrix}2\\3\end{bmatrix}\right) \rightarrow \begin{bmatrix}4\\5\end{bmatrix}$$

$$A\begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}4\\5\end{bmatrix}$$

$$\text{data}\left\{\begin{array}{l} \begin{bmatrix}3\\2\end{bmatrix} \xrightarrow{f} \underline{5} \\ \begin{bmatrix}4\\-6\end{bmatrix} \xrightarrow{f} \underline{3} \\ \vdots \\ \begin{bmatrix}5\\7\end{bmatrix} \longrightarrow \underline{-1} \end{array}\right.$$

$$X = \begin{bmatrix}3 & 2\\4 & -6\\5 & 7\end{bmatrix} \quad Y = \begin{bmatrix}5\\3\\-1\end{bmatrix}$$

$$\left[\begin{array}{l} 10f(x) = f(10x) \\ f(x+y) = f(x)+f(x) \end{array}\right]$$

f represented as a vector.

inner product.

$$\begin{bmatrix}\beta_1 & \beta_2\end{bmatrix}\begin{bmatrix}3\\2\end{bmatrix} = 5$$

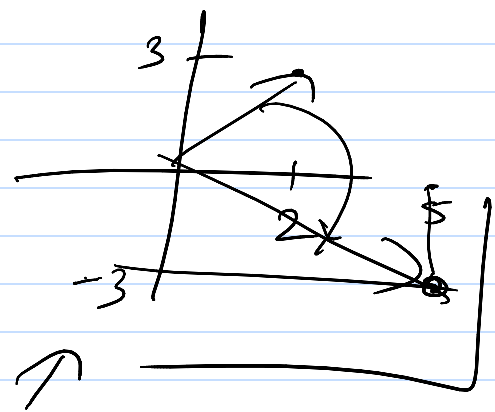
$$\rightarrow f(X) = \beta^T X$$

find a linear reg model $\equiv \hat{\beta} = \begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix}$ $\begin{bmatrix}0.1 \\ 3.3\end{bmatrix}$

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

$$\begin{bmatrix}0.1 & 3.3\end{bmatrix}\begin{bmatrix}2 \\ 3\end{bmatrix} =$$

$$\underset{2\times2}{A}\begin{bmatrix}2 \\ 3\end{bmatrix} = \begin{bmatrix}5 \\ -3\end{bmatrix}$$



$$\begin{bmatrix}2 & 5 \\ 2 & -1\end{bmatrix}$$

$$\beta = A^T \begin{bmatrix}2 \\ 3\end{bmatrix}$$

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

closeness

"goodness"

: Residual Sum of Squares $\leftarrow$ data

X $\quad Y = \begin{bmatrix} y_i \end{bmatrix}$ $N \times 1$

$\begin{bmatrix} x_i^T \end{bmatrix}$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \leftarrow$$

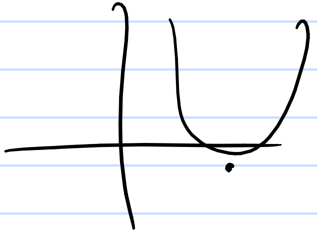
$$RSS(\beta) = \sum_{i=1}^{N} \left(y_i - \beta^T x_i\right)^2 \leftarrow$$

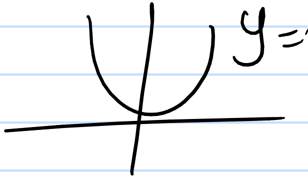
$$\sum_{i=1}^{N} \left(y_i - f(x_i)\right)^2 \leftarrow$$

$$\beta_{first} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

$$\sum_{i=1}^{3} \Big(5 - \underbrace{[5\ 3]\begin{bmatrix} 4 \\ 2 \end{bmatrix}}\Big)^2 + \dots$$

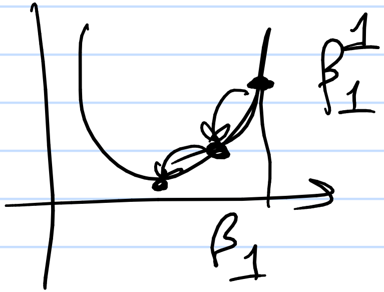
$$\begin{bmatrix} 4 & 2 \\ 3 & 1 \\ 5 & 5 \end{bmatrix} \qquad \begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix}$$





① $\hat{\beta} = \underbrace{\underbrace{(X^T X)^{-1}}_{p \times p} X^T Y}$ $\quad p \times n \quad n \times 1$

② $RSS(\beta)$

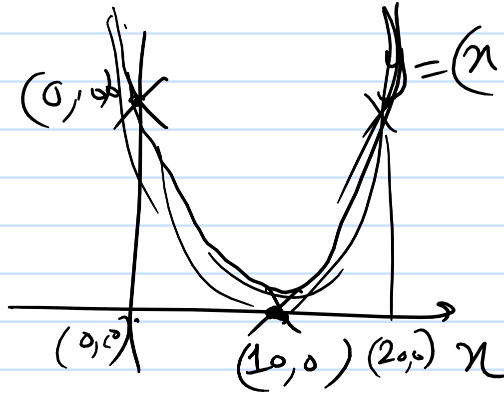


descent

gradient descent

$$\hat{\beta} = \underbrace{\arg\min_{\beta} \underbrace{RSS(\beta)}}$$

$f(x)$

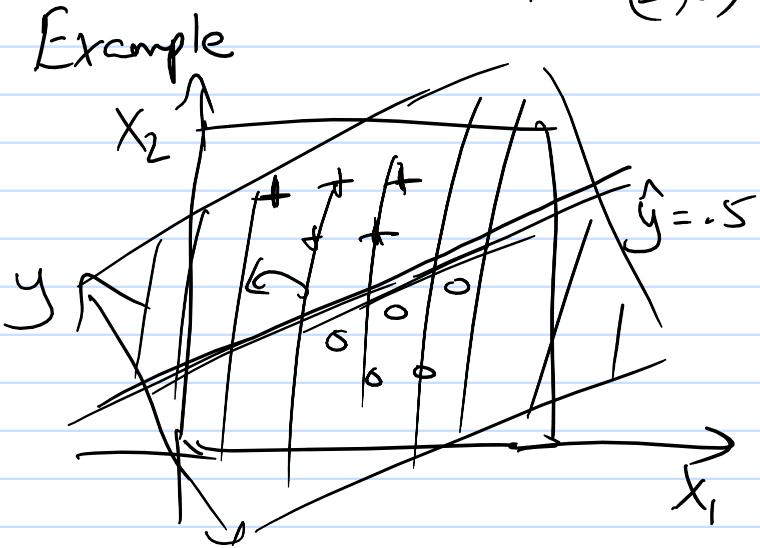


$$y = (x-10)^2 = f(x)$$

$$2(x-10) = 0$$

$$x = 10$$

Example

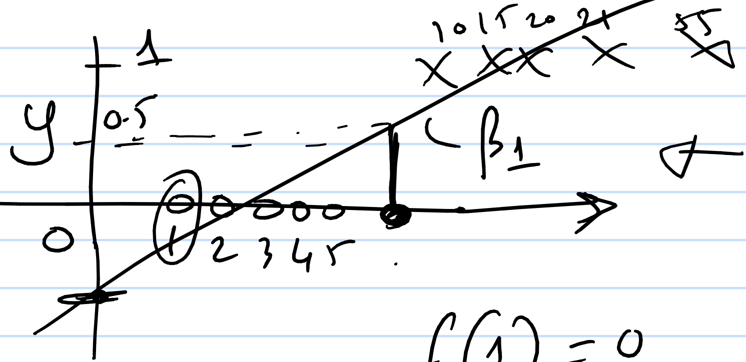


$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \qquad \vec{Y} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$\vec{X}$

$f(x)$

$\beta_1 x + C$



$f(1) = 0$

$f(2) = 0$

$\vdots$

$f(55) = 1$

K-nearest neighbor

$\hat{\beta}$ (K-n-n)

← models →

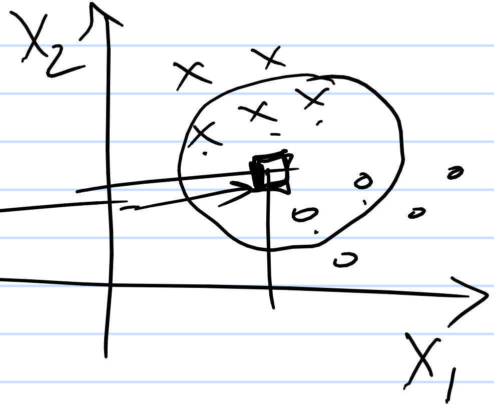
inflexible flexible

(K)-n-n model

: take the average targets of the "nearest" K observation in my data X

output

$$\begin{bmatrix}4\\4\end{bmatrix} \qquad \begin{bmatrix}5\\3\end{bmatrix}$$



$$\sqrt{(4-5)^2+(4-3)^2}$$

$k=5$

$\hat{y}=.5$