LECTURE 7 $\boxed{2/21/18}$

$$g = \mathcal{A}(\mathbb{D}, \mathcal{H})$$

$$\hat{y} = g(\vec{x}^*)$$

What if g located to the closest $\vec{x}_i \in \mathbb{D}$ to $\vec{x}^*$ and returned $\hat{y} = y_i$

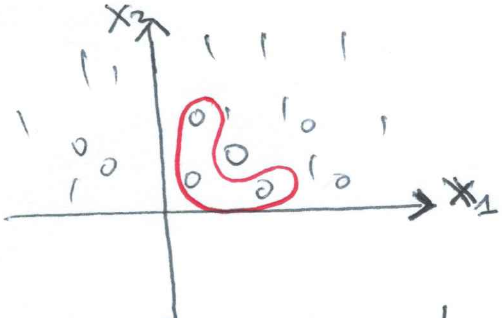
(i.e $\vec{x}^*$'s "nearest neighbour")

We need a notion of "difference" between two observations.

$$d(\vec{x}_i, \vec{x}_k) = \sum_{j=1}^{p} |x_{ij} - x_{kj}|$$ called L1 distance, or Manhattan distance.

take the square both sides:

$$\Rightarrow \sum^{p} (x_{ij} - x_{kj})^2 = \|\vec{x}_i - \vec{x}_k\|^2 : L2 \text{ distance}.$$



1 & 0 represent $\mathbb{D}$

what about g located to the closest K of $\vec{x}_i \in \mathbb{D}$.

$$\hat{y} = \text{mode}[y^*_{(1)}, \ldots, y_{(\ )}]$$

error being zero, you can ~~not~~ trust that.

650

so far $y = \{0, 1\}$: called "binary clasification"

Supervised binary algorithm, this create predict in this case area. If: "y is space possible responses"

$y = \{1, \dots, k\}$ where these values are "nominal" (no order). it is called "clasification."

What if $y = \mathbb{R}$ or $Y = \mathbb{R} \subset \mathbb{R}$?

This is called regression.

First Regression: Linear regression (Model).

Null Model : $g(\vec{x}) = \bar{Y} = \frac{1}{n} \sum y$

simple avg

LINEAR REGRESSION MODEL

Percept: $\mathcal{H} = \{ \mathbb{1}_{\vec{w} \cdot \vec{x} \geq 0} : \vec{w} \in \mathbb{R}$

$$\mathcal{H} = \{ \vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1} \}$$

$$\vec{x} = [1 \cdot x_1, \dots, x_p]$$

$$\vec{w} = [w_0, w_1 \dots, w_p]$$

$$= \{ w_0 + w_1 x + w_2 x + \dots + w_p x_p : w_0 \in \mathbb{R}, w_1 \in \mathbb{R}, \dots, w_p \in \mathbb{R}$$

This is called linear algorithm.

If $p = 1$ $\quad \mathcal{H} = \{ w_0 + w_1 x : w_0 \in \mathbb{R}, w_1 \in \mathbb{R} \}$

$h^*(x) = w_0^* + w_1^* x = \beta_0 + \beta_1 x$

is the best choice $\in \mathcal{H}$ because ~~it~~ is closest to f.

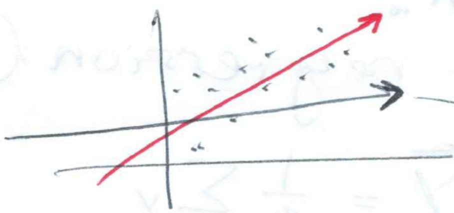
β : means best candidate closest to x.

$$y = h^*(\vec{x}) + \underbrace{\left(t(\vec{z}) - f(\vec{x})\right)}_{\text{ignorance error}} + \underbrace{\left(f(\vec{x}) - h^*(\vec{x})\right)}_{\text{missp error}}$$

called ε

$f(\vec{x})$ is not inside $\mathcal{H}$

$Y = g(\vec{x})$



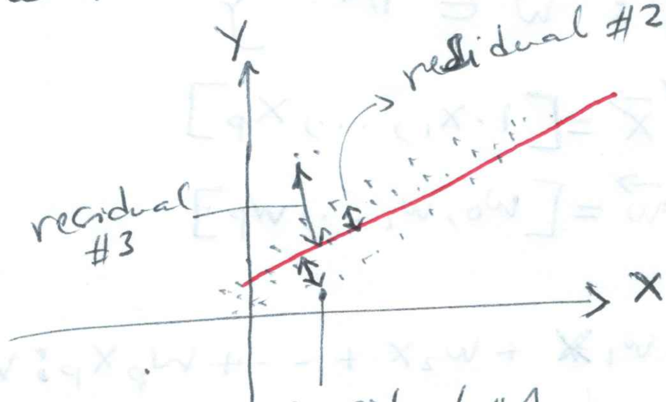
$g(x) = b_0 + b_1 x$

$$Y = g(\vec{x}) + \underbrace{\left(t(\vec{z}) - f(\vec{x})\right) + \left(f(\vec{x}) - h^*(\vec{x})\right) + \underbrace{\left(h^*(\vec{x}) - g(\vec{x})\right)}_{\text{estimate}}}_{e \text{ "residual"}}$$

Only thing we know is

$g = \mathcal{A}(\mathbb{D}, \mathcal{H})$

$\simeq h^*$



loss functions

$SAE := \sum_{i=1}^{n} |e_i|$

summable errors

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \underset{\{w_0, w_1\}}{\operatorname{argmin}} \{SSE\} = \sum e_i^2$$

→ sum square error

↳ called Ordinary Least square

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(Y_i - (w_0 + w_1 X)\right)^2$$

$$= \sum_{i=1}^{n} Y_i^2 + w_0^2 + w_1^2 \cdot \cancel{x}_i^2 - 2Y_i w_0 - 2Y_i w_1 x_i + 2w_0 w_1 x_i$$

$$= \sum y_i^2 + n w_0^2 + w_1^2 + \Sigma y_i^2 - 2w_0 n\bar{y} - 2w_1 \Sigma x_i y_i + 2w_0 w_1 n$$

$$\frac{\partial}{\partial w_0}[SSE] = 2n w_0 - 2n\bar{y} + 2w_1 n\bar{y} = 0$$

$$= 2n(w_0 - 2\bar{y} + w_1 \bar{x}) = 0$$

$$\Rightarrow w_0 - 2\bar{y} + w_1 \bar{x} = 0$$

$$\boxed{w_0 = \bar{y} - w_1 \bar{x}}$$

$$\frac{\partial}{\partial w_1}[SSE] = 2w_1 \Sigma x_i^2 - 2\Sigma x y_i + 2w_0 n\bar{x} = 0$$

$$= 2w_1 \Sigma x_i^2 - 2\Sigma x y_i + 2n\bar{x}(\bar{y} - w_1 \bar{x}) = 0$$

$$= 2w_1 \Sigma x_i^2 - 2\Sigma x y_i + 2n\bar{x}\bar{y} - 2n\bar{x}^2 w_1 = 0$$

$$w_1(2\Sigma x_i^2 - 2n\bar{x}^2) = 2\Sigma x y_i - 2n\bar{x}\bar{y}$$

$$\boxed{w_1 = \frac{\Sigma x y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2}} \Leftarrow \boxed{w_1 = \frac{2\Sigma x y_i - 2n\bar{x}\bar{y}}{2\Sigma x_i^2 - 2n\bar{x}^2}}$$

Two random variable X, Y

Recall:

$$\sigma_{XY} = Cov[X, Y] := E\left[(X - \mu_i)(Y - \mu_i)\right]$$

estimate by $S_{XY} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})$ — 2

$$\frac{\Sigma x_i y_i - \Sigma x_i \bar{y} - \Sigma y_i \bar{x} + n\bar{x}\bar{y}}{n-1} = \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{n-1}$$

c)

$$\rho = cov(x,y) = \frac{Cov(X,Y)}{SE(X)SE(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$S_X = \frac{1}{n-1}\sum(x_i-\bar{x})^2 = \frac{\Sigma x_i^2 - 2\Sigma x_i\bar{x} + \Sigma\bar{x}^2}{n-1} = \frac{\Sigma x_i^2 - n\bar{x}^2}{n-1}$$

$$r = \frac{S_{xy}}{S_x S_y} \Rightarrow S_{xy} = r S_x S_y$$

since $w_0 = \bar{Y} - w_1\bar{x} = \boxed{\bar{Y} - r\frac{S_y}{S_x}\bar{x} = b_0}$

$$w_1 = \frac{S_{xy}}{S_x^2} = \boxed{r = \frac{S_y}{S_x} = b_1}$$

single loss squares estimates.