Lecture 7

2/21/18

$g = \mathcal{A}(\mathbb{D}, \mathcal{H})$ and $\hat{y}^* = g(\vec{x}^*)$

What if g located the closest $\vec{x}_i \in \mathbb{D}$ to $\vec{x}^*$ and returned $\hat{y} = y_i$
(i.e. $\vec{x}^*$'s "nearest neighbor")

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

$$d(\vec{x}_i, \vec{x}_k) = \sum_{j=1}^{p} |x_{i_j} - x_{k_j}| \quad \text{L1 distance / Manhattan distance}$$

$$= \sum_{j=1}^{p} (x_{i_j} - x_{k_j})^2 = \|\vec{x}_j - \vec{x}_k\|^2 \quad \text{L2 distance (Sq'd Euclidian distance)}$$

what if g located the closest k $\vec{x}_i$'s in $\mathbb{D}$ to $\vec{x}^*$ and returned
$\hat{y} = \text{Mode}[y_1, \ldots, y_k] \leftarrow$ responses of the k closest $\vec{x}_i$'s
("k nearest neighbors")

You might want to standardize (scaling)
This sums up nearest neighbors

So far $\mathcal{Y} = \{0, 1\}$
Supervised learning algorithm that create predictions in this case are called "binary classification" algorithms.

If $\mathcal{Y} = \{1, \ldots, k\}$ where those values are "nominal" (no order) it's called "classification"

What if $\mathcal{Y} = \mathbb{R}$ or $\mathcal{Y} = \mathbb{R} \subset \mathbb{R}$? this is called "regression" for historical reasons
($\mathcal{Y}$ = space of possible responses)

Null Modol

doesn't use $\vec{x}$'s

$g(\vec{x}) = \bar{y} := \frac{1}{n} \sum y$ sample avg.

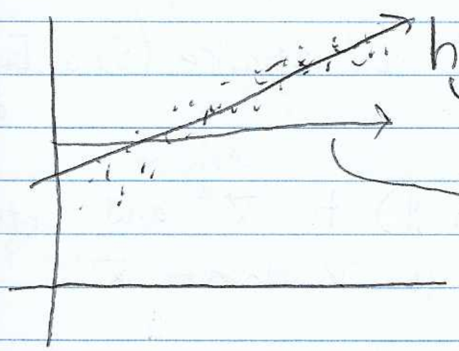
Linear Regression Model

$$\mathcal{H} = \{\vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1}\} = \{w_0 + w_1 x_1 + w_2 x_2 + \dots + w_p x_p : w_0 \in \mathbb{R}, w_1 \in \mathbb{R}, \dots, w_p \in \mathbb{R}\}$$

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

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

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



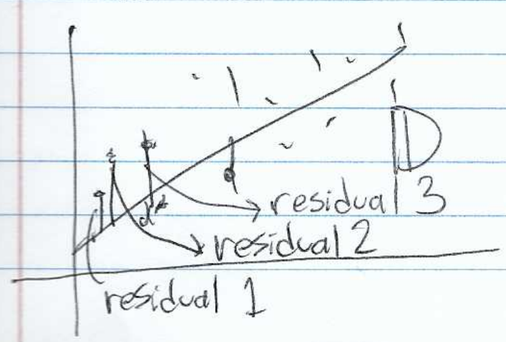
$h^*(x) = w_0^* + w_1^* x = \beta_0 + \beta_1 x$

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

$g(x) = b_0 + b_1 x$

$$\underbrace{y = h^*(\vec{x}) + \underbrace{(f(\vec{z}) - f(\vec{x}))}_{\text{ignorance error}} + \underbrace{(f(\vec{x}) - h^*(\vec{x}))}_{\text{misspecification error}}}_{\varepsilon \to \text{"error" or "noise"}}$$

$$y = g(\vec{x}) + \underbrace{(f(\vec{z}) - f(\vec{x})) + (f(\vec{x}) - h^*(\vec{x})) + \underbrace{(h^*(\vec{x}) - g(\vec{x}))}_{\text{estimation}}}_{e \text{ "residual"}}$$



$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \underset{\{w_0, w_1\}}{\operatorname{argmin}} \{SSE\} = \sum_{i=1} e_i^2$$

sum sqd. error

Ordinary least squares (OLS)

$\bar{y} = \frac{1}{n}\sum y_i \Rightarrow \sum y_i = n\bar{y}$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2$$

$$= \sum_{i=1}^{n} y_i^2 + w_0^2 + w_1^2 x_i^2 - 2y_i w_0 - 2y_i w_1 x_i + 2w_0 w_1 x_i$$

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

$$\frac{\partial}{\partial w_0}[SSE] = 2nw_0 - 2n\bar{y} + 2w_1 n\bar{x} \stackrel{\text{set}}{=} 0$$

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

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

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

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

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

$$\Rightarrow w_1 = \frac{\sum x_i y_i - n\bar{y}\bar{x}}{\sum x_i^2 - n\bar{x}^2} = \frac{(n-1) S_{xy}}{(n-1) S^2_x} = \frac{S_{xy}}{S^2_x} = \boxed{r\frac{S_y}{S_x} = b_1}$$

Recall $\sigma_{xy} = Cov[X,Y] := E[(X-\mu_x)(Y-\mu_y)]$

est. by $S_{xy} := \frac{1}{n-1}\sum (x_i - \bar{x})(y_i - \bar{y}) = \frac{\sum x_i y_i - \sum x_i \bar{y} - \sum y_i \bar{x} + n\bar{x}\bar{y}}{n-1} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{n-1}$

$$\rho := Cov[X,Y] = \frac{Cov[X,Y]}{SE[X]\,SE[Y]} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

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

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