Line is drawn in the min amount of error. * Dropping serial # when running data 02/14/2019.

Perceptron requires linearly separability in d.

$y = \{0, 1\}$ => Binary Classification Model

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

indicator

coerce all $h \in \mathcal{H}$ to have rang y (functions to return 0 and 1s).

Null Model — strongman to beat.

$g_0 = \bar{y}$ — best thing to minimize square expected error.

sample average.

$\mathbb{D} = \langle X, y \rangle$

$\mathbb{D} = y$

Our first non-trivial regression model

(functions) want them to be linear.

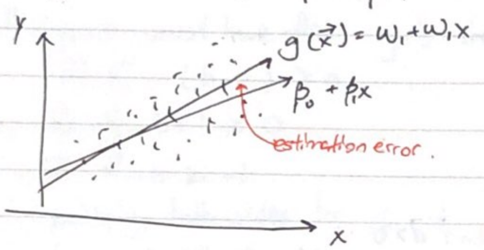
$$\mathcal{H} = \{ \vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1} \}$$
$$= \{ w_0 + w_1 x_1 + \dots + w_p x_p : w_0, w_1, \dots w_p \in \mathbb{R} \}$$

if $p = 1$

$$\mathcal{H} = \{ w_0 + w_1 x = w_0, w_1 \in \mathbb{R} \}$$

$$h^*(\vec{x}) = w_0^* + w^* x_1 + \dots + w_p^* x_p$$
$$= \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad \text{econometrics / Statistics}$$

$$y = h^*(\vec{x}) + \mathcal{E} = \beta_0 + \beta x_1 + \dots + \beta_p x_p + \mathcal{E} \leftarrow \text{misspecification / ignorance}$$



How to "fit" i.e. select an element of $\mathcal{H}$ which is an good model?

Need to specify an objective function that reflects error in the model.

Then select g that minimizes the error.

our metric to measure error. $\rightarrow$ $SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

sum squared error

$p = 1$ We want minimal SSE linear function

$$\hat{w_0} w_1 = \underset{w_0 w_1 \in \mathbb{R}}{\operatorname{argmin}} \{ SSE \} \qquad \hat{y}_i = g(x_i)$$

$$= \operatorname{argmin} \{ \Sigma (y_i - (w_0 + w_1 x_i))^2 \} \qquad \bar{y} = \frac{1}{n} \Sigma y_i$$

$$= \Sigma y_i^2 + w_0^2 + w_1^2 x_i^2 - 2 y_i w_0 - 2 y_i w_1 x_i + 2 w_0 w_1 x_i$$

$$SSE = \Sigma y_i^2 + n w_0^2 + w_1^3 \Sigma x_i^2 - 2 w_0 n \bar{y} - 2 w_1 \Sigma x_i y_i + 2 w_0 w_1 n \bar{x}$$

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

$$\beta_0 = w_0 = \bar{y} - w_1 \bar{x}$$

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

$$w_1 \Sigma x_i^2 = \Sigma x_i y_i - w_0 n \bar{x}$$

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

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

$\Rightarrow \hat{\beta}_1 \Rightarrow \hat{w}_1 = \dfrac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2} \Rightarrow \dfrac{(n-1) S_{xy}}{(n-1) S_x^2} = \dfrac{S_{xy}}{S_x^2} = r\dfrac{S_y}{S_x}$

$e = Corr[X, Y] = \dfrac{Cor[X, Y]}{SE[X] SE[Y]}$ correlation (unitless)

$\sigma_{xy} = Cov[X, Y] = E[(X - \mu_x)(Y - \mu_y)]$ $\boxed{\text{est by } r = \frac{S_{xy}}{S_x S_y} \quad r_o \frac{S_y}{S_x} = \frac{S_{xy}}{S_x^2}}$

Sample covariance $S_{xy} = \frac{1}{n-1}\Sigma(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}(\Sigma x_i y_i - \overbrace{\Sigma \bar{x} y_i}^{-n\bar{x}\bar{y}} - \overbrace{\Sigma x_i \bar{y}}^{-n\bar{x}\bar{y}} + \Sigma \bar{x}\bar{y})$

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

$\sigma_x^2 = Var(x)$ est by $S_x^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$

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

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

note: $\lim_{n \to \infty} b_0 \to \beta_0 \quad \lim_{n \to \infty} b_1 = \beta_1$

$= g(x) \to h^*(x)$ minimize estimation error but the other two error will still be there.

$\Rightarrow g(x) = b_0 + b_1 x$

How well does g predict?