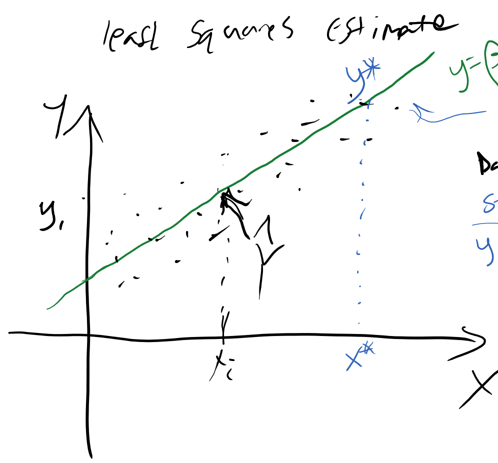
Pr. Qu
Saturday, May 19, 2018 2:08 PM

Linear algebra in statistics & Machine learning

	x	y
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
⋮	⋮	⋮
m	x_m	y_m

least squares estimate



statistical inference, ↖ confidence interval certainty

Data fitting

Statistical Model

$y_i \sim N(\theta_0 + \theta_i x_i, \sigma^2)$

Normal distribution

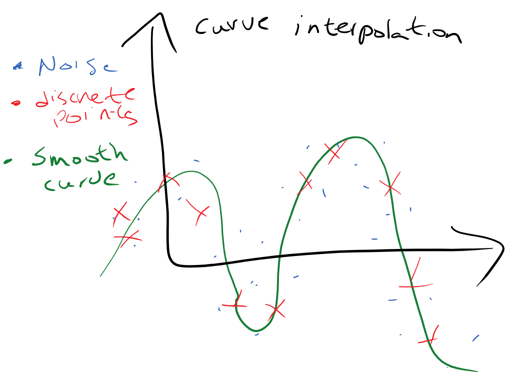
$\gamma_i = y_i - \hat{y}_i$

$= y_i - (\beta_0 + \beta_1 x_i)$

S.d. Quantifies uncertainty

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min \sum_{i=1}^{n} \gamma_i^2$$

$$= \arg\min_{\beta_0, \beta_1} \sum_{i}^{n} [y_i - (\beta_0 + \beta_1 x_i)]$$



* parametric model means this model depends on parameters.

Statistical inference

hypothesis testing.

$H_0: \beta_1 = 0$

$H: \beta_1 > 0$

	x_1	x_2	…	x_p	y_1
1	x_{11}	x_{12}	…	x_{1p}	
2	x_{21}	x_{22}		x_{2p}	
3		⋮			
4		⋮			
⋮					
m	x_{m1}	x_{m2}	…	x_{mp}	

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$

let $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$ $\beta_p = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$

* could be row vector or col vector as long as you are consistent.

RSS, residual sum of squares

let $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix}$ X ← stack all x_i's together

Capital $[x_i]$

residual sum of squares

Let $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ X capital X

$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}$

* could be row vector or col vector as long as you are consistent.

then $y_i = \beta_0 + \beta^T x_i$ ← transpose converts row vector to col →

$$rss = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta^T x_i) \right]^2 = (Y - X^T[\beta_0, \beta])^T [Y - X[\beta_0, \beta]]$$

we want to find β_0 vectors such that we minimize this function.

Linear Model Specific form

$$Y = \beta^T x + \varepsilon$$

given observations $\{y_i, x_i\}$, we can represent the regression model and data as $y = X\beta + \epsilon$, where X is the $n \times m$ matrix whose rows are the x_i's and ϵ is the vector of deviations of the observations from the functional model

$$y_1 = \beta^T x_1 + \varepsilon_1$$
$$y_2 = \beta^T x_2 + \varepsilon_2$$
$$y_m = \beta^T x_m + \epsilon_m$$

let $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$

$$y = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_m^T \end{bmatrix} \beta + \epsilon$$

intercepts

let $X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_M^T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mp} \end{bmatrix}$ regression matrix

ordinary least squares OLS $\hat{\beta} = x$

Solution β

Estimate

Estimator $E|\hat{\beta} - \beta|^2$

Squared OLS $\beta = x$

Solution β

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

Estimate

Estimator $E|\hat{\beta} - \beta|^2$

MSE mean squared error

on Avg. how far is the soln from the truth.

Linear model is linear on parameter β

---

still linear on β

linear logistical regression model

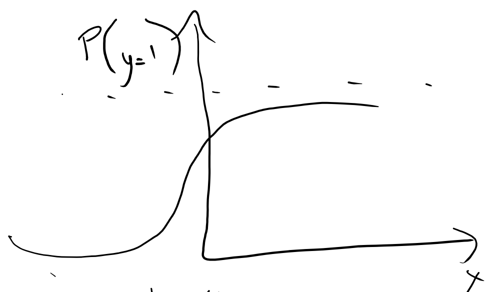
Mechanism for how y_i is generated is not needed. $\rightarrow$ $y_i \sim \text{Bernoulli}(\beta^T x_i)$

predictor

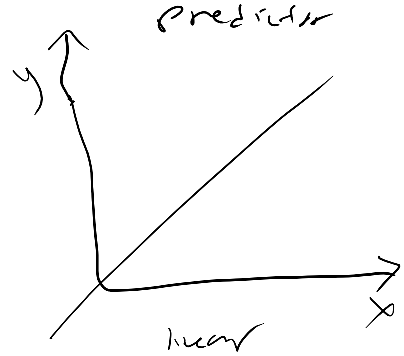
relationship between x and y is non-linear

i.e. non-linear on the predictor

Same model with different data format



Linear logistical



linear

True bernoulli model

$$P(y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 z_1 + \beta_2 z_2)}}$$

predictor
