Agenda

1. Linear model w/ Gaussian/normal error terms
2. OLS as maximum likelihood
3. Inference under normality (of error terms)

Last time: normal eq'ls

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

OLS estimator $\hat{\beta} = (X^T X)^{-1} X^T y$

---

1. Linear model w/ Gaussian / normal error terms

Assume (X, y) satisfies

$$\boxed{y \mid X} \sim N(X\beta_*, \sigma^2 I_n) \text{ for some } \beta_* \in \mathbb{R}^p$$

random vector in $\mathbb{R}^n$

$$X = \begin{pmatrix} - x_1^T - \\ \vdots \\ - x_n^T - \end{pmatrix}$$

x_i = vector of covariates for sample i; each row has p components

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Ex: 1. generate X from some dist. on covariates

2. compute: $\mu \leftarrow X\beta_*$ for some vector $\beta_* \in \mathbb{R}^p$

3. generate y from $N(\mu, \sigma^2 I_n)$

concretely: $y_i \overset{ind}{\sim} N(\mu_i, \sigma^2)$

$\sigma^2 I_n$:

$$\sigma^2 \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{bmatrix}$$

Equivalent statement of Assumption:

1. (linearity) $E[y \mid X] = X\beta_*$

2 (Gaussianity/normality) Define error terms: $\epsilon \leftarrow y - E[y|X] \overset{\text{by linearity}}{=} y - X\beta_*$

$\epsilon | X \sim N(0, \sigma^2 I_n)$

Aside (max. likelihood):

Data: observation of weight of chalk $x \in \mathbb{R}$

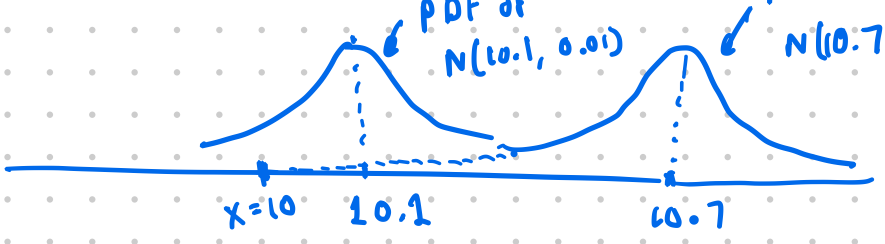
Assume $x \sim N(\underbrace{\mu}_{\text{unknown}}, 0.01)$

pdf of x is $p(x;\mu) = \frac{1}{\sqrt{2\pi}\,0.1} \exp\left(-\frac{(x-\mu)^2}{2 \cdot 0.01}\right)$

likelihood: $L(\mu) \triangleq p(x;\mu)$

maximum likelihood: $\hat{\mu}_{MLE} \leftarrow \arg\max_{\mu} L(\mu)$

Lets assume $x_i = 10.$



OLS as Max Likelihood under normality

likelihood of (X, y)

$L(\beta, \sigma) \triangleq p(\underbrace{y|X}_{\text{random vector } y|X}; \underbrace{\beta, \sigma}_{\text{parameters of dist of } y|X}) = \prod_{i=1}^{n} p(y_i | X; \beta, \sigma)$

$= \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2\right\}}_{\text{PDF of a } N(x_i^T\beta, \sigma^2) \text{ random var}}$

$\sum_{i=1}^{n} s_i = s_1 + \dots + s_n$

$\prod_{i=1}^{n} s_i = s_1 \cdot s_2 \cdot \dots \cdot s_n$

$\hat{\beta}, \hat{\sigma} = \arg\max_{\beta, \sigma} L(\beta, \sigma)$

$$(\hat{\beta}, \hat{\sigma}) \in \arg\max_{\beta,\sigma} \ell(\beta,\sigma)$$

$$= \arg\max_{\beta,\sigma} \prod_{i=1}^n p(y_i | x_i; \beta, \sigma)$$

$$= \arg\max_{\beta,\sigma} \log\left(\prod_{i=1}^n p(y_i | x_i; \beta, \sigma)\right)$$

$$= \arg\max_{\beta,\sigma} \sum_{i=1}^n \log p(y_i | x_i; \beta, \sigma)$$

$$= \arg\max_{\beta,\sigma} \sum_{i=1}^n \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2$$

plug $\hat{\beta}$ in

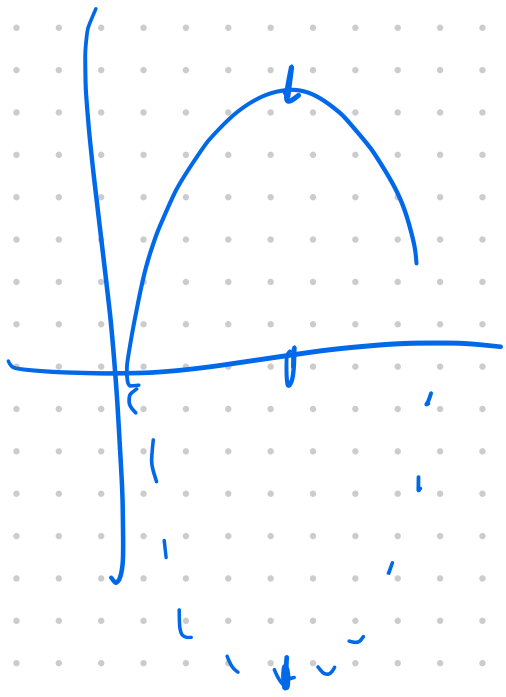
Focus on solving for $\hat{\beta}$

The argmax in β is

$$\arg\max_{\beta} \sum_{i=1}^n -\frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2$$

$$= \arg\max_{\beta} \sum_{i=1}^n -\frac{1}{2}(y_i - x_i^T\beta)^2$$

$$= \arg\min_{\beta} \underbrace{\sum_{i=1}^n \frac{1}{2}(y_i - x_i^T\beta)^2}_{SSR(\beta)}$$



Dist. of $\hat{\beta}$ under normality:

$$\hat{\beta} = (X^TX)^{-1}X^Ty \quad \text{(from normal eq)}$$

$$= (X^TX)^{-1}X^T(X\beta_* + \epsilon) \quad \text{(from linearity)}$$

$$= \cancel{(X^TX)^{-1}}\cancel{X^TX}\beta_* + (X^TX)^{-1}X^T\epsilon$$

$$= \beta_* + (X^TX)^{-1}X^T\epsilon$$

$$\hat{\beta} - \beta_* = (X^TX)^{-1}X^T\epsilon$$

If $z \sim N(\mu, \Sigma)$ then

$Az \sim N(A\mu, A\Sigma A^T)$

$$\sim N\left((X^TX)^{-1}X^T \cdot 0,\ \sigma^2 (X^TX)^{-1}X^T I_n X (X^TX)^{-1}\right)$$

$$= N\left(0,\ \sigma^2 (X^TX)^{-1}\right)$$