Lecture August 27

Linear Regression

input $X = \{x_0\, x_1 \ldots x_{n-1}\}$
output $y = \{y_0, y_1, \ldots y_{n-1}\}$

our data

Model $\tilde{y}(x)$

Basic assumption: There exists a continuous function $f(x)$

$$y(x) = f(x) + \varepsilon$$

stochastic variable but in ML not interested in PDF

statistical noise

$\varepsilon \sim N(0, \sigma^2)$
normal distributed noise

x is deterministic, non-stochastic

$$f(x) \simeq \tilde{y}(x)$$

$$f(x_i) \simeq \tilde{y}_i = \tilde{y}(x_i)$$

$$y_i = y(x_i) \simeq \tilde{y}_i + \varepsilon_i$$

Make a polynomial fit

$$\tilde{y}(x_i) = \sum_{j=0}^{p-1} \beta_j x_i^j$$

$$\boldsymbol{\beta} = \{\beta_0, \beta_1 \, - - - \, \beta_{p-1}\}$$

$\underline{\beta \in \mathbb{R}^p}$ (can have $p = n$)

$\underline{x \in \mathbb{R}^n}$

$\underline{y \in \mathbb{R}^n}$

$1 \qquad 2 \qquad p-1$

$$y_0 = \beta_0 + \beta_1 x_0^1 + \beta_2 x_0^2 + \cdots \beta_{p-1} x_0^{p-1}$$

$$y_1 = \beta_0 + \beta_1 x_1^1 + \beta_2 x_1^2 + \cdots \beta_{p-1} x_1^{p-1}$$

$$\vdots$$

$$y_{n-1} = \beta_0 + \beta_1 \underline{x_{n-1}^1} + \cdots + \beta_{p-1} \underline{x_{n-1}^{p-1}}$$

$$y = X\beta$$

$$X \in \mathbb{R}^{n \times p}$$

X = Feature/Design matrix

$$\begin{bmatrix} 1 & x_0^1 & x_0^2 & \cdots & x_0^{p-1} \\ 1 & x_1^1 & x_1^2 & \cdots & \\ 1 & \vdots & \ddots & & \\ 1 & \vdots & & \ddots & \\ \vdots & \vdots & & & \ddots \end{bmatrix}$$

$$\begin{bmatrix} 1 & & & \tilde{x}_{n-1}^{p-1} \end{bmatrix}$$

(Rows correspond to data. Each column represents a given feature (here power of x corresponding to one parameter β_j)

β is the unknown

Assess the quality of the model:

Loss/cost/Risk/Error ... function

$$C(X,\beta) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \tilde{y}_i)^2$$

$$\left(= \frac{1}{2}\sum_{i=0}^{n-1}(y_i - \tilde{y}_i)^2 \right)$$

$$= \mathbb{E}[(y-\tilde{y})^2]$$

Statistics

$$p(x)\,dx$$

$$\mathbb{E}[x] = \int_D x\,p(x)\,dx = \mu$$

$$\sigma^2 = \mathbb{E}[(x-\mu)^2]$$

$$= \int_D (x-\mu)^2\,p(x)\,dx$$

Discrete Probability

$$\mathbb{E}[x] = \sum_{i=0}^{n-1} x_i \cancel{p(x_i)} = \mu$$

?

$$\sigma^2 = \sum_{i=0}^{n-1} (x_i-\mu)^2 \cancel{p(x_i)}$$

sample mean

$$\mu_x = \mathbb{E}[x] = \frac{1}{n}\sum_{i=0}^{n-1} x_i$$

$$+\mu$$

true

$$C(X,\beta) = \mathbb{E}\left[(y-\tilde{y})^2\right]$$

optimal $\beta = \hat{\beta} = \boxed{\underset{\beta \in \mathbb{R}^p}{\arg\min}\ C(X,\beta)}$

$$C(X,\beta) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i-\tilde{y}_i)^2$$

$$\tilde{y}_i = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0\,p-1} \\ x_{10} & x_{11} & \cdots & x_{1\,p-1} \\ \vdots & & & \\ x_{i0} & x_{i1} & \cdots & x_{i\,p-1} \\ \vdots & & & \\ x_{n-1\,0} & \cdots & & x_{n-1\,p-1} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

$$= X_{i*}\beta$$

$$C(X, \beta) = \frac{1}{n} \sum_{i=0}^{n-1} \left(y_i - X_{i*} \beta\right)^2$$

$$\frac{\partial C(X, \beta)}{\partial \beta} = 0$$

$$C(X, \beta) = \frac{1}{n} \left(y - X\beta\right)^T \left(y - X\beta\right)$$

$$\frac{\partial C(X, \beta)}{\partial \beta} = -\frac{2}{n} X^T \left(y - X\beta\right)$$

$$= 0$$

$$\Rightarrow$$

$$X^T \left(y - X\beta\right) = 0$$

$$= X^T X \beta = X^T y \Rightarrow$$

$$\cancel{\hat{\beta} = \beta^{opt} = (X'X)^{-1}X'y}$$

$$\hat{\beta} \in \mathbb{R}^{p} \qquad X \in \mathbb{R}^{n \times p}$$

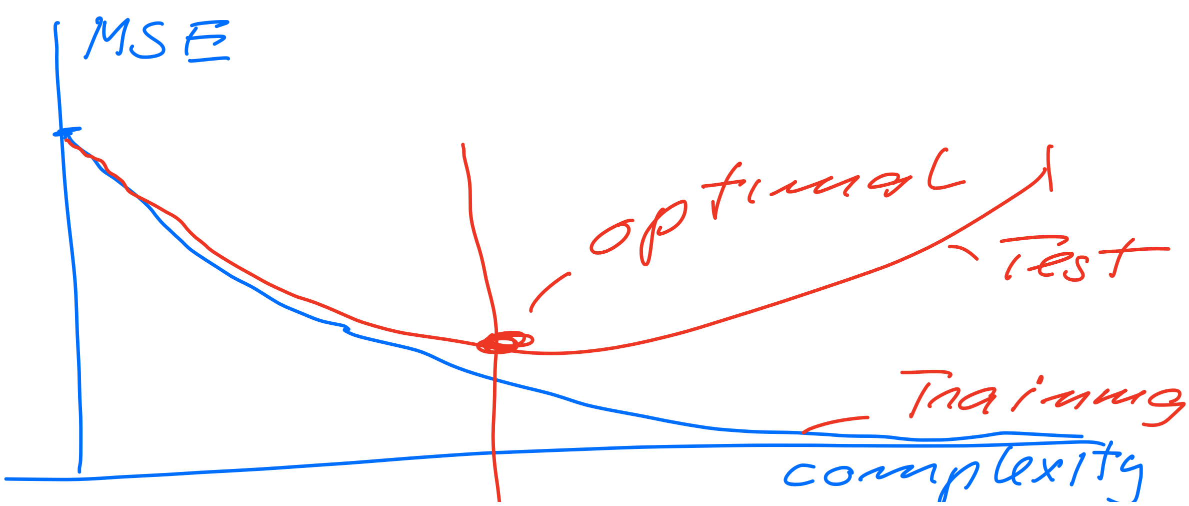
$$X^T X = \mathbb{R}^{p \times p}$$

Need to invert

$$X^T X \; ?$$

$$MSE(X, \beta) = MSE$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} \left(y_i - X_{i*}\beta\right)^2$$



of model

Split data into

Train (70%-80%)

Test (20-30%)