Lecture October 18

Basic elements:

- Data

$$D = \{ (x_0, y_0), (x_1, y_1) \dots (x_{n-1}, y_{n-1}) \}$$

inputs non-stochastic

outputs/ targets

- Model

- Assessing the model $\Rightarrow$ Cost/Loss function (Error, risk ...)

Regression problem (supervised learning)

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

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

μ

$f(x)$ = non-stochastic

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

$$\left(\sum_{i \in D} x_i\, p(x_i) \right)$$

$$\mathrm{var}[x] = \sigma_x^2 = \int_D (x - \mu_x)^2 p(x)\, dx$$

$$\mathrm{cov}(x, y) = \int_D (x - \mu_x)(y - \mu_y) \times p(x, y)\, dx\, dy$$

iid = independent and identically distributed

$$p(x, y) = p(x)\, p(y) \Rightarrow$$

$$\mathrm{cov}(x, y) = 0$$

Sample mean

$$\mathbb{E}[x] = \frac{1}{n} \sum_{i \in D} x_i = \bar{\mu}_x \neq \mu_x$$

$$\mathrm{var}[x] = \bar{\sigma}_2^2 = \frac{1}{n} \sum_{i \in D} (x_i - \bar{\mu}_x)^2$$

$$\mathrm{cov}[x, y] = \frac{1}{n} \sum_{i \in D} (x_i - \bar{\mu}_x)(y_i - \bar{\mu}_y)$$

($\Rightarrow$ leads to resampling techniques : (bootstra techniques: $\begin{cases} \text{bootstrap} \\ \text{cross-validation.} \end{cases}$

Derivation of OLS

Assess the error in our model $\tilde{y} = [\tilde{y}_0 \; \tilde{y}_1 \; \cdots \; \tilde{y}_{n-1}]^T$

$y = [y_0 \; y_1 \; \cdots \; y_{n-1}]^T$

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

$$\equiv \text{cost function}$$

$$= C$$

$f(x) + \varepsilon = y \qquad f(x) \simeq \tilde{y}(x)$

$$\tilde{y}(x_0) = \tilde{y}_0 = \sum_{j=0}^{p-1} \beta_j x_0^j$$

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

$$= \beta_0 x_{00} + \beta_1 x_{01} + \beta_2 x_{02} + \cdots$$

$$\tilde{y}(x_1) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \cdots + \beta_{p-1} x_1^{p-1}$$

$\vdots$

$;$

$$\tilde{y}(x_{n-1}) = \beta_0 + \beta_1 x_{n-1} + \dots + \beta_{p-1} x_{n-1}^{p-1}$$

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

$$\tilde{y} = X\beta$$

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

Design/feature

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^{p-1} \\ 1 & x_1 & x_1^2 & & \vdots \\ 1 & x_2 & x_2^2 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n-1} & x_{n-1}^2 & & x_{n-1}^{p-1} \end{bmatrix}$$

rows = Data inputs

columns = features (here

polynomial
degree)

More general

$$X = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0p-1} \\ \vdots & & & \\ \vdots & & & \\ \vdots & & & \\ x_{n-10} & x_{n-11} & \cdots & x_{n-1p-1} \end{bmatrix}$$

(our inputs)

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

data entries of D

# of features

unknown parameters β

$$C(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j \right)^2$$

$$= \frac{1}{n} \sum_{i=0} (y_i - y_i)$$

$$= \frac{1}{n} (y - \tilde{y})^T (y - \tilde{y})$$

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

$$= \mathbb{E}\left[(y - X\beta)^T (y - X\beta)\right]$$

optimal $\hat{\beta}$ which minimizes $C(\beta)$

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{n} (y - X\beta)^T (y - X\beta)$$

$$\frac{\partial C}{\partial \beta_j} = 0 = -\frac{2}{n} \sum_{i=0}^{n-1} x_{ij} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j\right)$$

$$\Rightarrow \quad = -\frac{2}{n} X^T (y - X\beta)$$

$$\Rightarrow \underbrace{X^T X}_{\in \mathbb{R}^{p \times p}} \hat{\beta} = \underbrace{X^T y}_{\in \mathbb{R}^p} \Rightarrow$$

$\mathbb{R}^1 \quad \mathbb{R}^p \quad \mathbb{R}^1$

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

what kind of problems can arise?

Solve using SVD and pseudoinverse.

In numpy we have pinv

Simply trick to avoid singular $X^T X = H$?

$$\begin{bmatrix} H_{00}+\lambda & x & x & x \\ x & H_{11}+\lambda & x & x \\ x & x & H_{22}+\lambda & \\ x & x & & \ddots \; H_{p-1p-1} \end{bmatrix}$$

$\lambda \sim 10^{-8}$

$X^T X + \lambda I_p \qquad I_p \in \mathbb{R}^{p\times p}$

$\Rightarrow$

$$\hat{\beta} = \left(X^T X + \lambda I_p\right)^{-1} X^T y$$

Ridge Regression.

$$C(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j\right)^2$$

$$+ \lambda \sum_{j=0}^{p-1} \beta_j^2$$

$$\lambda > 0$$

$$\sum_{j=0}^{p-1} \beta_j^2 < \infty$$

Regularization term.

Lasso Regression

$$C(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j\right)^2$$

$$+ \lambda \sum_{j=0}^{p-1} |\beta_j|$$

$$\lambda > 0$$

$$\sum_j |\beta_j| < t$$

$$\frac{d|x|}{dx} = \begin{cases} +1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$