Lecture-8

Linear Regression

$\hat{f}(x) = \beta_0 + \sum_j x_j \beta_j$

- Basis expansion

  $x_1^2, x_2^2, \ldots x_p^2$

  $x_1x_2, x_1x_3, \ldots$

  $\sin(x_1)$

Training data

$(x_1,y_1), (x_2,y_2), \ldots (x_n,y_n)$

Least square fit

$RSS(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2$

X is $\boxed{N \times (p+1)}$: rows are datapoints | 1st col is 1

$\boxed{f(x) = X\beta}$

$\beta : (p+1, 1)$

$\therefore Y : (N, 1)$

$RSS(\beta) = (Y - X\beta)^T (Y - X\beta)$

$\frac{\partial RSS(\beta)}{\partial \beta} = -2X^T(Y - X\beta)$

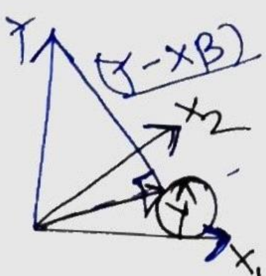
$\frac{\partial RSS(\beta)}{\partial \beta} = -2X^TX \rightarrow$ (Minima)

Minimizing

$X^T(Y - X\beta) = 0 \Rightarrow \hat{\beta} = (X^TX)^{-1}X^Ty$

$Y = \underbrace{X(X^TX)^{-1}X^T}_{\text{'HAT' matrix}} Y = X\hat{\beta}$

Geometric Interpretation of Y :- Assumption is X full Rank



$p = 2$, $n = 3$, $Y \rightarrow$ 3 dimension

$x_1 \rightarrow$ 3 dim

$x_2 \rightarrow$ 3 dim

$= \begin{bmatrix} x_1, x_2 \end{bmatrix}_{3\times 2}$

If some of the Columns are dependent, then X' wouldn't be able to span $(P+1)$ dimensional space, so prediction will be very bad.

If $P >> N$, in Case of Images
↓ Infinite solution (Not a unique) solution.

LECTURE-9 - Multivariate Regression

Multiple Regression :→ More than 1-feature used to predict output.

Univariate regression: ↓ one output variable, one input variable
$(Y = mx + c)$

$Y = X\beta + \epsilon$, w/o intercept (x, y) └ 1 dimension.

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$L = \sum_i (y_i - x_i\beta)^2$$

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow \sum_i 2(0 - x_i)(y_i - x_i\beta)$$

$$\Rightarrow \sum 2(-x_i y_i + x_i^2 \beta) = 0$$

$$\Rightarrow \sum x_i^2 \beta = \sum x_i y_i$$

$$\left(\beta = \frac{\sum x_i y_i}{\sum x_i^2}\right)$$

$Y = (y_1, y_2, \dots y_N)^T$

$X = (x_1, x_2, \dots x_N)^T$

$\langle x, y \rangle = x^T y$

$$\hat{\beta} = \frac{x^T y}{x^T x} = \frac{\langle x, y \rangle}{\langle x, x \rangle}, \quad (r = Y - X\beta)$$

$\boxed{\text{(Regress } Y \text{ on } X)}$

intercept + 1 variable $(\vec{1}, X)$ $\bar{x} = \frac{\sum x_i}{N}$

1. Regress X on $\vec{1}$, from the residual.

residual $Z = X - \bar{x}\vec{1}$
orthogonal x on $\vec{1}$

2. Regress Y on Z to give
$\hat{\beta}_1$

$Z_0 = X_0 = 1$

$J = 1, 2, \cdots p$

$\Rightarrow \boxed{\{\hat{\gamma}\}} = \frac{\langle z_i, z_\cdot \rangle}{\cdots}$

Regress x_j on z_0, z_1, z_2

$$Z_j = x_j - \sum_{k=0}^{J-1} \hat{\gamma}_{kj} Z_k$$

Regress Y on Z_p to give $\hat{\beta}_p = \frac{\langle Z_p, Y \rangle}{\langle Z_p, Z_p \rangle}$

$X = Z\Gamma$

$\Gamma = \begin{bmatrix} 0 & \hat{\gamma}_{kj} \end{bmatrix}$

$Z = [z_1, \cdots z_p]$

$D = \mathrm{diag}(\|z_j\|)$

$X = Z D^{-1} D \Gamma$
$= QR$