

Linear Regression ↔ Classification

- Nearest Neighbor hood Regression
- Kernel Regression → SVM
- Linear Regression by Least Squire Method

$$\text{Minimize } Z = \sum_{i=1}^n (y_i - w x_i)^2$$

Least Squares LR (Linear Regression)

- $y = w_0 + w_1 x$

$$\vec{w} = (w_0 \ w_1), \quad \vec{x} = (1 \ x)$$

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} \vec{x}_1 \\ \vdots \\ \vec{x}_i \\ \vdots \\ \vec{x}_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad W = \vec{w}^T = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

$$Y^T = \vec{w} X^T$$

$$Y = X W$$

- Minimization Error

$$Z = \sum_{i=1}^n (y_i - \vec{w} \cdot \vec{x}_i)^2$$

$$Z = \sum_{i=1}^n (y_i - w_0 - w_1 x_i)^2$$

- Differentiation

$$\frac{\partial Z}{\partial w_0} = \frac{\partial \sum_{i=1}^n (y_i - w_0 - w_1 x_i)^2}{\partial w_0} = \sum_{i=1}^n 2 (y_i - w_0 - w_1 x_i) (-1) = 0$$

$$\frac{\partial Z}{\partial w_1} = \frac{\partial \sum_{i=1}^n (y_i - w_0 - w_1 x_i)^2}{\partial w_1} = \sum_{i=1}^n 2 (y_i - w_0 - w_1 x_i) (-x_i) = 0$$

Derivation

$$\sum_{i=1}^n y_i = \sum_{i=1}^n (w_0 + w_1 x_i)$$

$$\sum_{i=1}^n y_i x_i = \sum_{i=1}^n (w_0 x_i + w_1 x_i^2)$$

$$\begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \cdot \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

$$\begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} \cdot \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \cdot \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \cdot \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \cdot \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

$$X^T Y = X^T X W$$

$$(X^T X)^{-1} (X^T X) W = (X^T X)^{-1} (X^T Y)$$

$$W = (X^T X)^{-1} (X^T Y)$$

$$W = (X^T X)^{-1} (X^T Y)$$

■ Exam

$$\vec{W} = (w_0 \ w_1 \ w_2 \ w_3 \ w_4 \ w_5)$$

$$\vec{X} = (1 \ x \ y \ x^2 \ xy \ y^2)$$

$$Z = w_0 + w_1 x + w_2 y + w_3 x^2 + w_4 xy + w_5 y^2$$

$$Z = X W$$

$$W = (X^T X)^{-1} (X^T Y)$$

■ $1^{q_0} x_1^{q_1} x_2^{q_2} \dots x_m^{q_m} \rightarrow (Q+m) C_Q$

m Input, Q degree, $[q_0, q_1, \dots, q_m] \sum_{i=0}^m q_i = Q$

$${}_{(2+2)} C_2 = \frac{4 * 3}{2 * 1} = 6$$

$$\begin{pmatrix} 1^{q_0} x^{q_1} y^{q_2} & 1 & x & y & x^2 & xy & y^2 \\ q_0 & 2 & 1 & 1 & 0 & 0 & 0 \\ q_1 & 0 & 1 & 0 & 2 & 1 & 0 \\ q_2 & 0 & 0 & 1 & 0 & 1 & 2 \end{pmatrix}$$

Multivariate Regression

$$y = w_1 x_1 + w_2 x_2 + \dots + w_m x_m$$

$$Y = X W$$

$$W = \vec{W}^T = \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}$$

$m \times 1$ matrix

$$\vec{X} = (x_1 \ \dots \ x_m)$$

m vector

$$X = \begin{pmatrix} \vec{X}_1 \\ \vdots \\ \vec{X}_R \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{R1} & \dots & x_{Rm} \end{pmatrix}$$

$R \times m$ matrix, R datapoints

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_R \end{pmatrix}$$

$R \times 1$ matrix

$$\text{Out}[X] = W^T X$$

$$W = (X^T X)^{-1} (X^T Y)$$

$$(X^T X)_{ij} \Rightarrow \sum_{k=1}^R x_{ki} x_{kj}$$

$m \times m$ matrix

$$(X^T Y)_i \Rightarrow \sum_{k=1}^R x_{ki} y_k$$

m - element vector

Bayesian Linear Regression

$$P[y \mid w, x] = \text{Normal}[\text{mean } w x, \text{var } \sigma^2]$$

■ Maximum likelihood estimation of w

$$P[w \mid y_1, y_2, \dots, y_n, x_1, x_2, \dots, x_n]$$

$$P[y_1, y_2, \dots, y_n \mid x_1, x_2, \dots, x_n, w] = \prod_{i=1}^n P[y_i \mid w, x_i]$$

- Maximum likelihood w == Minimizes sum of squares of residuals

$$E = \sum_i (y_i - w x_i)^2 = \sum_i y_i^2 - (2 w x_i y_i) w + \left(\sum_i x_i^2 \right) w^2$$

$$w = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\text{Out}[x] = w x$$

Maximum Likelihood estimation of w

$$\prod_{i=1}^n P[y_i \mid w, x_i] \text{ maximized}$$

$$\prod_{i=1}^n \text{Exp}\left[-\frac{1}{2} \left(\frac{y_i - w x_i}{\sigma}\right)^2\right] \text{ maximized}$$

$$\sum_{i=1}^n -\frac{1}{2} \left(\frac{y_i - w x_i}{\sigma}\right)^2 \text{ maximized}$$

$$\sum_{i=1}^n (y_i - w x_i)^2 \text{ minimized}$$

MLE (Maximum Likelihood Estimator)

- Regression with varying noise

$$y_i \sim N[w x_i, \sigma_i^2]$$

MLE estimate of w

- Minimize the weighted sum of squares

$$\underset{w}{\text{argmin}} \sum_{i=1}^R \frac{(y_i - w x_i)^2}{\sigma_i^2}$$

$$i \text{ datapoint weight} = \frac{1}{\sigma_i^2}$$