ML w2 multivariate liner regression

Multiple Features

Multiple Features

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \cdots + \theta_n x_n$$

$$\uparrow$$

$$x_0 = 1$$

$$X = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

Transpose

$$\underbrace{[\theta_0, \theta_1, \theta_2 \cdots \theta_n]}_{\theta^T} \times \underbrace{\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{X}$$

$(n+1) \times 1$ matrix

$$h_\theta = \theta^T X$$

Feature Scaling

$-1 \leq x_i \leq 1$ range.

- Deviding by maximum value.

  e.g. $\frac{x_2}{2000}$ size of feet < 2000

  $\frac{x_2}{5}$ bedrooms 5.

- make features approximately zero mean.

  e.g. $x_1 = \frac{size - 1000}{2000}$

  $x_2 = \frac{size - 2}{5}$

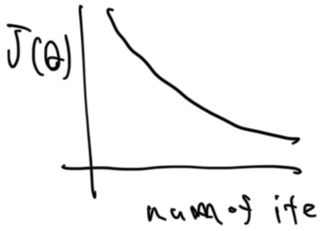
x μ_i ← average value of

$$x_1 \leftarrow \frac{x_1 - \mu_1}{s_1}$$

μ_1 ← average value of training set (x).

s_1 ← range (max − min)

Learning Rate.



To choose α, try

0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1

Feature Scaling with Polynomial Regression

$$h_\theta(x) = \theta_0 + \theta_1(\text{size}) + \theta_2\sqrt{(\text{size})}$$

※ size range : 1 ~ 1000

model.

$$h_\theta(x) = \theta_0 + \theta_1 \underbrace{x_1}_{\frac{\text{size}}{1000}} + \theta_2 \underbrace{x_2}_{\frac{\sqrt{\text{size}}}{32}}$$

← Feature scaling (without mean normalization)

Normal Equation. 正規方程式

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & \cdots \\ 1 & 1416 & 3 & 2 & \cdots \\ 1 & 1534 & 3 & 2 & \cdots \\ 1 & 852 & 2 & 1 & \cdots \end{bmatrix} \qquad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

↑ x_0 ↑ x_1

m examples $(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})$; n : features.

$$x^{(i)} = \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} \in \mathbb{R}^{n+1} \qquad X = \begin{bmatrix} \text{———} \; x^{(1)T} \; \text{———} \\ \text{———} \; x^{(2)T} \; \text{———} \\ \text{———} \; x^{(3)T} \; \text{———} \\ \text{———} \; x^{(m)T} \; \text{———} \end{bmatrix}$$

features (n)

design matrix

examples (m)

E.g.

$$x^{(i)} = \begin{bmatrix} 1 \\ x_1^{(i)} \end{bmatrix} \leftarrow x_0 \qquad X = \begin{bmatrix} 1 & x_1^{(1)} \\ 1 & x_1^{(2)} \\ 1 & x_1^{(3)} \\ \vdots & \vdots \\ 1 & x_1^{(m)} \end{bmatrix} \qquad y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

$y = \theta x$

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

$\theta = \frac{1}{x} y$?

inverse.

octave

pinv (X'*X) * X'*y

O(n³)

Normal equation Noninvertibility.

$X^T X$ non invertible.

- Redundant features. x_1 feet, x_2 : meter $\rightarrow x_1 = (3.28)^2 x_2$
- Too many features (e.g. $m \leq n$) delete features or use regularization

Vectorization

$$h\theta(x) = \sum_{j=0}^{n} \theta_j x_j \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} \cdot \quad x = \begin{bmatrix} x^0 \\ x^1 \\ x^2 \end{bmatrix}$$

prediction = theta' * x;

transpose

$\mathbb{R}^{n+1}$ (delta)

$$\theta := \theta - \alpha \delta$$

$\mathbb{R}^{n+1}$ (vector)

Real number

$\mathbb{R}^{n+1}$

vector delta

$$\text{where } f = \boxed{\frac{1}{m} \sum_{i=1}^{m} (h\theta(x^{(i)}) - y^{(i)}) x^{(i)}}$$

$n = 2$

$$\delta = \begin{bmatrix} \delta_0 \\ \delta_1 \\ \delta_2 \end{bmatrix} \qquad \begin{matrix} (h\theta(x^1) - y^{(1)}) \cdot x^{(1)} \\ (h\theta(x^2) - y^{(2)}) \cdot x^{(2)} \\ \vdots \end{matrix}$$