LR using GD for Single feature

LR using GD for multiple feature

LR using Normal eqⁿ for multiple features.

GD: random θ

loss

updating θ $\rightarrow$ $\theta = \theta - \eta \frac{\partial J(\theta)}{\partial \theta}$ $\quad \nabla_\theta J(\theta)$

NE: $\boxed{\theta = (X^T X)^{-1} X^T y}$?

Derivation of Normal Equation

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 \; \ldots\ldots \; \theta_n x_n$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \qquad \theta^T = [\theta_0 \;\; \theta_1 \;\; \theta_2 \;\; \ldots\ldots \;\; \theta_n] \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$x \rightarrow$ Single example

$$\theta^T x = [\theta_0 \;\; \theta_1 \;\; \ldots\ldots \;\; \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \theta_0 x_0 + \theta_1 x_1 \; \ldots\ldots \; \theta_n x_n$$

$$h_\theta(x) = \theta^T x$$

<u>Error fx^n:</u> $$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^i) - y^i\right)^{2} \times$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(\underline{h_\theta(x^i}) - y^i\right)$$

$$\begin{bmatrix} h_\theta(x^1) \\ h_\theta(x^2) \\ h_\theta(x^3) \\ h_\theta(x^4) \\ \vdots \\ h_\theta(x^m) \end{bmatrix} - \begin{bmatrix} y^1 \\ y^2 \\ y^3 \\ y^4 \\ \vdots \\ y^m \end{bmatrix} = \begin{bmatrix} \theta^T x^1 \\ \theta^T x^2 \\ \theta x^3 \\ \vdots \\ \theta^T x^m \end{bmatrix} - \begin{bmatrix} y^1 \\ y^2 \\ y^3 \\ y^4 \\ \vdots \\ y^m \end{bmatrix}$$

$$\theta^T x^1 = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix} \begin{bmatrix} x_0^1 \\ x_1^1 \\ x_2^1 \\ \vdots \\ x_n^1 \end{bmatrix}$$

$$= \theta_0 x_0^1 + \theta_1 x_1^1 + \theta_2 x_2^1 + \cdots \theta_n x_n^1$$

$$= \begin{bmatrix} \theta_0 x_0^1 + \theta_1 x_1^1 + \theta_2 x_2^1 + \cdots \theta_n x_n^1 \\ \theta_0 x_0^2 + \theta_1 x_1^2 + \theta_2 x_2^2 + \cdots \theta_n x_n^2 \\ \vdots \\ \theta_0 x_0^m + \theta_1 x_1^m + \theta_2 x_2^m \cdots \theta_n x_n^m \end{bmatrix} - \begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^m \end{bmatrix}$$

$m \times 1$ $m \times 1$

$X \cdot \Theta$

$$X: \begin{bmatrix} \text{---} \; x^1 \; \text{---} \\ \text{---} \; x^2 \; \text{---} \\ \vdots \\ \text{---} \; x^m \; \text{---} \end{bmatrix}$$

Same

$$X \cdot \Theta = \underset{m \times (n+1)}{\begin{bmatrix} x_0^1 & x_1^1 & x_2^1 & \cdots & x_n^1 \\ x_0^2 & x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & & & & \\ x_0^m & x_1^m & x_2^m & \cdots & x_n^m \end{bmatrix}} \underset{(n+1) \times 1}{\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}}$$

$$= \underset{m \times 1}{\begin{bmatrix} \theta_0 x_0^1 + \theta_1 x_1^1 + \theta_2 x_2^1 + \cdots \; \theta_n x_n^1 \\ \theta_0 x_0^2 + \theta_1 x_1^2 + \theta_2 x_2^2 + \cdots \; \theta_n x_n^2 \\ \vdots \\ \theta_0 x_0^m + \theta_1 x_1^m + \cdots + \theta_n x_n^m \end{bmatrix}}$$

$X\Theta - y$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\Big(\underbrace{h_\theta(x^i) - y^i}_{X\theta - y}\Big)$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^i) - y^i\right)^2$$

$\underline{h_\theta(x^i)} - y^{(i)}$

$$\begin{bmatrix} - \\ - \\ - \\ - \end{bmatrix} - \begin{bmatrix} - \\ - \\ - \\ - \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

$$(X\theta - y)^2 = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}^2$$

Square your matrix

$$\begin{pmatrix} a^2 \\ b^2 \\ c^2 \\ d^2 \end{pmatrix}$$

Square each element of your matrix

$$\underbrace{\begin{bmatrix} a & b & c & d \end{bmatrix}}_{X\theta - y} \left.\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}\right\} X\theta - y$$

$a^2 + b^2 + c^2 + d^2$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\underline{\left(h_\theta(x^i) - y^i\right)^2}$$

$$J(\theta) = \frac{1}{m}\underline{(X\theta - y)^T (X\theta - y)}$$

$$(X\theta - y)^T (X\theta - y)$$

$$((X\theta)^T - y^T)(X\theta - y)$$

$$(X\theta)^T X\theta \; - \underline{(X\theta)^T y} \; - \underline{y^T (X\theta)} \; + y^T \cdot y$$

$$\downarrow$$

$$X \cdot \theta$$

$$\underset{m\times(n+1)}{\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}} \quad \underset{(n+1)\times 1}{\begin{bmatrix} \theta_0 \\ \vdots \\ \theta_n \end{bmatrix}}$$

$$a^T \cdot b = b^T a$$

$$X\theta = \begin{bmatrix} \; \end{bmatrix}_{m\times 1} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} m\times 1$$

$$(X\theta)^T \cdot y = (1\times m)(m\times 1)$$
$$= 1\times 1$$

$[\;]$ ↳ single value

$$y^T \cdot X\theta = (1\times m)(m\times 1)$$
$$= 1\times 1 = [\;]$$

↳ single value

$$J(\theta) = (X\theta)^T X\theta \; - \underline{(X\theta)^T y} \; - \underline{y^T (X\theta)} \; + y^T \cdot y$$

$$J(\theta) = \underbrace{(X\theta)^T X\theta}_{Q} - \underbrace{2(X\theta)^T y}_{P} + y^T \cdot y$$

(cost / error / loss)

$$P = 2(X\theta)^T y$$

$$X = \begin{bmatrix} x_0^1 & x_1^1 & x_2^1 & \cdots & x_n^1 \\ x_0^2 & x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots \\ x_0^m & x_1^m & x_2^m & \cdots & x_n^m \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$X\theta = \begin{bmatrix} \theta_0 x_0^1 + \theta_1 x_1^1 + \theta_2 x_2^1 + \cdots + \theta_n x_n^1 \\ \theta_0 x_0^2 + \theta_1 x_1^2 + \theta_2 x_2^2 + \cdots \theta_n x_n^2 \\ \vdots \\ \theta_0 x_0^m + \theta_1 x_1^m + \theta_2 x_2^m + \cdots \theta_n x_n^m \end{bmatrix}_{m \times 1}$$

$$(X\theta)^T = \begin{bmatrix} \_\_\_, & \_\_\_, & \cdots & , \_\_\_ \end{bmatrix}_{1 \times m}$$

$$(X\theta)^T \cdot y = \begin{bmatrix} \_\_\_, & \_\_\_, & \cdots & , \_\_\_ \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\begin{aligned}(X\theta^T) \cdot y = {} & (\theta_0 x_0^1 + \theta_1 x_1^1 + \theta_2 x_2^1 + \cdots + \theta_n x_n^1)\, y_1 \\ & + \\ & (\theta_0 x_0^2 + \theta_1 x_1^2 + \theta_2 x_2^2 + \cdots \theta_n x_n^2)\, y_2 \\ & + \\ & \vdots \\ & (\theta_0 x_0^m + \theta_1 x_1^m + \theta_2 x_2^m + \cdots \theta_n x_n^m)\, y_m \end{aligned}$$

$$\underbrace{2(X\theta^T)\cdot y}_{P} = 2(\theta_0 x_0^1 + \theta_1 x_1^1 + \theta_2 x_2^1 + \cdots + \theta_n x_n^1)\, y_1$$

$$+$$

$$2(\theta_0 x_0^2 + \theta_1 x_1^2 + \theta_2 x_2^2 + \cdots \theta_n x_n^2)\, y_2$$

$$+$$

$$\vdots$$

$$2(\theta_0 x_0^m + \theta_1 x_1^m + \theta_2 x_2^m + \cdots \theta_n x_n^m)\, y_m$$

<u>differentiate wrt θ</u>

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$\frac{\partial P}{\partial \theta_0} = 2x_0^1 y_1 + 2x_0^2 y_2 + \cdots\cdots 2x_0^m y_m$$

$$\frac{\partial P}{\partial \theta_1} = 2x_1^1 y_1 + 2x_1^2 y_2 + \cdots\cdots + 2x_1^m y_m$$

$$\vdots$$

$$\frac{\partial P}{\partial \theta_n} = 2x_n^1 y_1 + 2x_n^2 y_2 + \cdots\cdots + 2x_n^m y_m$$

$$\frac{\partial P}{\partial \theta} = \begin{bmatrix} \frac{\partial P}{\partial \theta_0} \\ \frac{\partial P}{\partial \theta_1} \\ \vdots \\ \frac{\partial P}{\partial \theta_n} \end{bmatrix} = 2 \begin{bmatrix} x_0^1 y_1 + x_0^2 y_2 + \cdots\cdots x_0^m y_m \\ x_1^1 y_1 + x_1^2 y_2 + \cdots\cdots + x_1^m y_m \\ x_n^1 y_1 + x_n^2 y_2 + \cdots\cdots + x_n^m y_m \end{bmatrix}$$

$X^T y$

$$X = \begin{bmatrix} x_0^1 & x_1^1 & x_2^1 & \cdots & x_n^1 \\ x_0^2 & x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & & & & \\ x_0^m & x_1^m & x_2^m & \cdots & x_n^m \end{bmatrix} \quad m \times (n+1)$$

$$X^T = \begin{bmatrix} x_0^1 & x_0^2 & \cdots & x_0^m \\ x_1^1 & x_1^2 & & x_1^m \\ x_2^1 & x_2^2 & & x_2^m \\ \vdots & \vdots & & \vdots \\ x_n^1 & x_n^2 & & x_n^m \end{bmatrix} \quad (n+1) \times m$$

$$X^T \cdot y = \underbrace{\begin{bmatrix} x_0^1 & x_0^2 & \cdots & x_0^m \\ x_1^1 & x_1^2 & & x_1^m \\ x_2^1 & x_2^2 & & x_2^m \\ \vdots & \vdots & & \vdots \\ x_n^1 & x_n^2 & & x_n^m \end{bmatrix}}_{(n+1) \times m} \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_{(m \times 1)}$$

$$X^T \cdot y = \begin{bmatrix} x_0^1 y_1 + x_0^2 y_2 + \cdots x_0^m y_m \\ x_1^1 y_1 + x_1^2 y_2 + \cdots + x_1^m y_m \\ x_n^1 y_1 + x_n^2 y_2 + \cdots + x_n^m y_m \end{bmatrix} \quad (n+1) \times 1$$

$$\frac{\partial P}{\partial \theta} = 2 \cdot X^T \cdot y$$

$$Q = (X\theta)^T \times \theta$$

$$(AB)^T = B^T \cdot A^T$$

$$Q = \theta^T X^T X \theta$$

$$X^T \cdot X = \underbrace{\begin{bmatrix} x_0^1 & x_0^2 & \cdots & x_0^m \\ x_1^1 & x_1^2 & \cdots & x_1^m \\ \vdots & & & \\ x_n^1 & x_n^2 & \cdots & x_n^m \end{bmatrix}}_{(n+1)\times m} \cdot \underbrace{\begin{bmatrix} x_0^1 & x_1^1 & \cdots & x_n^1 \\ x_0^2 & x_1^2 & \cdots & x_n^2 \\ \vdots & & & \\ x_0^m & x_1^m & \cdots & x_n^m \end{bmatrix}}_{m\times(n+1)}$$

$$= \begin{array}{c|c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ n \end{array} & \begin{bmatrix} x_0^1 x_0^1 + x_0^2 x_0^2 + \cdots x_0^m x_0^m, & \overbrace{x_0^1 x_1^1 + x_0^2 x_1^2 + \cdots x_0^m x_1^m}^{X_{01}}, \cdots \cdots x_0^1 x_n^1 + x_0^2 x_n^2 + \cdots x_0^m x_n^m \\ \underbrace{x_1^1 x_0^1 + x_1^2 x_0^2 + \cdots x_1^m x_0^m}_{X_{10}}, & x_1^1 x_1^1 + x_1^2 x_1^2 + \cdots x_1^m x_1^m, \cdots \cdots x_1^1 x_n^1 + x_1^2 x_n^2 + \cdots x_1^m x_n^m \\ \vdots & \end{bmatrix} \end{array}$$

$(n+1)\times(n+1)$

Symmetric Matrix

$$r,c = x_r^1 x_c^1 + x_r^2 x_c^2 + x_r^3 x_c^3 + \cdots\cdots x_r^m x_c^m$$

$$r,c = \sum_{i=1}^{m} x_r^i x_c^i = X_{rc}^2$$

$$X^T X = \begin{bmatrix} X_{00}^2 & X_{01}^2 & \cdots\cdots & X_{0n}^2 \\ X_{10}^2 & X_{11}^2 & \cdots\cdots & X_{1n}^2 \\ \vdots & & & \end{bmatrix}$$

$$
\begin{bmatrix}
\vdots & & & \\
x_{n0}^2 & x_{n1}^2 & \cdots & x_{nn}^2
\end{bmatrix}_{(n+1)\times(n+1)}
$$

$$
X^T \cdot X \cdot \theta = \begin{bmatrix}
x_{00}^2 & x_{01}^2 & \cdots & x_{0n}^2 \\
x_{10}^2 & x_{11}^2 & \cdots & x_{1n}^2 \\
\vdots & & & \\
x_{n0}^2 & x_{n1}^2 & \cdots & x_{nn}^2
\end{bmatrix}_{(n+1)\times(n+1)}
\begin{bmatrix}
\theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n
\end{bmatrix}_{(n+1)\times 1}
$$

$$
X^T \cdot X \cdot \theta = \begin{bmatrix}
x_{00}^2\theta_0 + x_{01}^2\theta_1 + \cdots \; x_{0n}^2\theta_n \\
x_{10}^2\theta_0 + x_{11}^2\theta_1 + \cdots \; x_{1n}^2\theta_n \\
\vdots \\
x_{n0}^2\theta_0 + x_{n1}^2\theta_1 + \cdots \; x_{nn}^2\theta_n
\end{bmatrix}_{(n+1)\times 1}
$$

$$
\theta^T X^T \cdot X \cdot \theta = \underset{1\times(n+1)}{\begin{bmatrix}\theta_0 & \theta_1 & \cdots & \theta_n\end{bmatrix}}
\begin{bmatrix}
x_{00}^2\theta_0 + x_{01}^2\theta_1 + \cdots \; x_{0n}^2\theta_n \\
x_{10}^2\theta_0 + x_{11}^2\theta_1 + \cdots \; x_{1n}^2\theta_n \\
\vdots \\
x_{n0}^2\theta_0 + x_{n1}^2\theta_1 + \cdots \; x_{nn}^2\theta_n
\end{bmatrix}_{(n+1)\times 1}
$$

$$
\underline{\underline{\theta^T X^T X \theta}} = \theta_0\left(x_{00}^2\theta_0 + x_{01}^2\theta_1 + \cdots \; x_{0n}^2\theta_n\right)
$$

Q

$$
+
$$

$$
\theta_1\left(x_{10}^2\theta_0 + x_{11}^2\theta_1 + \cdots \; x_{1n}^2\theta_n\right)
$$

$$
\vdots
$$

$$
\theta_n\left(x_{n0}^2\theta_0 + x_{n1}^2\theta_1 + \cdots \; x_{nn}^2\theta_n\right)
$$

differentiate w.r.t $\theta \longrightarrow \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$

$$\frac{\partial Q}{\partial \theta_0} = 2x_{00}^2\theta_0 + \underline{x_{01}^2}\theta_1 + \cdots \quad \underline{x_{0n}^2}\theta_n$$
$$+ \underline{x_{10}^2}\theta_1 + x_{20}^2\theta_2 + \cdots \quad \underline{x_{n0}^2}\theta_n$$

$$\frac{\partial Q}{\partial \theta_0} = 2x_{00}^2\theta_0 + 2x_{01}^2\theta_1 + \cdots \quad 2x_{0n}^2\theta_n$$

$$\frac{\partial Q}{\partial \theta_1} = \underline{x_{10}^2}\theta_0 + 2x_{11}^2\theta_1 + \cdots \quad \underline{x_{1n}^2}\theta_n$$
$$+ \underline{x_{01}^2}\theta_0 + \cdots \quad \underline{x_{n1}^2}\theta_n$$

$$\frac{\partial Q}{\partial \theta_1} = 2x_{01}^2\theta_0 + 2x_{11}^2\theta_1 + \cdots \quad 2x_{1n}^2\theta_n$$

$\vdots$

$$\frac{\partial Q}{\partial \theta_n} = 2x_{0n}^2\theta_0 + 2x_{m1}^2\theta_1 + \cdots \quad 2x_{nn}^2\theta_n$$

$$\frac{\partial Q}{\partial \theta} = \begin{bmatrix} \frac{\partial Q}{\partial \theta_0} \\ \frac{\partial Q}{\partial \theta_1} \\ \vdots \\ \frac{\partial Q}{\partial \theta_n} \end{bmatrix} = \begin{bmatrix} 2x_{00}^2\theta_0 + 2x_{01}^2\theta_1 + \cdots \quad 2x_{0n}^2\theta_n \\ 2x_{10}^2\theta_0 + 2x_{11}^2\theta_1 + \cdots \quad 2x_{1n}^2\theta_n \\ \\ 2x_{n0}^2\theta_0 + 2x_{m1}^2\theta_1 + \cdots \quad 2x_{nn}^2\theta_n \end{bmatrix}$$

$$= 2 \begin{bmatrix} x_{00}^2\theta_0 + x_{01}^2\theta_1 + \cdots & x_{0n}^2\theta_n \\ x_{10}^2\theta_0 + x_{11}^2\theta_1 + \cdots & x_{1n}^2\theta_n \\ \vdots & \\ x_{n0}^2\theta_0 + x_{n1}^2\theta_1 + \cdots & x_{nn}^2\theta_n \end{bmatrix}$$

$$= 2 \underbrace{\begin{bmatrix} x_{00}^2 & x_{01}^2 & \cdots & x_{0n}^2 \\ x_{10}^2 & x_{11}^2 & \cdots & x_{1n}^2 \\ \vdots & & & \\ x_{n0}^2 & x_{n1} & \cdots & x_{nn} \end{bmatrix}}_{2X^T \cdot X} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$\frac{\partial Q}{\partial \theta} = X^T \cdot X \cdot \theta$$

$$J(\theta) = \underbrace{(X\theta)^T X\theta}_{Q} - \underbrace{2(X\theta)^T y}_{P} + y^T \cdot y$$

(cost/error/loss)

$$\frac{\partial J}{\partial \theta} = \frac{\partial Q}{\partial \theta} - \frac{\partial P}{\partial \theta}$$

$$\frac{\partial J}{\partial \theta} = 2X^T \cdot X \cdot \theta - 2X^T y$$

minimize

$$2X^T X \cdot \theta - 2X^T y = 0$$

$$X^T \cdot X \cdot \theta = X^T y$$

multiply both sides by $(X^T X)^{-1}$

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