CLASS EXERCISES

$f(x,y,z) = x^3 + yz$

1.

$$\nabla f(x,y,z) = [3x^2, z, y]$$

$v = [v_1, v_2 \dots v_d]$ $\quad f: \mathbb{R}^d \to \mathbb{R}$

$\theta = [\theta_1, \theta_2 \dots \theta_d]$

$f(\theta) = \langle v, \theta \rangle$ dot product

a) Show $\nabla f(\theta) = v$?

$$f(\theta) = v_1 \theta_1 + v_2 \theta_2 \dots + v_d \theta_d$$

$$= \sum_{i=1}^{d} v_i \theta_i$$

$$\nabla f(\theta) = \left[\frac{\partial f(\theta)}{\partial \theta_1}, \frac{\partial f(\theta)}{\partial \theta_2} \dots \frac{\partial f(\theta)}{\partial \theta_d} \right]$$

$= [v_1, v_2 \dots, v_d] = V$ //

b) $f: \mathbb{R}^d \to \mathbb{R}$

$$f(\theta) = \underset{1\times d}{\theta^T} \underset{d\times d}{M} \underset{d\times 1}{\theta} \underset{1\times 1}{\in \mathbb{R}} \quad \theta =$$

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix}$$

$\theta_{1\times D}$ $\theta^T_{D\times 1}$

$\theta^T_{D\times 1}$ $\theta^T_{1\times D}$

$D\times 1$ 1×1 $\cancel{D\times 1}$ $1\times D$

$$M = \begin{bmatrix} M_{11} & \dots & M_{1d} \\ M_{21} & & \\ \vdots & & \\ M_{d1} & \dots & M_{dd} \end{bmatrix}$$

$A \cdot B \neq B \cdot A$

$A_{m\times n} \cdot B_{n\times r} = C_{m\times r}$

$$f(\theta) = [\theta_1, \theta_2 \dots \theta_d] \cdot \underset{d\times d}{\begin{bmatrix} M_{11} & \dots & M_{1d} \\ \vdots & & \vdots \\ M_{d1} & \dots & M_{dd} \end{bmatrix}}$$

$$\underbrace{\cdot \begin{bmatrix} \theta_1 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}}_{} \quad d\times 1 \to \underline{d\times 1}$$

$1\times d \cdot d\times 1 \to 1\times 1$

$$= [\tilde{\Theta}_1 \ldots \tilde{\Theta}_d] \cdot \begin{bmatrix} \sum_{i=1}^{d} M_{1i} \tilde{\Theta}_i \\ \vdots \\ \sum_{i=1}^{d} M_{di} \tilde{\Theta}_i \end{bmatrix}$$

$$= \sum_{i=1}^{d} \tilde{\Theta}_j \sum_{j=1}^{d} M_{ji} \cdot \tilde{\Theta}_i =$$

$$f(\tilde{\Theta}) = \sum_{i=1}^{d} \sum_{j=1}^{d} M_{ij} \cdot \tilde{\Theta}_i \cdot \tilde{\Theta}_j$$

$$(\nabla f(\tilde{\Theta}))_k = \frac{\partial f(\tilde{\Theta})}{\partial \tilde{\Theta}_k} =$$

$$= \frac{\partial}{\partial \tilde{\Theta}_k} \left(M_{kk} \tilde{\Theta}_k^2 \right) + \sum_{\substack{j=k \\ j \neq k}}^{d} M_{kj} \cdot \tilde{\Theta}_j$$

$$+ \sum_{\substack{i \neq k \\ j = k}}^{d} M_{ik} \tilde{\Theta}_i = \ldots$$

(
1. derivative of a square function
2. two sums i = 1..d
)

$$= \sum_{i=1}^{d} (M_{ik} + M_{ki}) \tilde{\Theta}_i$$

$$M = \begin{bmatrix} M_{11} & \cdots & M_{1d} \\ \vdots & & \vdots \\ M_{d1} & & M_{dd} \end{bmatrix} \quad M^T = \begin{bmatrix} M_{11} & \cdots & M_{d1} \\ \vdots & & \vdots \\ M_{1d} & \cdots & M_{dd} \end{bmatrix}$$

$$(M + M^T)_{ik} = (M + M^T)_{ki}$$

$$(M^T)^T = M$$

$$((M + M^T)\Theta)_k$$

k-th value of $\nabla f(\Theta)$

$\Downarrow$

$$\underline{(M + M^T)\Theta = \nabla f(\Theta)}$$

Properties to remember

M^T transpose of a matrix

$\nabla f(\Theta)$ rate of change

matrix multiplication | $A \cdot B \neq B \cdot A$

$m \times n \cdot n \times r \to m \times r$

Linear regression

1. $\mathcal{L}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(f_\theta(x_i) - y_i\right)^2$

$\boxed{f} : \boxed{x} \rightarrow \boxed{y}$

$f_\theta(x) = \theta_1 x + \theta_2$

a) write MSE in matrix form

$$\frac{1}{n}\|X\theta - y\|_2^2$$ square

L2 NORM

$V = [x, y, z]$ L2 NORM

$$\|V\|_2 = \sqrt{x^2 + y^2 + z^2}$$

$$\frac{1}{n}\|X\theta - y\|_2^2$$

$1 = [x_1, x_2, x_3 \ldots x_d]$

$\vdots$

$n = \ldots\ldots$

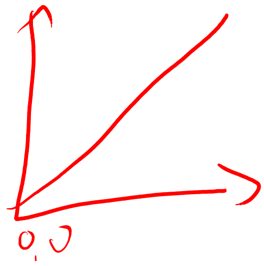
$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$(X\Theta - y)_i = \Theta_1 x_i^1 \qquad x_{n \times d}$$

$$+ \Theta_2 x_i^2 + \dots \Theta_d x_i^d - y_i =$$

$$= f_\Theta(x_i) - y_i$$

$$// \; f_\Theta(x) = \Theta_1 x$$



$$= \frac{1}{n} \sum_{i=1}^{n} \left(f_\Theta(x_i) - y_i \right)^2$$

2. $\nabla L(\theta) = ?$ $\quad 1\times d \cdot d\times 1$

$L(\theta) = \|X\theta - y\|_2^2 = (X\theta - y)^T (X\theta - y)$

$$= \Big(\underbrace{(X\theta)^T}_{= \theta^T X^T} - y^T\Big)(X\theta - y)$$

$$= \theta^T X^T X \theta - \theta^T X^T y - y^T X \theta + y^T y$$

$$\nabla L(\theta) = -X^T y + \left(X^T X + (X^T X)^T\right)\theta$$

$$- X^T y$$

$$= -2X^T y + 2X^T X \theta =$$

$$= -2\left(X^T y + X^T X \theta\right)$$

- symmetric matrix

$$M = M^T$$

$$\Rightarrow \nabla f(\theta) = 2M\theta$$

$$\nabla L(\tilde{\theta}^*) = 0$$

$$/ (X^T X) \cdot (X^T X)^{-1}$$

$$= I$$

$$-2X^T y + 2X^T X \tilde{\theta}^* = 0$$

$$X^T X \tilde{\theta}^* = X^T y \quad / \cdot (X^T X)^{-1}$$

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