Page No.:
Date:

Week = 1

$$A w = b \qquad w = A^{-1} b$$

3.1 $$A^{-1} = \frac{\text{adj}(A)}{|A|} \qquad A^{-1} = \frac{1}{11}\begin{bmatrix} 3 & -1 \\ -1 & 4 \end{bmatrix}$$

$$w = A^{-1} b = \frac{1}{11}\begin{bmatrix} 3 & -1 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{11}\begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

Verification
$$A w = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} \frac{1}{11} \\ 7/_{11} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = b$$

3.2 Part A $f(x) = x^2 + 3x + 1 \qquad \frac{df}{dx} = 2x + 3$

Part B $g(x_1, x_2) = x_1^2 + 2x_1 x_2 + 3x_2^2$

$$\frac{\partial g}{\partial x_1} = 2x_1 + 2x_2 \qquad \frac{\partial g}{\partial x_2} = 6x_2 + 2x_1$$

Part C $h(x) = \begin{bmatrix} x^2 \\ 3x+1 \end{bmatrix} \qquad \frac{dh}{dx} = \begin{bmatrix} 2x \\ 3 \end{bmatrix}$

Part D $\nabla_x g = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 6x_2 \end{bmatrix}$

Part E $g(x) = [x_1 \; x_2]\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$g(x) = [x_1, x_2]\begin{bmatrix} ax_1 + bx_2 \\ bx_1 + cx_2 \end{bmatrix} = ax_1^2 + 2b x_1 x_2 + cx_2^2$$

$$x_1^2 + 2x_1x_2 + 3x_2^2 = ax_1^2 + 2b x_1 x_2 + 3x_2^2$$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$$

$$\frac{\partial}{\partial x}(x^T A x) = (A + A^T)x$$

Part G (a) $\frac{\partial}{\partial x}(a^T x) = a$

$$a^T x = \sum_{i=1}^{n} a_i x_i$$

~~Now~~
Now differentiating component wise

$$\cancel{\frac{\partial}{\partial x_j}(a^T x) = a_j}$$

$$\frac{\partial}{\partial x_j}(a_j x_j) = a_j$$

$$\frac{\partial}{\partial x}(a^T x) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a.$$

(b) $\frac{\partial}{\partial x}(x^T x) = 2x$

$$x^T x = \sum x_i^2$$

differentiating ~~column~~ component-wise :-

$$\frac{\partial}{\partial x_i}(x_i^2) = 2x_i$$

$$\frac{\partial}{\partial x}(x^T x) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} = 2x.$$

(c) $$\frac{\partial}{\partial x}(x^T A x) = (A + A^T)x$$

First Scalar expansion:-

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j$$

$$\frac{\partial}{\partial x_k}(x^T A x) = \sum A_{kj} x_j + \sum A_{ik} x_k$$

$$= Ax + A^T x$$

$$= (A + A^T) x$$

$$\frac{\partial}{\partial x}(x^T A x) = (A + A^T) x$$

4.1 $P(D) = 0.01$ $\quad P(+|D) = 0.99$, $P(+|\bar{D}) = 0.05$

$$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.99(0.05)}$$

$$= \frac{\cancel{0.99} \times 0.01}{\cancel{0.99}(0.06)} = \frac{1}{6}$$

4.1.1 likelihood $L(\mu, \sigma^2)$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

4.1.2 log likelihood

$$l = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2$$

4.1.3 differentiating w.r.t μ

$\frac{dl}{d\mu} = 0$ value of μ to get max. Maximum $= \frac{1}{n} \sum x_i$

4.1.4 $\sigma^2 = \frac{1}{n} \sum (x_i - \hat{\mu})^2$

x ——— x ——— x

(5) Linear Regression

1. ans = 10

2. $y = 2x$ pattern was there in data set first thought and intuition

the cubic fxn is $f(x) = (x-2)(x-4)(x-9) - 2x$ and

$f(5)$ equals to $\underline{-22}$

4. I assumed the data-generating process to be linear, linearity provides better generalization when data set is ~~small~~ scarce. prefering simple models.

Hypothesis

Single-value hypothesis model is that the output y is a linear fxn of x.

$$y = a + bx$$

we try to find the best fit values of $\underline{a, b}$

<u>multivalues</u>

when each data point has d features, ~~the~~ we take output to be linear of all features

$$y = a_1 x_1 + a_2 x_2 - - - a_d x_d + a_0 x_0 \quad \{x_0 = 1\}$$

$$y = A^T X$$

Considering (2,4) (4,8) (9,18) to be the data set

<u>Loss is</u> $L(w,b) = \Sigma (wx_i + b - y_i)^2$

~~Loss~~ $L(w,b) = (2w+b-4)^2 + (4w+b-8)^2 + (9w+b-18)^2$

$$\frac{\partial L}{\partial b} = 2(2w+b-4) + 2(4w+b-8) + 2(9w+b-18) = 0$$

$$\cancel{15w+3b} \cdot \boxed{5w + b = 10}$$

$$\frac{\partial L}{\partial w} = 2(2w+b-4) \cdot 2 + 2(4w+b-8)4 + 2(9w+b-18) \cdot 9 = 0$$

$$\boxed{101w + 15b = 202}$$

Solving the 2 eq^n we get $w = 2, b = 0$

11 $\hat{y} = Xw$

12 $L(w) = |Xw - y|^2$

~~13~~ $\cancel{w = (X^T X)^{-1} X^T} y$

$$L(w) = \|Xw - y\|^2$$

(13) $\hat{y} = Xw$

$$L(w) = (Xw - y)^T (Xw - y)$$

$$L(w) = (Xw)^T (Xw) - 2y^T(Xw) + y^T y$$

$$= w^T X^T X w - 2w^T X^T y + y^T y \quad \{ y^T(Xw) \text{ is scalar} \}$$

differentiate w.r.t w

$$\nabla_{w} L(w) = 2X^T X w - 2X^T y$$

Set gradient to zero. $\cancel{2} X^T X w = \cancel{2} X^T y$

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