Question 2 : Probability Refresher

Q2):

$$P(H|H50) = P(T|H50) = 0.5$$

$$P(H|H60) = 0.6 \quad , \quad P(T|H60) = 0.4$$

$$P(H50) = P(H60) = 0.5$$

a) (i) $P(H50|T) = ?$

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)} \rightarrow (1)$$

From (1), we can say that :

$$P(H50|T) = P(T|H50) \cdot P(H50) / P(T)$$

$$P(T|H50) = 0.5 \quad , \quad P(H50) = 0.5$$

$$P(T) = P(T|H50) \cdot P(H50) + P(T|H60) \cdot P(H60)$$

$$= 0.5 \times 0.5 + 0.4 \times 0.5 = 0.5 \times 0.9 = 0.45$$

$P(T) = 0.45$

$$P(H50|T) = \frac{0.5 \times 0.5}{0.45} = 25/45 = 5/9$$

Hence $P(H50|T) = 5/9$

a) (ii) The Sample size $S = \{T, H, H, H\}$

$$P(H50|S) = \frac{P(S|H50) \cdot P(H50)}{P(S)}$$

where $P(S) = P(S|H50) \cdot P(H50) + P(S|H60) \cdot P(H60)$

$$P(S) \Rightarrow (0.5)^4 \cdot (0.5) + (0.4)(0.6)^3(0.5)$$

$$P(H50|S) = \frac{(0.5)^4 \cdot (0.5)}{((0.5)^4 \cdot (0.5) + (0.4)(0.6)^3(0.5))}$$

$P(H50|S) = 0.42$

a) (iii) $P(H50) = P(H55) = P(H60) = 1/3$

Let S be the sample size describing the event of 9 Heads & 1 Tail, the order of the same does not matter.

We need to calculate the following:

$P(H50|S)$, $P(H55|S)$, $P(H60|S)$

$$P(H50|S) = P(S|H50) \cdot P(H50) / P(S)$$

$$P(S) = P(S|H50) \cdot P(H50) + P(S|H55) \cdot P(H55) + P(S|H60) \cdot P(H60)$$

$$P(S|H50) = (0.5 * 0.5^9)$$

$$P(S|H55) = (0.45 * 0.55^9)$$

$$P(S|H60) = (0.6)^9 (0.4)$$

$$P(H60|S) = \frac{1/3 \cdot (0.4) \cdot (0.6)^9}{1/3 \left((0.5 \cdot (0.5)^9) + (0.45) \cdot (0.55)^9 + (0.4) \cdot (0.6)^9\right)}$$

$$P(H60|S) = 0.596$$

$$P(H55|8) = \frac{1/3\left((0.45)\cdot(0.55)^9\right)}{1/3\left((0.5\cdot(0.5)^9) + (0.45)\cdot(0.55)^9 + (0.4)\cdot(0.6)^9\right)}$$

$$P(H55|8) = 0.293$$

$$P(H50|8) = \frac{1/3\left((0.5)\cdot(0.5)^9\right)}{1/3\left((0.5\cdot(0.5)^9) + (0.45)\cdot(0.55)^9 + (0.4)\cdot(0.6)^9\right)}$$

$$P(H50|8) = 0.148$$

Hence,

$P(H50|8) \simeq 0.148$, $P(H55|8) \simeq 0.293$, $P(H60|8) \simeq 0.596$

---

2) a) $P(\text{Positive Pregnant}) = 0.99$, $P(\text{Negative}|\text{Pregnant}) = 0.01$

$P(\text{Positive, Not Pregnant}) = 0.10$, $P(\text{Negative}|\text{Not Pregnant}) = 0.9$

$P(\text{Pregnant}) = 0.01$, $P(\text{Not Pregnant}) = 0.99$.

We consider X to denote if the women is pregnant (1) or not (0).

and Y be the test result i.e Positive (1) and Negative (0).

$$P(X=1 | Y=1) = \frac{P(Y|X).P(X)}{P(Y)}$$

$$= \frac{(0.99 \times 0.01)}{(0.99)(0.01) + (0.01)(0.99)} = 0.09$$

$P(X=1|Y=1) = 0.09$

Given that the 99 % of the female population is not pregnant and the model false detects (10 %) of this total population. So, the result that we obtain makes sense as the number of false detections are many more than actual pregnant women. Hence the model's confidence is less and is a bad test.

---

2) c) $E[\alpha f(x) + \beta g(x)] = \alpha . E[f(x)] + \beta . E[g(x)]$

$E[Ax + b] = E[Ax] + E[b]$

$$E[Ax] = E\left(\sum_{j=1}^{n} A_{i,j} x_j\right) = \left(\sum_{j=1}^{n} A_{i,j} E(x_j)\right)$$

$$E(Ax) = \left(\sum_{j=1}^{n} A_{i,j} \cdot E(x)_j \right) = A \cdot E(x)$$

Hence $E(Ax) = A \cdot E(x)$

So, $\underline{E(Ax+b) = A \cdot E(x) + b}$

2) d) We know that :

$$\operatorname{cov}(x) = E\left((x - Ex)(x - Ex)^T\right)$$

$$\operatorname{cov}(Ax+b) = E\left((Ax + \cancel{b} - A \cdot E(x) - \cancel{b})(Ax + \cancel{b} - A \cdot E[x] - \cancel{b})^T\right)$$

$$\operatorname{cov}(Ax+b) = E\left(\left(A(x - E(x))\right)\left(A(x - E(x))\right)^T\right)$$

$$\Rightarrow E\left(A(x - E(x))\left((x - E(x))\right)^T \cdot A^T\right)$$

$$\Rightarrow A \cdot E\left((x - E(x))\left((x - E(x))\right)^T\right) \cdot A^T$$

$$\Rightarrow A \cdot \operatorname{cov}(x) \cdot A^T$$

Hence $\underline{\operatorname{cov}(Ax+b) = A \cdot \operatorname{cov}(x) \cdot A^T.}$

Question 3 : Multivariate Derivatives

3) a) Given $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{n \times m}$

$$\underset{(1 \times n)}{x^T} \; \underset{(n \times m)}{A} \; \underset{(m \times 1)}{y}$$

$$\nabla_x x^T A y = \frac{\partial x^T A y}{\partial x}$$

$$\Rightarrow \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \underset{N \times M}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & \\ & & \cdots & a_{nm} \end{bmatrix}} \underset{M \times 1}{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}$$

$$\Rightarrow \underset{1 \times N}{\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}} \underset{N \times 1}{\begin{bmatrix} a_{11} y_1 + a_{12} \cdot y_2 + \cdots a_{1m} \cdot y_m \\ \vdots \\ a_{n1} y_1 + \cdots \quad \cdots + a_{nm} \cdot y_m \end{bmatrix}}$$

Hence $x^T A y = x_1 (a_{11} y_1 + a_{12} \cdot y_2 + \cdots a_{1m} \cdot y_m) + \cdots + x_n (a_{n1} y_1 + \cdots a_{nm} \cdot y_m)$

$$x^T A y = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i (a_{ij} \cdot y_j)$$

$$\nabla_x x^T A y = \frac{\partial \left(\sum \sum x_i (a_{ij} \cdot y_j) \right)}{\partial x} = \sum_{i=1}^{n} \sum_{j=1}^{m} (a_{ij} \cdot y_j)$$

$$\nabla_x x^T A y = Ay$$

Hence $\nabla_x x^T A y = Ay$

3) b) From the previous question, we know that:-

$$x^T A y = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i (a_{ij} \cdot y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} y_j (a_{ij} \cdot x_i)$$

$$\nabla_y x^T A y = \frac{\partial \left(\sum_{i=1}^{n} \sum_{j=1}^{m} y_j (a_{ij} \cdot x_i) \right)}{\partial y} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \cdot x_i$$

Converting into a matrix form we obtain:

$$\nabla_y x^T A y = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A^T x.$$

Hence $\nabla_y x^T A y = A^T x.$

3 c) $\nabla_A x^T A y = \begin{bmatrix} \frac{\partial x^T A y}{\partial a_{1,1}} & \frac{\partial x^T A y}{\partial a_{1,2}} & \cdots & \frac{\partial x^T A y}{\partial a_{1,m}} \\ \vdots & & & \\ \frac{\partial x^T A y}{\partial a_{n,1}} & \cdots & \cdots & \frac{\partial x^T A y}{\partial a_{n,m}} \end{bmatrix}$

where $\partial \left(\sum_{i=1}^{n} \sum_{j=1}^{m} (x_i a_{ij}) \cdot y_j \right) / \partial a$

Hence $\nabla_A x^T A y = x y^T$.

3) d) Given $A \in \mathbb{R}^{n \times n}$ and $f = x^T A x + b^T x$

$\nabla_x f = \partial f / \partial x = \partial (x^T A x + b^T x) / \partial x$

$\nabla_x f = \partial (x^T A x) / \partial x + \partial (b^T x) / \partial x$

We clearly know that: $\partial (b^T x) / \partial x = b^T \rightarrow 1$

We could solve $\partial (x^T A x) / \partial x$

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

$$\partial (x^T A x) / \partial x_k = \sum_{j=1}^{n} a_{kj} \cdot x_j + \sum_{i=1}^{n} a_{ik} \cdot x_i$$

$\forall \; k = 1, 2 \ldots n$

Hence $\partial (x^T A x) / \partial x_k = (A^T + A) x$

$$\nabla_x x^T A x = x (A^T + A) + b$$

3) e) Given: $f = tr(AB)$

$A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times m}$ Where $A \times B \in \mathbb{R}^{m \times m}$

$$\nabla_A f = \partial f / \partial A = \partial (tr(AB)) / \partial A$$

$$AB = \begin{array}{cc} (M \times N) & (N \times M) \\ \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} & \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & & \\ b_{n1} & \cdots & b_{nm} \end{bmatrix} \end{array}$$

$$= \begin{bmatrix} - a_1 - \\ - a_2 - \\ \vdots \\ - a_m - \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_r \\ | & | & & | \end{bmatrix}$$

$$= tr\left(\begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_r \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_r \\ \vdots & & & \\ a_M^T b_1 & a_M^T b_2 & \cdots & a_M^T b_r \end{bmatrix}\right)$$

$$= \sum_{i=1}^{N} a_{1i} b_{i1} + \sum_{i=1}^{N} a_{2i} \cdot b_{i2} + \cdots$$

$$\partial(tr(AB)) / \partial a_{ij} = b_{ji}$$

Hence $\nabla_A tr(AB) = B^T$

Question 4 : Deriving Least Squares with matrix derivatives

We shall first differentiate the objective function wrt to W

$$\frac{1}{2}\sum_{i=1}^{n} \|y_i - Wx_i\|^2 = \frac{1}{2}\sum_{i=1}^{N} (y_i - Wx_i)^T (y_i - Wx_i)$$

$$\Rightarrow \frac{1}{2}\sum_{i=1}^{N} \left(y_i^T y_i - y_i^T W x_i - x_i^T . W^T y_i + x_i^T W^T W x_i \right)$$

$$= \frac{1}{2}\sum_{i=1}^{N} \left(y_i^T y_i - 2 y_i^T W x_i + x_i^T . W^T W x_i \right)$$

We know that $\langle y_i^T, Wx_i \rangle = tr(y_i^T W x_i)$

$$= \sum_{i=1}^{N} \left(\frac{1}{2} y_i^T y_i - tr(y_i^T W x_i) + \frac{1}{2} tr(x_i^T W^T W x_i) \right)$$

$$= \sum_{i=1}^{N} \left(\frac{1}{2} y_i^T y_i - tr(W x_i y_i^T) + \frac{1}{2} tr(W x_i x_i^T W^T) \right)$$

$$= \sum_{i=1}^{N} \frac{1}{2} y_i^T y_i - tr\left(W \sum_{i=1}^{N} x_i y_i^T\right) + \frac{1}{2} tr\left(W \sum (x_i x_i^T) W^T\right)$$

$$= \sum_{i=1}^{N} \frac{1}{2} y_i^T y_i - tr(W X Y^T) + \frac{1}{2} tr(W X X^T W^T)$$

$$\nabla_W h = 0 - YX^T + \frac{1}{2}(WXX^T + WXX^T) = 0$$

$$\Rightarrow \quad -yx^T + Wxx^T = 0$$

Solving for W, we get:

$$W = yx^T(xx^T)^{-1}$$