ML Assignment
(Mathematics for ML)

– Manoj Sampath
17110043
5th yr, MSC, SCS.

I) MINIMUM BACKGROUND TEST

A) VECTORS and MATRICES

① $X = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}_{2\times2}$ $y = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ $z = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

$y^T z$ (inner product) $= \begin{bmatrix} 1 & 3 \end{bmatrix}_{1\times2} \begin{bmatrix} 2 \\ 3 \end{bmatrix}_{2\times1}$

$= 1(2) + 3(3)$
$= \underline{\underline{11}}$

② Xy

$\begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}_{2\times2} \begin{bmatrix} 1 \\ 3 \end{bmatrix}_{2\times1} = \begin{bmatrix} 2(1)+4(3) \\ 1(1)+3(3) \end{bmatrix} = \begin{bmatrix} 14 \\ 10 \end{bmatrix}$

③ A matrix is invertible if its non-singular ($\det \neq 0$)

$X = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}$ $\det X = 6-4$
$= \underline{\underline{2}}$
$\neq 0$ ∴ invertible

$X^{-1} = \frac{\text{Adj } X}{\det X}$

$= \frac{1}{2} \begin{bmatrix} 3 & -4 \\ -1 & 2 \end{bmatrix}$

$= \begin{bmatrix} 3/2 & -2 \\ -1/2 & 1 \end{bmatrix}$

④ Rank of X = Order of X (since $\det X \neq 0$)
$= \underline{\underline{2}}$

[As the rows or columns cannot be expressed in the form of other rows or columns are they are independent]

(II) Calculus

1) $y = x^3 + x - 5$

$$\frac{d}{dx}(y) = \frac{d}{dx}(x^3 + x - 5) = 3x^2 + 1$$

2) $f(x_1, x_2) = x_1 \sin(x_2) e^{-x_1}$

$$\nabla f(x) = \begin{pmatrix} \partial_{x_1} f \\ \partial_{x_2} f \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix}$$

$$= \begin{pmatrix} \sin(x_2) e^{-x_1} - x_1 \sin(x_2) e^{-x_1} \\ x_1 \cos(x_2) e^{-x_1} \end{pmatrix}$$

$$= \begin{pmatrix} \sin(x_2) e^{-x_1} (1 - x_1) \\ x_1 \cos(x_2) e^{-x_1} \end{pmatrix}$$

(III) Probability and Statistics

1) $S = \{1, 1, 0, 1, 0\}$

$$\text{Sample mean} = \frac{1+1+0+1+0}{5} = 3/5$$

2) $$\text{Sample variance} = s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$$

$$= \frac{1}{4}\left[\left(1-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(0-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(0-\frac{3}{5}\right)^2\right]$$

$$= \frac{1}{4}\left[\left(\frac{2}{5}\right)^2 + \left(\frac{2}{5}\right)^2 + \left(\frac{-3}{5}\right)^2 + \left(\frac{2}{5}\right)^2 + \left(\frac{-3}{5}\right)^2\right]$$

$$= \frac{1}{4}\left[\frac{4}{25} + \frac{4}{25} + \frac{9}{25} + \frac{4}{25} + \frac{9}{25}\right]$$

$$= \frac{30}{100} = \frac{3}{10}$$

3) Probability of observing this data (unbiased)

$$P(S) = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

4) Given $P(x=1) \neq 0.5$, to maximize we should write a function and then find where it reaches maximum.

Let $P(x=1) = p$

So now $P(S) = \prod_{i=1}^{5} p^{x_i}(1-p)^{1-x_i}$

$$(f(p)) = p^{\sum_{i=1}^{5} x_i}(1-p)^{5-\sum_{i=1}^{5} x_i} \quad \text{(function of p)}$$

Applying log, $\log(f(p)) = \left(\sum_{i=1}^{5} x_i\right)\log p + \left(5-\sum_{i=1}^{5} x_i\right)\log(1-p)$

This would reach max when $\frac{d}{dp}(\log(f(p))) = 0$.

i.e $0 = \frac{1}{p}\sum_{i=1}^{5} x_i - \frac{1}{1-p}\left(5-\sum_{i=1}^{5} x_i\right)$

$$= \frac{\left[\sum_{i=1}^{5} x_i\right][1-p] - 5p + p\sum_{i=1}^{5} x_i}{p(1-p)}$$

$$= \frac{\sum_{i=1}^{5} x_i - \cancel{p\sum_{i=1}^{5} x_i} - 5p + \cancel{p\sum_{i=1}^{5} x_i}}{p(1-p)}$$

$\Rightarrow 5p = \sum_{i=1}^{5} x_i$

$\Rightarrow \boxed{p = \frac{1}{5}\sum_{i=1}^{5} x_i}$

Hence at this p, the function maximizes.

$\therefore$ Required $P = \frac{1}{5}[1+1+0+1+0]$

$= 3/5$

5)

		y		
		a	b	c
z	T	0.2	0.1	0.2
	F	0.05	0.15	0.3

• What is $p(z=T \text{ and } y=b) = 0.1$

• What is $p(z=T \mid y=b)$ (conditional probability) $= \dfrac{p(z=T \text{ and } y=b)}{p(y=b)}$

$$= \frac{0.1}{0.1+0.15} = \frac{0.1}{0.25} = 0.4$$

IV Big-O-Notation

(f, g) below, state whether each is true: $f(n) = O(g(n))$, $g(n) = O(f(n))$

1) $f(n) = \ln_e(n)$, $g(n) = \lg(n) = \log_2 n$

$$\left[\log_a x = \frac{\log_b x}{\log_b a}\right]$$

A) $f(n) = \log_e(n)$ | $g(n) = \log_2 n = \dfrac{\log_e n}{\log_e 2}$

$\Rightarrow$ Since functions are related by a constant, both are true.

2) $f(n) = 3^n$, $g(n) = n^{100}$

$g(n) = O(f(n))$ as $f(n)$ grows rapidly for large n

3) $f(n) = 3^n$, $g(n) = 2^n$

$g(n) = O(f(n))$ as $f(n)$ grows rapidly than 2^n for large n.

4) $f(n) = 1000n^2 + 2000n + 4000$, $g(n) = 3n^3 + 1$

$f(n) = O(g(n))$ as $g(n)$ grows rapidly for large n.

Medium Background Test

1) Algorithms

Since it is given that all 0's occur before 1's it is assured that there is a single point of transition. Given a subarray from the actual array, we can determine if this point of transition lies inside or outside the subarray. Further, if the point lies inside this subarray, we can divide array into 2 parts and recursively find the part which contains the transition point.

Explanation

→ We have the low value, i, and high value j. Array [i] would be 0 and Array [j] would be 1 which means that definitely the point lies inside this sub array.

Case 1	Case 2
If $i \geq j$	If $i < j$
• If i equals or greater than j means that this is transition point, end of algorithm	a) If mid value is 0, means the transition point lies in right half of sub array, recursively call algorithm for $\boxed{low = mid}$, $\boxed{high = j}$ b) If mid value is 1, means transition point lies in the left half of sub array, recursively call algorithm for $\boxed{low = i}$, $\boxed{high = mid}$

Time complexity

For a sub array of size 1, the execution time is constant

$T(1) = 1$

For a sub array of size n, the execution time would be an offset and the time taken for the next sub array of size $n/2$

$T(n) = k + T(n/2)$

Which evaluates to $O(\log n)$ by masters theorem.

2) Probability and Random Variables

(a) $P(A \cup B) = P(A \cap (B \cap A^c))$

$= P((A \cap B) \cap (A \cap A^c))$

$= P(A \cap B) \cap \phi) \neq P(A \cup B)$

False.

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

True

(as $P(A \cap B)$ is counted twice in $A \cup B$)

(c) $P(A) = P(A \cap B) + P(A^c \cap B)$

False

(d) $P(A/B) = \dfrac{P(A \cap B)}{P(B)}$

$P(B/A) = \dfrac{P(A \cap B)}{P(A)}$ $\Rightarrow P(A/B) \neq P(B/A)$

False

(e) $P(A_1 \cap A_2 \cap A_3) = P(A_3 | (A_2 \cap A_1)) \, P(A_2 | A_1) \, P(A_1)$

$= \dfrac{P(A_3 \cap (A_2 \cap A_1))}{\cancel{P(A_2 \cap A_1)}} \cdot \dfrac{\cancel{P(A_2 \cap A_1)}}{\cancel{P(A_1)}} \cdot \cancel{P(A_1)}$

$= P(A_1 \cap A_2 \cap A_3)$ True

3) Discrete and Continuous Distributions

Matching

(a) Multivariate gaussian : $\frac{1}{\sqrt{(2\pi)^d|\Sigma|}} e^{\left(-\frac{1}{2}-(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}$ A

(b) Bernoulli : $p^x(1-p)^{1-x}$ e

(c) Uniform : $\frac{1}{b-a}$ when $a \le x \le b$ else 0 f

(d) Binomial : $\binom{n}{x} p^x (1-p)^{n-x}$ g

4) Mean, Variance and Entropy :

(a) given X random variable with finite expectation $EX < \infty$. Recall that the variance of a random variable is defined as $Var(X) = E[(X-EX)^2]$. Prove that $Var(X) = E(X^2) - E(X)^2$

A) $Var(X) = E[(X-E(X))^2]$

$= E[X^2 - 2XE[X] + (E[X])^2]$

$= E[X^2] - 2E[X E[X]] + E[X]^2$

$= E[X^2] - 2E[X]^2 + E[X]^2$

$= E[X^2] - E[X]^2$

(b) Mean $= p$

Variance $= p(1-p)$

Entropy $= -(1-p)\log(1-p) - p\log p$

of a Bernoulli (p) random variable

5) Law of Large Numbers and Central Limit Theorem

1(a) Die rolled 6000 times, number of times 3 shows up is close to 1000

A) Since its unbiased, $P(3) = \frac{1}{6}$, for 6000 throws.. close to 1000

from Law of Large Numbers

(b) Fair coin tossed n times and $\bar{X}$ denotes avg number of heads, then distribution of $\bar{X}$ satisfies

$$\sqrt{n}(\bar{x}-1/2)] \xrightarrow{n\to\infty} N(0, 1/4)$$

From Central limit theorem for large n LHS tends to RHS.

6) LINEAR ALGEBRA

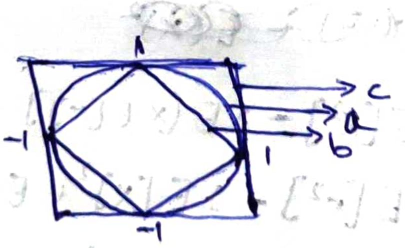
Vector norms

(a) $\|x\|_2 \le 1$ (Recall $\|x\|_2 = \sqrt{\sum_i x_i^2}$)

(b) $\|x\|_1 \le 1$ (Recall $\|x\|_1 = \sum_i |x_i|$)

c) $\|x\|_\infty \le 1$ (Recall $\|x\|_\infty = \max_i |x_i|$)

Soln



Geometry

(a) Show that the vector w is orthogonal to the line $w^T x + b = 0$ (x_1, x_2 lie on line, $w^T(x_1 - x_2)$ inner product

A) Consider x_1 and $_2$ on $w^T x + b = 0$

then $x_1 - x_2$ is vector ||el to this line

now,

$$w^T x_1 + b = 0$$
$$w^T x_2 + b = 0 \quad \Rightarrow \quad w^T x_1 = w^T x_2$$
$$\Rightarrow w^T (x_1 - x_2) = 0$$

$\Rightarrow$ w is orthogonal to our line.

(b) Distance from origin to line is $\frac{b}{\|w\|}$

A) Let x be a point on $w^T x + b = 0$

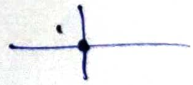
The distance from origin to the hyperplane is computed by the projection of x onto the normal vector of the hyperplane given by w.

Hence this distance $= \frac{|w^T x|}{\|w\|_2} = \frac{|-b|}{\|w\|_2} = \frac{|b|}{\|w\|_2}$

7) Programming Skills

from the graphs in the code attached

(a) Plotted



(b) Since mean moves from [0,0] to [-1,1] the data move up and left

(c) Since variance is doubled, we see more spacing

(d) Data gets directed towards (bottom left to top right)

(e) Data gets directed towards (bottom right to top left)