Sanjana Agarwal.

MACHINE LEARNING ASSIGNMENT 1.

1. Given : $\Omega_x = \{a, b, c\}$

$p_x(a) = 0.1$, $p_x(b) = 0.2$ & $p_x(c) = 0.7$

$$f(x) = \begin{cases} 10 & \text{if } x = a \\ 5 & \text{if } x = b \\ 10/7 & \text{if } x = c \end{cases}$$

(a.) We know, $E(x) = \sum_{x \in \Omega} f(x) \cdot p_x(x)$

$$E[f(x)] = f(a) \cdot p_x(a) + f(b) \cdot p_x(b) + f(c) \cdot p_x(c)$$

$$= 10 \times 0.1 + 5 \times 0.2 + \frac{10}{7} \times 0.7$$

$$= \boxed{3}$$

(b) To find $E[1/p_x(x)]$

Now, $E[1/p_x(x)] = \sum_{x \in \Omega} \frac{1}{p_x(x)} \cdot p_x(x)$

$$= \frac{1}{p_x(a)} \cdot p_x(a) + \frac{1}{p_x(b)} \cdot p_x(b) + \frac{1}{p_x(c)} \cdot p_x(c)$$

$$= \frac{1}{0.1} \times 0.1 + \frac{1}{0.2} \times 0.2 + \frac{1}{0.7} \times 0.7$$

$$= \boxed{3}$$

(c.) For an arbitrary pmf $p_x(x)$, the $E\left[1/p_x(x)\right]$ will depend on the cardinality of Ω.

As seen in 1.(b) for the given sample space $\Omega_x = \{a, b, c\}$, the expected value of $1/p_x(x)$ is 3.

Thus, it depends on the cardinality of Ω.

2.(b.) $X = a_1 x_1, a_2 x_2, \ldots, a_m x_m$

$$\mathrm{cov}(X) = \mathrm{cov}(X, X)$$

$$X = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} x_1, & x_2, & \ldots & x_m \end{bmatrix}$$

$$\mathrm{cov}[X, X] = \mathrm{var}[X]^{[1]}$$

$$= \mathrm{Var}\left(\sum_{i=1}^{m} a_i x_i\right)$$

$$= \sum_{i=j}^{m} a_i a_j \mathrm{cov}[x_i, x_j]$$

$$= \sum_{i=1}^{m} a_i^2 \mathrm{Var}[x_i] + \sum_{i \neq j} a_i a_j \mathrm{Cov}[x_i, x_j]$$

Since $x_1, \ldots, x_m$ are independent, we get,

$$\mathrm{Cov}[x_i, x_j] = 0 \quad \text{where } i \neq j$$

$$\therefore \begin{bmatrix} \mathrm{Cov}[x_1, x_1] & 0 & \cdots & 0 \\ 0 & \mathrm{Cov}[x_2, x_2] & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \mathrm{cov}[x_m, x_m] \end{bmatrix}$$

$\because$ For multivariate Gaussian, variance is Σ,

$\mathcal{N}(\mu, \Sigma)$, $\mathrm{Cov}[x, x] = \sum_{i=j}^{m} a_i^2 \Sigma$

For part 2, if $x_1, \dots x_M$ are not independent then $cov[x_i, x_j]$ where $i \neq j$, $\neq 0$

Given, $Cov[x_1, x_2] = \lambda$

We know, $Cov[X, Y] = Cov[Y, X]^T$

$\therefore$ $Cov[x_2, x_1] = \lambda^T$

Thus,

$$Cov[x, x] = \sum_{i,j=1}^{M} a_i^2 \Sigma + a_1 a_2 \lambda + a_1 a_2 \lambda^T$$

Thus, dimension of $Cov[x]$ is $d \times d$.

2.1a) Given $X = a_1 X_1 + a_2 X_2 + \dots a_M X_M$.

$X_1, \dots X_M$ are independent multivariate Gaussian random variables.

To find $E[x]$.

As per the notes, we know,

$$E[x] = E[\sum_{i=1}^{M} a_i x_i]$$

$$E[x] = \sum_{i=1}^{M} E[a_i x_i]$$

$$E[x] = \sum_{i=1}^{M} a_i E[x_i] \quad (\because E[cx] = c \cdot E[x])$$

The expected value for Gaussian is μ.

$$\therefore E[x] = \sum_{i=1}^{M} a_i \mu_i$$

The dimension of $E[X]$ is $1 \times d$

3.(a) When the code is run for dim $=1$, $\sigma = 1.0$:

$n = 10$

~~A~~ We get $\mu = -0.087$

$n = 100$

$\mu = -0.05$

$n = 1000$

$\mu = 0.004$

When the code is run for dim $=1$, $\sigma = 10$

$n = 10$

$\mu = -2.21$

$n = 100$

$\mu = 1.647$

$n = 1000$

$\mu = 0.358$

We notice that ~~the~~ as the number of samples increases, the mean approaches more & more towards 0.

With the increase in value of σ, the sample mean approaches more towards the average mean.

(b) The covariance for dim = 3 is

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Since all the other elements i.e. $cov(x, y)$, $cov(y, z)$, etc. are 0 & $cov(x, x)$, $cov(y, y)$ & $cov(z, z)$ is 1, this means that they are independent.

Now, the covariance changes to,

$$\Sigma = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

We see that the $cov(x, z) = cov(z, x) = 1$
Thus, all the samples have dependent x & z values.

4. Prior density $= f(\lambda) = \theta \cdot e^{-\theta\lambda}$

Given : Poisson distribution $= \dfrac{\lambda^k \cdot e^{-\lambda}}{k!}$

where $\lambda \in (0, \infty)$

(a) Since we do not have any observed data, we do not know what distribution does the data follow. So we have to find the most likely value for λ.

$$\lambda = \text{argmax}_\lambda\, f(\lambda)$$

$$= \text{argmax}_\lambda\, \theta \cdot e^{-\theta\lambda}$$

$\because \lambda \in (0, \infty)$

it will be a value very close to 0.

However if we assume, $\lambda \in [0, \infty)$,

$\lambda = 0$.

(b) ~~Assumption of~~ ~~Most~~ Maximum likelihood for Model M is, [2]

$$M_{ML} = \underset{m \in \mathcal{M}}{\text{argmax}} \{ p(D|M) \}$$

D is data, substituting M with λ,

$$\lambda_{ML} = \underset{\lambda \in (0,\infty)}{\text{argmax}} \{ p(D|\lambda) \}$$

We can write the likelihood function as,

$$p(D|\lambda) = p(\{x_i\}_{i=1}^{n} | \lambda)$$

$$= \prod_{i=1}^{n} p(x_i|\lambda)$$

$$= \frac{\lambda^{\sum_{i=1}^{n} x_i} \cdot e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

To find λ that maximizes the likelihood, we will first take a logarithm, find its derivative, w.r.t λ, & finally equate it with 0, to find the maximum.

We express log-likelihood $ll(D,\lambda) = \ln p(D|\lambda)$ as $ll(D,\lambda)$

$$= \ln\lambda \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \ln(x_i!)$$

and proceed with the first derivative as

$$\frac{\partial ll}{d\lambda}(D,\lambda) = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0$$

$$\therefore \lambda_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Given: 79 accidents over 9 days

$$\therefore \sum_{i=1}^{n} x_i = 79 \quad \& \quad n = 9.$$

$$\therefore \lambda_{ML} = \frac{79}{9} = \boxed{8.78}$$

4.(c) Given 79 accidents over 9 days.

$$M_{MAP} = \underset{m \in \mathcal{M}}{\operatorname{argmax}} \{ p(D|M) \cdot p(M) \}$$

$$\therefore \lambda_{MAP} = \underset{\lambda \in (0,\infty)}{\operatorname{argmax}} \{ p(D|\lambda) \cdot p(\lambda) \}$$

D is nothing but $\sum_{i=1}^{n} x_i$

Thus, $p(D|\lambda) \cdot p(\lambda) = \prod_{i=1}^{n} p(x_i|\lambda) \cdot p(\lambda)$

$$= \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \cdot \theta \cdot e^{-\theta\lambda}$$

{Using pmf & prior}

To find λ_{MAP}, we first take the log then take first derivative & equate it to zero.

$$= \ln \lambda^{\sum_{i=1}^{n} x_i} + \ln e^{-n\lambda} + \ln \theta + \ln e^{-\theta\lambda} - \ln \prod_{i=1}^{n} x_i!$$

$$= \left(\sum_{i=1}^{n} x_i\right) \ln \lambda - \lambda(n+\theta) + \ln\theta - \ln \prod_{i=1}^{n} x_i!$$

$(\because \ln e = 1)$

After taking first derivative, we get, and equating it to zero

$$\frac{\sum_{i=1}^{n} x_i}{\lambda} - (n+\theta) = 0$$

$$\lambda_{MAP} = \frac{\sum_{i=1}^{n} x_i}{n+\theta}$$

In this case, we get,

$$\lambda_{MAP} = \frac{79}{9+0.5}$$

$$= 8.32$$

4(d) In the above problems, we found the maximum likelihood estimate for 9 days.

So for 10 days we can find as,

$$\lambda_{ML} = \frac{79 + x_{10}}{10}$$

$\therefore$ λ_{ML} for 9 days, is;

$$\frac{79}{9} = \frac{79 + x_{10}}{10}$$

$$\frac{790}{9} - 79 = x_{10}$$

$\therefore$ $x_{10} = 8.78$

$\therefore$ Predicted value for 10^{th} day is 8.78

λ_{MAP} for 10th day, using λ_{MAP} equation, will be,

$$\lambda_{MAP} = \frac{\sum_{i=1}^{n} x_i}{n+\theta}$$

$$\frac{79}{9.5} = \frac{79 + x_{10}}{10 + 0.5}$$

$$x_{10} = 79/9.5$$

$$\therefore x_{10} = 8.32$$

(e.) The importance of prior is that it allows us to treat λ as a random variable & we can then calculate its posterior distribution using Bayes' Theorem.[3]

Thus we can have the probability distribution of an uncertain quantity before having any proper evidence.

This method then estimates λ as the mode of the posterior distribution of the random variable.

4.(f.) Considering exponential distributions, the equation to consider is,

$$f(\lambda) = \theta \cdot e^{-\theta\lambda}$$

Thus, to reflect that the safety measures are effective & the number of accidents per day should sharply decrease, we need to <u>increase θ</u>.

5. (a) To formulate as a maximum likelihood problem:

Let M & N be two discrete random variables.
M represents : Sunny & Not Sunny.

∴

M	Value
0	Not Sunny
1	Sunny

Likewise, N represents Table free & Not free.

N	Value
0	Not free
1	Free

As we have two possible outcomes, we can use Bernoulli Distribution.

∴ Let n be no. of days. ~~table free~~
& k be when table is free
∴ (n-k) means table is not free.

Let p be the parameter estimate for N.

$$\therefore N \rightarrow p^k \cdot (1-p)^{n-k}$$

Similarly, we can formulate a Bernoulli distribution for <u>sunny</u> ~~can~~ weather.

$$\therefore M \rightarrow q^a \cdot (1-q)^{n-a}$$

Thus, we can apply Bayes theorem, as,

$$P(N=1, M=0) = P(M=0 | N=1) \cdot P(N).$$

Likewise other equations.

So, to predict whether table will be free means,

when $M=0$ & $M=1$.

And $N=1$

Thus, we get,

$$P(N=1 | M=0, M=1) = P(N)\{[P(M=0 | N=1)] \cdot [P(M=1 | N=1)]\}$$

Since, $N = p^k (1-p)^{n-k}$

$$f(N) = \prod_{t \to 0} p^k \cdot (1-p)^{n-k}$$

which is the maximum likelihood.

We thus, determine the values of p & q using the estimate value.

Thus, once we get these values, we can find the conditional probabilities as well as k. & a.

5.(b) Given : $n = 10$, Day is Sunny $\therefore M = 1$.
To find : $N = 1$?
$\therefore$ To find $P(N=1, M=1) = ?$

As we used Bayes Theorem earlier, applying it similarly, will give

$$P(N=1, M=1) = P(M=1 \mid N=1) \cdot P(N)$$
$$= \underline{P(M=1 \mid N=1) \cdot p^k}$$

$$\left(\because N \rightarrow p^k \cdot (1-p)^{10-k}\right)$$

5.(c) Let T be the third random variable which represents time of the day.
$\therefore$ T can take values

T	Time of the day
0	Morning
1	Afternoon
2	Evening

$\therefore$ We get a trinomial distribution, which is

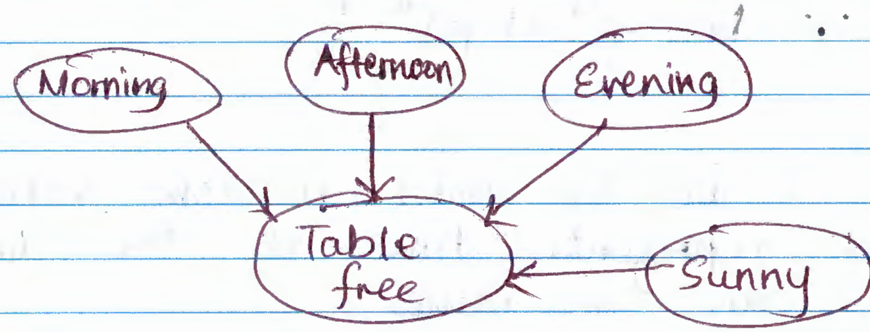
$$T \rightarrow r_0^m r_1^a (1 - r_0 - r_1)^{n-m-a}$$

where r is the parameter estimate value. m, a are 'morning' & 'afternoon' respectively and $(n-m-a)$ is ~~evene~~ 'evening'.

Thus, similar to the Bayes Net problem [4] of Alarm - Burglary, estimating T would depend on all of these values.

Also, prediction of whether table will be free or not free, will depend on T as well.

Thus, to show it diagramatically,



6. P (Coin A) = P(A) = Coin A has heads = 0·75
P(Coin B) = P(B) = Coin B has heads = 0·5
P(Coin C) = P(C) = Coin C has heads = 0·25

(a.) The sample space will be :
{ HHH, HHT, HTH, HTT, THT, TTH, THH, TTT }

P(Getting 0 heads) = { TTT } = $\frac{1}{4} \times \frac{1}{2} \times \frac{3}{4}$

$= \frac{3}{32}$

P(Getting 1 head) = { HTT, THT, TTH }

$= \left[\frac{3}{4} \times \frac{1}{2} \times \frac{3}{4}\right] + \left[\frac{1}{4} \times \frac{1}{2} \times \frac{3}{4}\right] + \left[\frac{1}{4} \times \frac{1}{2} \times \frac{1}{4}\right]$

$= \frac{13}{32}$

P(Getting 2 heads) = { HHT, HTH, THH }

$= \left[\frac{3}{4} \times \frac{1}{2} \times \frac{3}{4}\right] + \left[\frac{3}{4} \times \frac{1}{2} \times \frac{1}{4}\right] + \left[\frac{1}{4} \times \frac{1}{2} \times \frac{1}{4}\right]$

$= \frac{9+3+1}{32} = \frac{13}{32}$

P(Getting 3 heads) = { HHH } = $\frac{3}{4} \times \frac{1}{2} \times \frac{1}{4}$

$= \frac{3}{32}$

$$f(x) = \begin{cases} 0 & \text{when } 0 \text{ heads} \\ 1 & \text{when } 1 \text{ head} \\ 2 & \text{when } 2 \text{ heads} \\ 3 & \text{when } 3 \text{ heads} \end{cases}$$

Thus, $E[X] = \sum_{x \in \Omega} f(x) \cdot p_X(x)$

$$= 0 \times \frac{3}{32} + 1 \times \frac{13}{32} + 2 \times \frac{13}{32} + 3 \times \frac{3}{32}$$

$$= \frac{13 + 26 + 9}{32}$$

$$= \frac{48}{32}$$

$$\therefore E[X] = \boxed{\frac{3}{2}}$$

(b) Let $P(C)$ be probability that coin C is chosen.

Let $P(X)$ be probability that 3 of 5 flips result in heads.

Using conditional probability,

$$P(C|X) = \frac{P(C \cap X)}{P(X)}$$

By Bayes Net, this equals,

$$\frac{P(X|C) \cdot P(C)}{P(X)}$$

Let $P(A)$ & $P(B)$ be respective probabilities of A & B to be chosen.

$$P(X) = P(X|A) \cdot P(A) + P(X|B) \cdot P(B) + P(X|C) \cdot P(C)$$

Now, $P(X|A)$ = Choosing 5 coins, where 3 heads and 2 tails come out (Coin A)

i.e. from 2^5 coins, we can get 3 heads in ${}^5C_3$ ways.

$$= \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{{}^5C_3}{2^5}$$

$$= \frac{3^3 \times 10}{4^5 \cdot 2^5}$$

Similarly, $P(X|B) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{10}{32}$

[Same logic as above]

$$= \frac{1}{2^5} \times \frac{10}{2^5}$$

$$P(X|C) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{10}{32}$$

$$= \frac{9}{4^5} \cdot \frac{10}{2^5}$$

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

$$\therefore P(X) = \frac{1}{3}\left[P(X|A) + P(X|B) + P(X|C) \right]$$

$$= \frac{1}{3}\left[\frac{3^3 \cdot 10}{4^5 \cdot 2^5} + \frac{10}{2^5 \cdot 2^5} + \frac{3^2 \cdot 10}{4^5 \cdot 2^5} \right]$$

Now, we know, that, ~~P(A)~~

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

$$= \frac{\frac{9}{4^5} \times \cancel{\frac{10}{2^5}} \times \cancel{\frac{1}{3}}}{\cancel{\frac{1}{3}} \times \cancel{\frac{10}{2^5}} \times \left[\frac{3^3}{4^5} + \frac{1}{2^5} + \frac{3^2}{4^5} \right]}$$

$$= \frac{\dfrac{9}{48 \cdot 4^3}}{\dfrac{1}{\cancel{4}^2}\left[\dfrac{27}{64} + \dfrac{1}{2} + \dfrac{9}{64}\right]}$$

$$= \frac{9}{\cancel{4}^3 \cdot \left[\dfrac{27+32+9}{64}\right]}$$

$$= \boxed{\frac{9}{68}}$$

References:

[1] For Q.2.(b) www.wikipedia.~~com~~ org /wiki/Variance

[2] For Q.4, referred to notes.

[3] For Q.4(e), wikipedia.org/wiki/Maximum-a-posteriori-estimation.

[4] For Q.5.(c), book Artificial Intelligence A Modern Approach