CS 484 - Introduction to Machine Learning
Assignment - 2

Problem 1:-

1) Compute $P(X=1)$

From the question, we have below values

$P(X=1|Y=1, Z=1) = 0.6$
$P(X=1|Y=1, Z=0) = 0.1$
$P(X=1|Y=0) = 0.2$
$P(Y=1) = 0.9$
$P(Z=1) = 0.8$

we have the values of $P(Y=1)$ and $P(Z=1)$. so, from that we can get $P(Y=0)$ and $P(Z=0)$.

$P(Y=0) = 1 - P(Y=1) = 1 - 0.9 = 0.1$
$P(Z=0) = 1 - P(Z=1) = 1 - 0.8 = 0.2$

Formula for Law of Total Probability is:

$$P(X) = \sum_i P(X|Y=y_i) P(Y=y_i)$$

From the above formula we can write $P(X=1)$ as:

$$P(X=1) = P(X=1 \mid Y=1, Z=1)\, P(Y=1) P(Z=1) +$$
$$P(X=1 \mid Y=1, Z=0)\, P(Y=1)\, P(Z=0) +$$
$$P(X=1 \mid Y=0, Z=1)\, P(Y=0)\, P(Z=1) +$$
$$P(X=1 \mid Y=0, Z=0)\, P(Y=0)\, P(Z=0)$$

So now, Let us substitute the above values in the question in the above formula of $P(X=1)$

$$P(X=1) = (0.6)(0.9)(0.8) + (0.1)(0.9)(0.2) + (0.2)(0.1)(0.8) + (0.2)(0.1)(0.2)$$

$$= 0.432 + 0.018 + 0.016 + 0.004$$
$$= 0.47$$

So, the value of $P(X=1)$ is 0.47.

2) compute the expected value $E[Y]$

The Formula for Expected value is:

$$E[Y] = \sum_i y_i \cdot P(Y=y_i)$$

In the question given, we that Y is a binary variable.

So, Y takes only 0 or 1.

so, now let us substitute 0 and 1 in the Expected value formula.

$$E[Y] = 1 * P(Y=1) + 0 * P(Y=0)$$
$$= 1 * P(Y=1) + 0$$
$$= 1 * P(Y=1)$$
$$= P(Y=1)$$
$$= 0.9$$

so, the expected value of $E[Y]$ is 0.9

3) Compute the expected value $E[Y]$ where Y takes 115 and 20, instead of 0 and 1

Given, $P(Y=115) = 0.9$

so, $P(Y=20) = 1 - P(Y=115)$
$$= 1 - 0.9$$
$$P(Y=20) = 0.1$$

The Formula for Expected value is

$$E[Y] = \sum_i y_i \cdot P(Y=y_i)$$

Let us substitute 115 and 20 in the above formula.

$$E[Y] = 115 * P(Y=115) + 20 * P(Y=20)$$

$$= 115 * 0.9 + 20 * 0.1$$

$$= 103.5 + 2$$

$$= 105.5$$

Hence, the expected value of $E[Y]$ is 105.5

<u>Problem 2</u>:-

In the question it is given that three Factories A, B, C produces 20%, 30%, 50% of the phones with 2%, 1%, 0.05% being defective respectively.

Now, Let us calculate the Probabilities of each one:

Probability of Phones Produced:

Factory A: $P(A) = 0.2$

Factory B: $P(B) = 0.3$

Factory C: $P(C) = 0.5$

Probability of Defective Phones Produced:

Factory A: $P(D/A) = 0.02$

Factory B: $P(D/B) = 0.01$

Factory C: $P(D/C) = 0.0005$

1) So Now, Let us calculate the Probability of Phone being Defective.

Here, we use the Law of Total Probability:

$$P(\text{Defective}) = P(D/A)\,P(A) + P(D/B)\,P(B) + P(D/C)\,P(C)$$

Let us substitute the values into above formula

$$P(\text{Defective}) = (0.02)(0.2) + (0.01)(0.3) + (0.0005)(0.5)$$

$$= 0.004 + 0.003 + 0.00025$$

$$= 0.00725$$

The Probability of a phone produced being defective is 0.00725 (or) 0.725%

2) The Probability of Phone being defective is manufactured at factory A:

$$P(A/D) = \frac{P(A)\,P(D/A)}{P(D)}$$

$$= \frac{0.2 * 0.02}{0.00725}$$

$$= \frac{0.004}{0.00725}$$

$$= 0.5517$$

So, $P(A/D) = 0.552$ (or) 55.2%

3) The Probability of Phone being defective is manufactured at factory B:

$$P(B/D) = \frac{P(B)\,P(D/B)}{P(D)}$$

$$= \frac{0.3 * 0.01}{0.00725}$$

$$= \frac{0.003}{0.00725}$$

$$= 0.413$$

So, $P(B/D) = 0.413$ (or) 41.3%

4) The Probability of Phone being defective is manufactured at factory C:

$$P(C/D) = \frac{P(C)\,P(D/C)}{P(D)}$$

$$= \frac{0.5 * 0.0005}{0.00725}$$

$$= \frac{0.00025}{0.00725}$$

$$= 0.0344$$

So, $P(C/D) = 0.0344$ (or) 3.44%

Problem 3 :-

1) Yes, 1-D transformation is possible for this dataset.

Transformation Expression :-

$$\phi(x_1) = x_1^2$$

By squaring the values, all the points that are negative and those points that are positive but close to zero will be mapped to positive side, which will allow to separa the two classes.

2) Yes, 2-D transformation is possible for this 1-D dataset

Transformation Expression :-

$$\phi(x_1) = (x_1, x_1^2)$$

Here, the transformation would be possible by keeping original x_1 co-ordinate and adding a second co-ordinate by squaring the original co-ordinate.

3) Yes, 1-D transformation is possible for this 2-D dataset.

Transformation Expression :-

$$\phi(x_1, x_2) = x_1^2 + x_2^2$$

The above expression would map the inner circle to lower values and outer circle to higher values which will separate linearly

4) Yes, 2-D transformation of this 2-D dataset is possible.

Transformation Expression:-

$$\phi(x_1, x_2) = (x_1^2 + x_2^2, \tan^{-1}(\frac{x_2}{x_1}))$$

In the above expression, $x_1^2 + x_2^2$ is the radical component which separates the points based on how far they are from the center point. $\tan^{-1}(x_2/x_1)$ is the component which makes the angle of each point relative to origin. Here it makes a way to distinguish points which lie at same distance from the origin but in various different directions.

Kernel of Not:-

1) $K(x, z) = (xz + 1)^2$

Here to prove the above as valid we need to verify that its symmetric and positive semi-definite.

It is clearly symmetric as $(xz+1)^2 = (zx+1)^2$. Also when we expand it as: $x^2z^2 + 2xz + 1$, which is of degree 2.

So, $K(x, z) = (xz+1)^2$ is a valid kernel.

2) $K(x,z) = (xz-1)^3$

The above kernel is symmetric as $(xz-1)^3 = (zx-1)^3$ but it is not positive semi-definite. For example, $x = z = 0$ then

$K(0,0) = (-1)^3 = -1$

so, $K(x,z) = (xz-1)^3$ is not a valid kernel.

Problem 4 :-

1) Given geometric distribution,

$$P(y;\phi) = (1-\phi)^{y-1}\phi \; ; \; y = 1,2,3,\ldots$$

so, now we need to show that the above can be expressed in the form of exponential family :

$$P(y;n) = b(y)\exp(n^T T(y) - a(n))$$

Now let us take log to the geometric distribution on the both sides.

$$\log P(y;\phi) = \log((1-\phi)^{y-1}\phi)$$
$$= \log((1-\phi)^{y-1}) + \log\phi$$
$$= (y-1)\log(1-\phi) + \log\phi$$
$$= y\log(1-\phi) - \log(1-\phi) + \log\phi$$
$$= y\log(1-\phi) - \log\left(\frac{1-\phi}{\phi}\right)$$
$$P(y;\phi) = \exp\left(y\log(1-\phi) - \log\left(\frac{1-\phi}{\phi}\right)\right)$$

Now, we need to identify components by comparing the above with form of exponential family.

$$b(y) = 1$$

$$T(y) = y$$

$$\eta = \log(1-\phi) \quad \text{Here, } 1-\phi = e^{\eta}$$

$$\Rightarrow \phi = 1-e^{\eta}$$

$$a(\eta) = \log\left(\frac{1-\phi}{\phi}\right) = \log\left(\frac{e^{\eta}}{1-e^{\eta}}\right)$$

2) Given training set,

$$\{(x_n, y_n)\}_{n=1}^{N}$$

Also, given the log - likelihood of an example be:

$$l(\omega) = \log p(y_n \mid x_m; \omega)$$

Here, we are taking an assumption of standard GLM

$$\eta = \omega^T x$$

Now,

$$l_n(\omega) = \log\left\{\exp\left(\omega^T x_n \cdot y_n - \log\left(\frac{e^{\omega^T x_n}}{1-e^{\omega^T x_n}}\right)\right)\right\}$$

$$l_n(\omega) = \log\left\{\exp\left(\omega^T x_n \cdot y_n - \log\left(\frac{1}{e^{-\omega^T x_n}-1}\right)\right)\right\}$$

$$l_n(\omega) = \omega^T x_n \cdot y_n + \log(e^{-\omega^T x_n} - 1)$$

Now for the above equation differentiate on both sides.

$$\frac{\partial}{\partial w} l_n(w) = x_n y_n - \frac{1}{1-e^{w^T x_n}} x_n$$

$$= \left(y_n - \frac{1}{1-e^{w^T x_n}}\right) x_n$$

Now, Deriving the stochastic gradient ascent rule.

$$w := w + \alpha \frac{\partial l_n(w)}{\partial w}$$

$$w := w + \alpha \left(y_n - \frac{1}{1-e^{w^T x_n}}\right) x_n$$

5) Problem 5 code and Plots are attached in separate file.