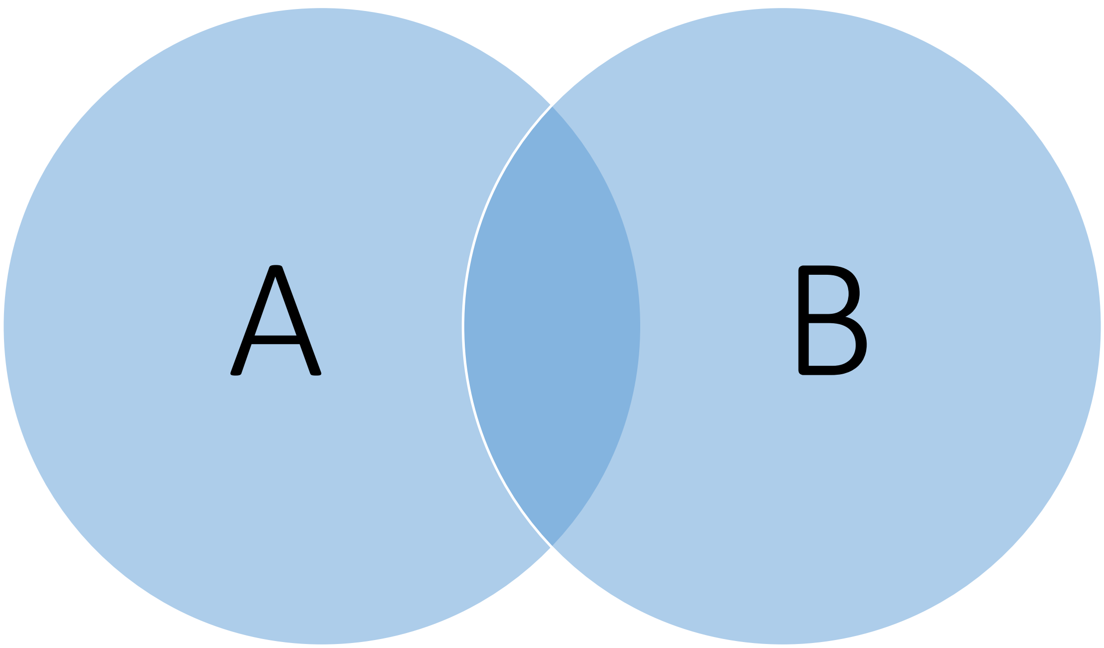
Homework Assignment # 1
B555- Machine Learning

NAME: - FNU ANIRUDH (aanirudh@umail.iu.edu)

Solution 1:-

P(A U B)= P(A)+P(B)-P(A n B)



Proof: - Refer to Figure. Both A U B and A n B= $(A^c \cup B^c)^c$ are events because A and B are events. Similarly, A n B^c and B n A^c are also events.

Notice that A n B^c , B n A^c, and A n B are pairwise disjoints events. Hence,

P(A) + P(B) – P(A n B)

= P ((A n B^c) U (A n B)) + P ((B n A^c) U (A n B)) – P (A n B)
= P (A n B^c) + P (A n B) + P (B n A^c) + P (A n B) – P (A n B)
= P (A n B^c) + P (A n B) + P (B n A^c)
= P((A n B^c) U (A U B) U (B n A^c))
= P(AUB)

Hence Proved P (A)+ P (B)- P(A n B) = P(AUB).

Solution 2:- $P(A)=P(A|B)+P(A|B^c)$

This expression is not true and can be proved easily by following example:-

Consider two fair dices are rolled and Event A and B are mentioned below

Let A: - Sum of two dice is 7 i.e. { (1, 6) (2, 5) (3, 4) (4, 3) (5, 2) (6, 1) }

B: - Sum of two dice is prime number i.e. {2, 3, 5, 7, and 11}

$P(A)= 1/6$

$P(B)= 5/12$

$P(A \cap B)= 6/36= 1/6$

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

$$= \frac{1/6}{5/12}$$

$$= 2/5$$

$$P(A/B^c)= \frac{P(A \cap B^C)}{P(B^c)}$$

$$= \frac{\text{P (Probability sum is 7 and sum is not prime number)}}{\text{P (Sum is not prime number)}}$$

$$= \frac{0}{7/12}$$

$$= 0$$

$1/6 \neq 2/5+ 0$

Hence **$P(A)\neq P(A|B)+P(A|B^c)$**.

Solution 3: -

(a) Let p be the additional wins of Player 1 and q be the additional wins for Player 2

Probability (Player 1 Wining) = P (Player one gets 8 Heads before Player 2 gets 8 tails)

= P (p=8-4 before q=8-6) (Mentioned in problem)
= P (p= 4 before q=2)

Case (i):- P(p=4 and L=0) i.e 4 Heads in a row

$$= 1/2^4 = \frac{1}{16}$$

Case (ii):- P(p=4 and q=1)

= {THHHH, HTHHH, HHTHH, HHHTH} **Note:-**HHHHHT is not to be considered

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

$$= \frac{1}{8}$$

Probability (Player one wins) = Case (i) + Case (ii)

$$= \frac{1}{16} + \frac{1}{8}$$

$$= \frac{3}{16}$$

(b) General Expression for n, m and l
Assuming $0 \leq m < n$ and $0 \leq l < n$.

Let p be number of wins by Player One
Let q be number of wins by Player Two

P (Player one wins) = P (p=n before q=n)

= P [(p=n and q=l) OR (p=n and q= l+1) OR (p=n and q=n-1)]

= $\sum^{n-1}_{i=1}$ P (p=n and q=i)

Now $0 \leq i < n$

First, We need to complete a win by player one which implies getting n-m more heads.

Since tosses are independent, so this probability is $(1/2)^{n-m}$.

Next, There will be i tails which gives a probability of $(1/2)^{i}$.

Finally, tails can occur in i of exactly n-m+1 different places to have heads and complete a win, to which we must add i places for the fact that last toss must always be head. Since order does not matter, there are $\binom{n-m+i-1}{i}$ places to put i tails. Putting all this together, we conclude that for given i:

$$P(p=n \text{ and } q=i) = {}^{n-m+i-1}C_i \, (1/2)^{n-m} (1/2)^{i}$$

$$= {}^{n-m+i-1}C_i \, (1/2)^{n-m} (1/2)^{n-m+i}$$

Final answer is sum of all of i where $0 \leq i < n-(l+1)$.

$$\text{Hence } P(\text{Player One Wins}) = \sum^{n-(l+1)}_{i=0} P(p=n \text{ and } q=i)$$

$$= \sum^{n-(l+1)}_{i=0} {}^{n-m+i-1}C_i \, (1/2)^{n-m+i}$$

Veracity of this expression can be tested by substituting values from (a) and checking result

n=8 , m=4 and l=6

$$= \sum_{i=0} {}^{3}C_i \, (1/2)^{4+i} = \sum_{i=0} {}^{3}C_0 \, (1/2)^{4+0}$$

$$= {}^{3}C_0 \, (1/2)^{4+0} + {}^{4}C_1 \, (1/2)^{4+1}$$

$$= {}^{3}C_0 \, (1/2)^{4} + {}^{4}C_1 \, (1/2)^{5}$$

$$= 1 \times (1/16) + 4 \times (1/32) = 3/16$$

Hence General expression derived holds good.

Solution 5:-

(a) Given P (A)= 0.75 --- Coin 1 showing heads
P (B)= 0.50---- Coin 2 showing heads
P (C)= 0.25---- Coin 3 showing heads

Since all the three coins are flipped together

Coin A	Coin B	Coin C	Probability of Events
T	T	T	0.25 x 0.5 x 0.75
T	T	H	0.25 x 0.5 x 0.25
T	H	T	0.25 x 0.5 x 0.75
T	H	H	0.25 x 0.5 x 0.25
H	T	T	0.75 x 0.5 x 0.75
H	T	H	0.75 x 0.5 x 0.25
H	H	T	0.75 x 0.5 x 0.75
H	H	H	0.75 x 0.5 x 0.25

Let X be the number of heads obtained by flipping all the three coins together then

$$E[X] = \Sigma^{3}_{X=0} X * P(X)$$

X	P(X)
0	P(TTT)
1	P(TTH) + P(THT)+ P(HTT)
2	P(THH)+ P(HTH)+ P(HHT)
3	P(HHH)

E [X]= 0 * P(TTT) + 1 * ((P(TTH) + P(THT) + P(HTT)) + 2 ((P(HHT) + P(HTH)+ P(THH)) + 3 * P(HHH)
= 0+ 1 ((0.25 x 0.5 x 0.25)+ (0.25 x 0.5 x 0.75) + (0.75 x 0.5 x 0.75) + 2 ((0.75 x 0.5 x 0.75) + (0.75 x 0.5 x 0.25) + (0.25 x 0.5 x 0.25) + 3 * (0.75 x 0.5 x 0.25)
= 0 + 0.40625+ 2 x (0.40625)+ 3 x (0.09375)= 1.5

(b)

Probability of selecting any of the three coins i.e. P(A)= P(B)=P(C)=1/3

Given that Probability of head occurring in Coin A, P(H/A)= 0.75

Given that Probability of head occurring in Coin B, P(H/B)= 0.50

Given that Probability of head occurring in Coin B, P(H/C)= 0.25

Given that 3 of the 5 flips result in heads i.e. P (H)= 3/5

We need to find that Coin C was selected i.e. P(C/H)

From Bayes Rule

$$P(C/H) = \frac{P(H/C)\ .\ P(C)}{P(H)} = \frac{0.25 \text{ X } (1/3)}{0.60}$$

$$= 0.1388$$

Solution 6:-

$$P_{x|y}(X=x \mid Y=y) = \sum_{z \in \Omega z} P_{x|yz}(X=x \mid Y=y, Z=z)\ P_{z|y}(Z=z \mid Y=y)$$

This expression holds good.

Let us start with Right hand side of the equation,

$$\sum_{z \in \Omega z} P_{x|yz}(X=x \mid Y=y, Z=z)\ P_{z|y}(Z=z \mid Y=y)$$

$$= \frac{\sum_{z \in \Omega z} P_{x|yz}(X=x, Y=y, Z=z)}{P_{y|z}(Y=y \mid Z=z)}\ .\ P_{z|y}(Z=z \mid Y=y)$$

$$= \frac{\sum_{z \in \Omega z} P_{x|yz}(X=x, Y=y, Z=z)}{P_{z|y}(Z=z \mid Y=y)\ .\ P_{y}(Y=y)}\ .\ P_{z|y}(Z=z \mid Y=y)$$

$$= \frac{\sum_{z \in \Omega z} P_{x|yz}(X=x, Y=y, Z=z)}{P_{y}(Y=y)}$$

$$= \frac{\sum_{z \in \Omega z} P_{x|yz}(X=x, Z=z \mid Y=y)\ .\ P_{y}(Y=y)}{P_{y}(Y=y)}$$

$$= \sum_{z \in \Omega z} P_{x|yz}(X=x, Z=z \mid Y=y)$$

$= P_{x|y}(X=x, \mid Y=y)$ By total probability.

Solution 7:-

A Player wins the game if (6,6) shows up at least once when two dices are rolled 24 times.

Instead of calculating Probability of player winning let's calculate probability of him loosing.

P (Player winning)+ P (Player loosing)= 1

P (Player winning)= 1- P (Player loosing)

$= 1 - (35/36)^{24}$

$= 1 - 0.5085$

$= 0.4915$

Solution 8:-

P (Ω) = 1 since given function satisfies unit measure property.

The given function satisfies σ- additivity for all elements in B(Ω).

Suppose A= (0, 0.2) , B= (0.2, 0.4) , C= (0.4, 0.6) , D= (0.6, 0.8) and E= (0.8, 1)
Then we can say that A, B, C ,D and E are mutually exclusive event

P (A U B U C U D U E)= P (A) +P (B) + P (C) + P(D) + P (E)= 1

Since this set satisfies axiom of probability it can be said that P is a probability function.

Solution 9:-

a)
$P(Z = 1|X = 1) = P(Z=1 |X=1, Y=0)\ P_{Y|X}(Y=0, X=1) + P(Z=1 |X=1, Y=1)\ P_{Y|X}(Y=1, X=1)$

From Figure 1.5 in textbook we have,

$P(Z = 1|X = 1) = d(1-c) + ec$

$= d + ec - dc$

b)

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

$= P(Z=1 \mid X=0, Y=0)\, P_{XY}(X=0, Y=0) + P(Z=1 \mid X=0, Y=1)\, P_{XY}(X=0, Y=1) + P(Z=1 \mid X=1, Y=0)\, P_{XY}(X=1, Y=0) + P(Z=1 \mid X=1, Y=1)\, P_{XY}(X=1, Y=1)$

Since $P(A \cap B) = P(A/B)\, P(B)$

$= d\,(P_{XY}(X=0, Y=0) + P_{XY}(X=1, Y=0)) + e\,(P_{XY}(X=0, Y=1) + P_{XY}(X=0, Y=1))$

$= d\,(1- (P_{XY}(X=0, Y=0) + P_{XY}(X=1, Y=0)) + e\,(P_{XY}(X=0, Y=1) + P_{XY}(X=0, Y=1))$

$= d\,(1- (P_{XY}(X=0, Y=1) + P_{XY}(X=0, Y=1)) + e\,(P_{XY}(X=0, Y=1) + P_{XY}(X=1, Y=1))$

$= d + (e - d)\, P_{XY}(X=0, Y=1) + P_{XY}(X=1, Y=1))$
From table 2 in Figure 1.5 we can say that

$P_{XY}(X=0, Y=1)) = (1 - a)\, b$

$P_{XY}(X=1, Y=1) = ca$

Therefore we can substitute values

$= d + (e - d)\,(b\,(1 - a) + ca)$
$= d + (e - d)\,(b - ba + ca)$
$= d + (e - d)\,(a\,(c - b) + b)$

Solution 10:-

a)

Number of samples= 10	Number of samples= 100	Number of samples= 1000
-0.228175606307	-0.237196881685	-0.00845576861218
0.328644212754	0.0203716989438	0.0356716746178
0.043803377201	0.250811011679	-0.0273885360947
-0.332743049464	-0.0444164898311	0.00099734822636
0.173012250392	-0.142981596548	0.0256970513246

After calculating mean of mean samples for all the cases , we can say that mean tends to reach zero when number of samples increase.

b) When dim=1 and sigma= 10 we calculate average mean in same way as in a)

We can say that mean tends to zero when number of samples increase and when sigma increases mean moves away from zero.

Bonus Question

Solution:-

a) Volume of a hypersphere in n-dimensions of radius a is given by:

$$B_a(n) = \frac{\Pi^{n/2} a^n}{\Gamma(n/2+1)} \qquad \text{from Reference}$$

Where Γ: - Gamma Function

Volume of a hypercube in n-dimension with side 2a is given by

$$C_a(n) = (2a)^n$$

$$R(n) = \frac{\text{Volume of Hypersphere of radius a}}{\text{Volume of Hypercube with side 2a}}$$

$$R(n) = \frac{\dfrac{\Pi^{n/2} a^n}{\Gamma(n/2+1)}}{(2a)^n}$$

$$= \frac{\Pi^{n/2} a^n}{\Gamma(n/2+1)\,(2a)^n}$$

$$= \frac{\Pi^{n/2}}{\Gamma(n/2+1)\,2^n}$$

For High Dimensional spaces take limit n to ∞.

$$\text{Lim}_{n \to \infty} R(n) = \text{Lim}_{n \to \infty} \frac{\Pi^{n/2}}{\Gamma(n/2+1)\,2^n} = 0$$

Volume of sphere is inundated by volume of cube for higher values of n hence we can say that most of the volume of cube is concentrated in corners.

Let D= distance to one of the corners from center of Hypercube of side 2a

For n dimensional

$$D = \sqrt{(2a)^2 + (2a)^2 + \_\_\_\_\_\_\_ + (2a)^n} = \sqrt{n\,(2a)^2} = 2a\sqrt{n}$$

Let d= the perpendicular distance to one of the edges. This is given by:

d= 2a

$$\text{Therefore, } R = \frac{D}{d} = \frac{2a\sqrt{n}}{2a} = \sqrt{n}$$

$$\text{Lim}_{n \to \infty} R = \text{Lim}_{n \to \infty} \sqrt{n} = \infty$$

This implies that distance from the center of the hypercube to one of the corners grows much faster than perpendicular distance to one of the edges. Hence High dimension corners of the hypercube become very long "Spikes".

b) Volume of a Sphere of radius a in d dimension

$V_d(a) = C_d a^d$ where C_d is a constant that depends on d alone and not a.

The fraction of volume of sphere which lies at values of the radius between a-e for 0 < €< a is

$$\frac{V_d(a) - V_d(a-€)}{V_d(a)} = \frac{C_d a^d - C_d (a-€)^d}{C_d a^d}$$

$$= \frac{a^d - (a-€)^d}{a^d}$$

$$= 1 - \frac{(a-€)^d}{a^d}$$

The following table summarizes the value of this fraction for €= 0.01a and for €=0.5a for d € { 2,3,10,100 }:

d -→	2	3	10	100
€= 0.01a	0.0199	0.95617	0.95617	0.633968
€= 0.5a	0.75	0.9990	0.9990	≈ 1

From above as d value increases i.e Fraction of points that lie in thin shell approaches to 1 which means that almost all points are there. To complete this argument just note that the fraction tends to 1 as d approaches infinity since

$$= \lim_{n\to\infty} 1 - \frac{(a - €)^d}{a^d}$$

$$= \lim_{n\to\infty} 1 - \lim_{n\to\infty} \frac{(a - €)^d}{a^d}$$

$$= 1 - 0 = 1$$

Where the second limit is due to the fact that $0<(a - €)/a <1$ and hence, it approaches 0 as d approaches infinity. This proves the result most of the points are in a thin shell, but only for spaces in high dimensions.

Note:- Question 4 not attempted.

List of References

1. Introduction to Statistical Inference and Its Application with R :- Michael W. Trosset
2. Pattern Recognition and Machine Learning :- Christopher M. Bishop
3. https://en.wikipedia.org/wiki/Volume_of_an_n-ball
4. http://www.math.bas.bg/~jeni/Discrete.pdf
5. http://mark.reid.name/blog/warning-high-dimensions.html