CSE 575: Statistical Machine Learning Assignment #1
Sushant Trivedi (1213366971)

PROBABLITY

1. **Independent and disjoint**
   - **If A and B are INDEPENDENT, and P(A) > 0, P(B) > 0, what is the value of P(A|B)?**

Ans. As A and B are independent events, P(A∩B) = P(A)P(B). Hence,

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

- **If A and B are DISJOINT events, and P(A) > 0, P(B) > 0, what is the value of P(A|B)?**

Ans. As A and B are disjoint events, P (A∩B) = 0. Hence, $P\left(\frac{A}{B}\right) = 0$

2. Suppose X is a random variable with the following PDF:

$$\text{f(x)} = \begin{cases} \mathrm{c}(3 + \mathrm{x}^3), \text{for } 0 \leq \mathrm{x} \leq 2 \\ 0, otherwise \end{cases}$$

- **What is the value of c ?**

Ans. The summation of the PDF over Real Space is going to be 1. Thus we integrate it using those limits. But since PDF exists only (0,2) , these would be integration bounds.

$$\int_0^2 3c + cx^3 = 3cx + \frac{cx^4}{4} = 1$$

⇒ 6c + 4c = 1
⇒ c = 0.1

- **What is the value of P(X<=2) ?**

Ans.

$$P(X \leq 2) = \int_{-\infty}^{2} 0.1(x^3 + 3)dx \ = 0.1\left(\frac{x^4}{4} + 3x\right) = \frac{4+6}{10} = 1$$

3. Suppose you are getting tested for a heart condition. The probability that the test comes out with accurate results is 0.98 if the patient has the heart condition. However, the probability that the test is falsely positive is 0.0004. Suppose that on average 1 in 1000 people have this heart condition

- **Can you partition the sample space into 2 events? What are they?**

Ans. Yes, There are two events are follows :

Event A -> Whether the tests is correct or not
Event B -> Whether there is heart condition or not

- **What is the probability that you actually have the disease if the test comes back positive?**

$$P\left(test = \frac{yes}{heart_{condition}}\right) = 0.98$$

$$P\left(\frac{yes}{no_{heart_{condition}}}\right) = 0.0004$$

$$P(heart\ condition) = 0.001$$

$$P\left(\frac{heart_{condition}}{yes}\right) = \frac{0.98 * 0.001}{9.8 * 10^{-4} + (4 * 10^{-4} * 0.999)} = 0.7103$$

MAXIMUM LIKELIHOOD ESTIMATION

1. **Given a set of i.i.d samples x1, x2, · · · , xn ≤ θ following the uniform distribution Uniform(0, θ). Find the maximum likelihood estimation of the parameter θ.**

Ans. The likelihood for a single RV is as follows

$$f(x) = \begin{cases} \frac{1}{\theta}, for\ 0 \le x \le \theta \\ 0\ ,\ otherwise \end{cases}$$

Thus the total likelihood function is as follows:

$$L(\theta) = \begin{cases} \frac{1}{\theta^n}\ ,\ \ for\ 0 \le x_i \le \theta\ \ (\mathrm{i} = 1,2,3 \ldots) \\ 0, \qquad Otherwise \end{cases}$$

Inorder to find MLE, we need to take log of L(θ) and then differentiate it to get:

$$\frac{d(\ln(L(\theta)))}{d\theta} = -\frac{n}{\theta} = 0$$

This is a decreasing function which would give a MLE estimate θ such that θ ≥ xi for i = 1, · · · , n. This value is θ = max(x1, · · · , xn).

2. Suppose that X is a discrete random variable with the following probability mass function: P(X = 0) = 3θ/4; P(X = 1) = θ/4; P(X = 2) = 2(1 − θ)/3; P(X = 3) = (1 − θ)/3 where θ ∈ [0, 1]. We have the following 10 i.i.d samples taken from such a distribution: (3,0,2,1,3,2,1,0,2,1). What is the maximum likelihood estimation of the parameter θ?

Ans. For sample (3,0,2,1,3,2,1,0,2,1), we get

$$L(\theta) = P(x = 3) * P(x = 0) * P(x = 2) * P(x = 1) * P(x = 3) * P(x = 2) * P(x = 1) * P(x = 0) * P(x = 2) * P(x = 1)$$

Simplifying and putting values, we get

$$\mathrm{L}(\theta) = \prod_{i=1}^{n} P(x_i) = \left(\frac{3\theta}{4}\right)^2 \left(\frac{\theta}{4}\right)^3 \left(\frac{2(1-\theta)}{3}\right)^3 \left(\frac{1-\theta}{3}\right)^2$$

Taking Log we get

$$= 2\left(log\frac{3}{4} + log\theta\right) + 3\left(log\frac{1}{4} + \log\theta\right) + 3\left(log\frac{2}{3} + \log(1-\theta)\right) + 2\left(log\frac{1}{3} + \log(1-\theta)\right)$$

Taking differential and simplifying we get,

$$\frac{d(ln(L(\theta)))}{d\theta} = \frac{5}{\theta} - \frac{5}{1-\theta} = 0$$

Simplifying we get θ = 0.5
Thus MLE of θ = 0.5

BAYES CLASSFIER

3. **CONTINUOUS** - We want to build a Bayes Classifier for a binary classification task (y = 1 or y = 2) with a 1-dimensional input feature (x). We know the following quantities: (1) P(y = 1) = 0:8; (2) P(x|y = 1) = 0:5 for 0 <= x <= 2 and P(x|y = 1) = 0 otherwise; and (3) P(x|y = 2) = 0:5 for 0 <= x <= 4 and P(x|y = 2) = 0 otherwise.

- **What is the prior of the class label y = 2?**

Ans. P(Y=2) = 1 – 0.8 = 0.2

- **What is P(Y=1/X) ?**

Ans.

$$P\left(\frac{y=1}{x}\right) = \frac{P\left(\frac{x}{y=1}\right) * P(y=1)}{P(x)}$$

For x ∈(0,2) ,

$$P\left(\frac{y=1}{x}\right) = \frac{\frac{1}{2} * 0.8}{\left(\frac{1}{2} * 0.8\right) + \left(\frac{1}{4} * 0.2\right)} = \frac{8}{9}$$

$$P\left(\frac{y=2}{x}\right) = 1 - \frac{8}{9} = \frac{1}{9}$$

For x ∈(2,4) ,

$$P\left(\frac{y=1}{x}\right) = \frac{0 * 0.8}{(0 * 0.8) + \left(\frac{1}{4} * 0.2\right)} = 0$$

$$P\left(\frac{y=2}{x}\right) = 1$$

For all else,

$$P\left(\frac{y=1}{x}\right) and\ P\left(\frac{y=2}{x}\right)\ are\ undefined$$

- **For x = 1, what is class label your classifier will assign? What is the risk of this decision?**

Ans.

Looking from P(y=1/x) and P(y=2/x) values calculated above, then
We see that P(y=1/x=1) = 8/9 and P(y=2/x=1) = 1/9
Thus the label will assign "y=1" as the label and risk = 1/9

- **What is the decision boundary of your Bayes classifier?**

Ans.

As 0<x<2, we classify as y=1
And 2<x<4, we classify as y=2
Thus, the decision boundary is at x=2

- **What is the Bayes error of your Bayes classifier?**

Ans.

$\in$ = area under the 0<x<2 in the $\Pi\in$ curve
Thus the bayes error = 0.05*2 = 0.1

4. **DISCRETE BAYES CLASSIFIER**

- **What is the prior of the class label y = 2?**

Ans.

P(y=2) = 1 – 0.6 = 0.4

- **What is $P(Y=1/x_1,x_2)$?**

Ans.

$\pi_1 p_1$

	X1 = 0	X1= 1
X2 = 0	0.12	0.06
X2 = 1	0.24	0.18

$\pi_2 p_2$

	X1 = 0	X1= 1
X2 = 0	0.04	0.12
X2 = 1	0.08	0.16

$$P\left(\frac{y=1}{x_1,\ x_2}\right) = \frac{P\left(\frac{x_1,\ x_2}{y=1}\right) * P(y=1)}{P(x_1,\ x_2)}$$

- CASE1: X1 = 0, X2 = 0

$$\mathrm{P}(\frac{y=1}{\mathrm{x1}=0;\mathrm{x2}=0}) = \frac{P(x_1=0,x_2=0|y=1)P(y=1)}{P(x_1=0,x_2=0|y=1)P(y=1)+\ P(x_1=0,x_2=0|y=2)P(y=2)}$$

$$= \frac{0.2\times0.6}{(0.2\times0.6)+(0.4\times0.1)} = \frac{3}{4}$$

- CASE2: X1 = 0, X2 = 1

$$\mathrm{P}\ (\frac{y=1}{\mathrm{x1}=0;\mathrm{x2}=1}) = \frac{0.4\times0.6}{(0.4\times0.6)+(0.4\times0.2)} = \frac{3}{4}$$

- CASE1: X1 = 1, X2 = 0

$$\mathrm{P}\left(\frac{y=1}{\mathrm{x1}=1;\mathrm{x2}=0}\right) = \frac{0.1\times0.6}{(0.1\times0.6)+(0.3\times0.4)} = \frac{1}{3}$$

- CASE1: X1 = 1, X2 = 1

$$\mathrm{P}(\frac{y=1}{\mathrm{x1}=1;\mathrm{x2}=1}) = \frac{0.3\times0.6}{(0.3\times0.6)+(0.4\times0.4)} = \frac{9}{17}$$

- **For an example with the following features x1 = 1; x2 = 0, what is class label your classifier will assign? What is the risk of this decision?**

Ans.

$$P\left(\frac{y=1}{x_1=1, x_2=0}\right) = \frac{0.1 \times 0.6}{(0.1 \times 0.6) + (0.3 \times 0.4)} = \frac{1}{3}$$

$$P\left(\frac{y=2}{x_1=1, x_2=0}\right) = \frac{0.3 \times 0.4}{(0.1 \times 0.6) + (0.3 \times 0.4)} = \frac{2}{3}$$

Therefore the label classified will be **"y=2" and the risk = 1/3**

- **What is the decision boundary of your Bayes classifier?**

Ans.

Comparing the $\pi_1 p_1$table with $\pi_2 p_2$, we get the following classification labels estimates

	X1 = 0	X1 = 1
X2 = 0	Y=1	Y=2
X2 = 1	Y=1	Y=1

From above we get the decision boundary limits

- **What is the Bayes error of your Bayes classifier?**

Ans.

We get the values from table above

Bayes error = min ($\pi_1 p_1$, $\pi_2 p_2$) = 0.04 + 0.06 + 0.08 + 0.16 = 0.34

Input Feature X

Class Label Y

Sky	Temp	Humid	Wind	Water	Forest	Enjoy?
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Rainy	Cold	High	Mild	Warm	Same	No
Sunny	Warm	High	Mild	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes
Sunny	Cold	Normal	Mild	Warm	Same	Yes
Rainy	Cold	Normal	Mild	Cool	Change	No

5. **NAÏVE BAYES**

- How many independent parameters are there in your Naive Bayes classifier? What are they? ONLY list the independent parameters. Justify your answer.

Ans.

For Binary classifier,

No of independent parameters = 2N+1

For this case N = no of features = 6,
Therefore,

No of independent parameters = 2N+1 = 13

PS: For these look below where their estimations have been tabulated

- What are your estimations for these parameters?

Ans.

$$\boldsymbol{MLE\ Estimations - P\left(\frac{x = x_i}{y = y_i}\right) = \frac{count(\,x_i\ \cap\ y_i)}{count(y_i)}}$$

1	**P(y = yes) = 4/7**
2	$\boldsymbol{P\left(\frac{sky = sunny}{y =\ yes}\right) = 1}$
3	$\boldsymbol{P\left(\frac{sky = sunny}{y =\ no}\right) = 0}$
4	$\boldsymbol{P\left(\frac{temp = warm}{y =\ yes}\right) = 3/4}$
5	$\boldsymbol{P\left(\frac{temp = warm}{y =\ no}\right) = 0}$
6	$\boldsymbol{P\left(\frac{humid = normal}{y =\ yes}\right) = 1/2}$
7	$\boldsymbol{P\left(\frac{humid = normal}{y =\ no}\right) = 1/3}$
8	$\boldsymbol{P\left(\frac{wind = strong}{y =\ yes}\right) = 1/2}$
9	$\boldsymbol{P\left(\frac{wind = strong}{y =\ no}\right) = 1/3}$
10	$\boldsymbol{P\left(\frac{water = warm}{y =\ yes}\right) = 3/4}$
11	$\boldsymbol{P\left(\frac{water = warm}{y =\ no}\right) = 2/3}$
12	$\boldsymbol{P\left(\frac{forest = same}{y =\ yes}\right) = 3/4}$
13	$\boldsymbol{P\left(\frac{forest = same}{y =\ no}\right) = 1/3}$

- Now, given a new (test) example x = (sunny; cold; high; strong; cool; same), what is P(y = 1jx)? Which class label will the naive Bayes classifer assign to this example? Justify your answer.

Ans.

$$P(enjoyspt = yes|x)$$

$$P\left(\frac{enjoyspt = yes}{x}\right) = \frac{\begin{array}{c} P(sky = sunny|enjoyspt = yes)P(temp = cold|enjoyspt = yes) \\ P(humid = strong|enjoyspt = yes)P(wind = strong|enjoyspt = yes) \\ P(water = cool|enjoyspt = yes)P(forest = same|enjoyspt = yes) \\ P(enjoyspt = yes) \end{array}}{P(x)}$$

$$= \frac{(1 \times 1/4 \times 1/2 \times 1/2 \times 1/4 \times 3/4 \times 4/7)}{\left(1 \times \frac{1}{4} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{4} \times \frac{3}{4} \times \frac{4}{7}\right) + 0} = 1$$

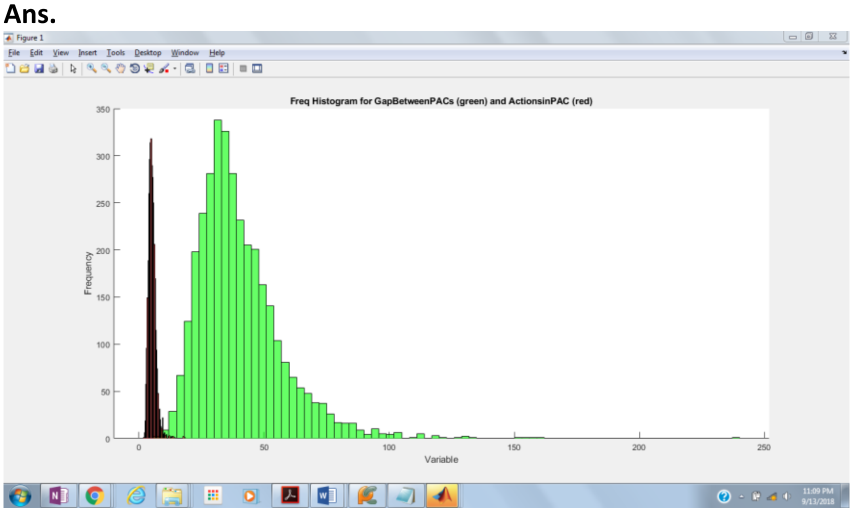
$$P\left(\frac{enjoyspt = no}{x}\right) = 0$$

Thus the estimated label is y = yes.

EXPLORING DATA

- **Plot the frequency histogram on 'GapBetweenPACs' and 'ActionsInPAC' attributes, respectively. What is your observation regarding the figures?**

Ans.



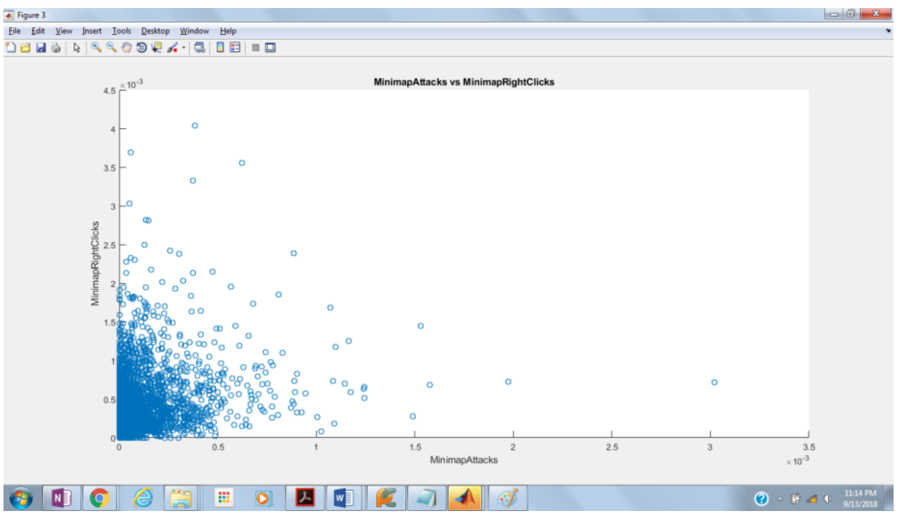
Looking at the two frequency histogram, we can observe that the variance of one of the curve – red one (actioninPAC) – is much less than the green one (GapbetweenPACs).

- **Plot the scatter figure for the following pairs of features: (1) UniqueHotkeys vs. AssignToHotkeys, and (2) MinimapAttacks vs. MinimapRightClicks, respectively.**

Ans.

UniqueHotkeys vs. AssignToHotkeys

MinimapAttacks vs. MinimapRightClicks



- **For all the attributes from the 6-th attribute to the 20-th attribute, compute the Pearson Correlation Coefficient (PCC) between each two different attributes and save the results as a matrix in a .mat file or in a .txt file (each line of which is separated by spaces). What are the maximum and minimum PCC values and what are corresponding attribute pairs? Plot the scatter figures for these attribute pairs. What is the main difference between these two scatter figures?**

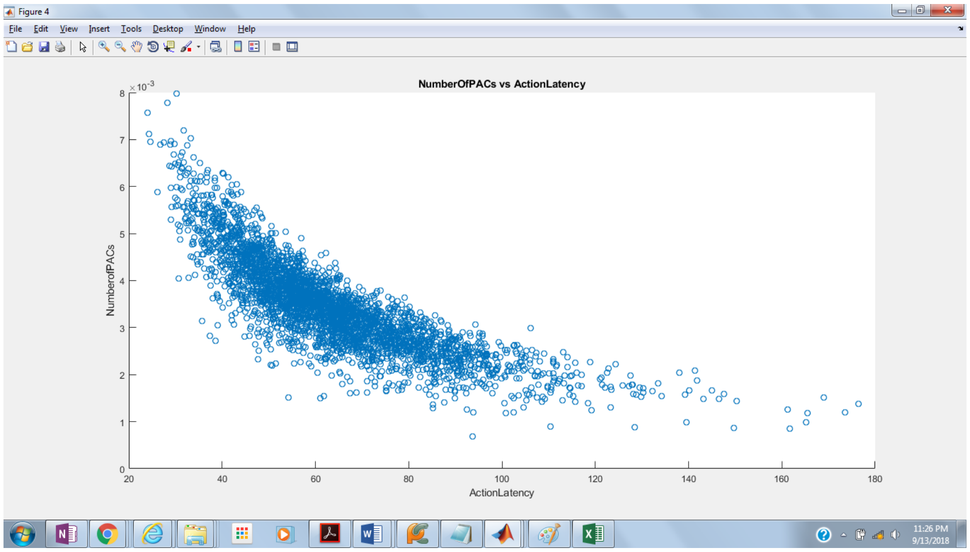
Ans.

Having found the PCC contribution pairs, we said the

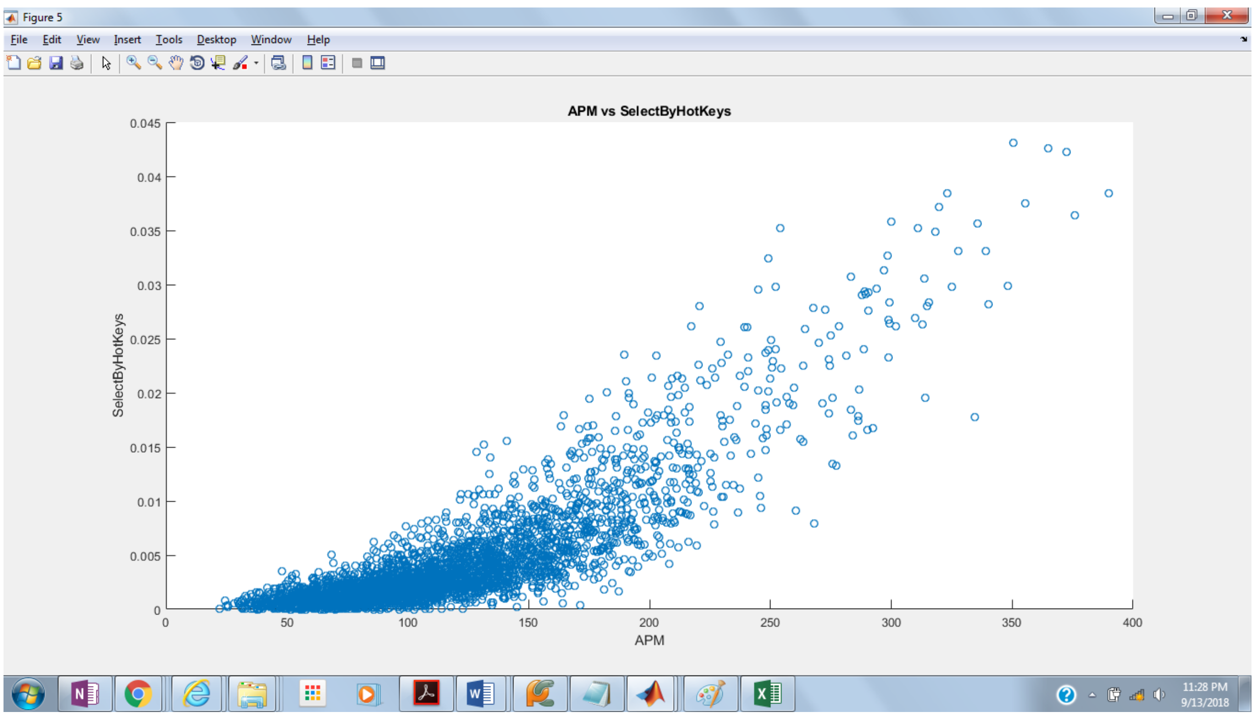
PCC_max = 0.8406 (APM vs SelectByHotkeys)

PCC_min = -0.8203 (NumberOfPACs vs ActionLatency)

<u>MINIMUM CURVE (NumberOfPACs vs ActionLatency)</u>



MAXIMUM CURVE (APM vs SelectByHotkeys)



We see an increasing scatter with a positive slope Maximum Curve showing positive linear dependency.

Whereas we see a decreasing scatter with a negative slope Minimum curve showing inverse linear dependency