ML Assignment-2

Akshala Bhatnagar
2018012

1. a) Principal Component Analysis (PCA) is used for dimensionality reduction. Many times numerous features are available in a dataset and it is not feasible to consider all of them. Suppose we have n features, then we can use feature extraction to create k new features using these n features. PCA is used for feature extraction. It tries to maximize variance and preserves large pairwise distance. New features are generated which are not correlated to each other. Essentially PCA finds the k vectors on which the data can be projected such that the projection error is minimized. Before running PCA the data needs to be normalized so that all features are compared on the same scale.

Covariance matrix $= \frac{1}{m} X^T X$

m: No. of training examples, X: normalized feature matrix

U: Eigenvectors of covariance matrix.
The first k columns of U are considered.

$U_k = U[:k]$

$Z = U_k^T X$

Z: new feature matrix

b) Singular Value Decomposition (SVD) states that a matrix A can be written as $A = U\Sigma V^T$, where U and V are orthogonal matrices with orthonormal eigenvectors which are taken from AA^T and A^TA. Σ is a diagonal matrix, where entries are positive singular values equal to root of positive eigenvalues of AA^T or A^TA, sorted in decreasing order.

$A = U\Sigma V^T$

$V^TV = I$

$U^TU = I$

If A is an m*n matrix then U is m*m, Σ is m*n and V is n*n. U and V are invertible and orthonormal.

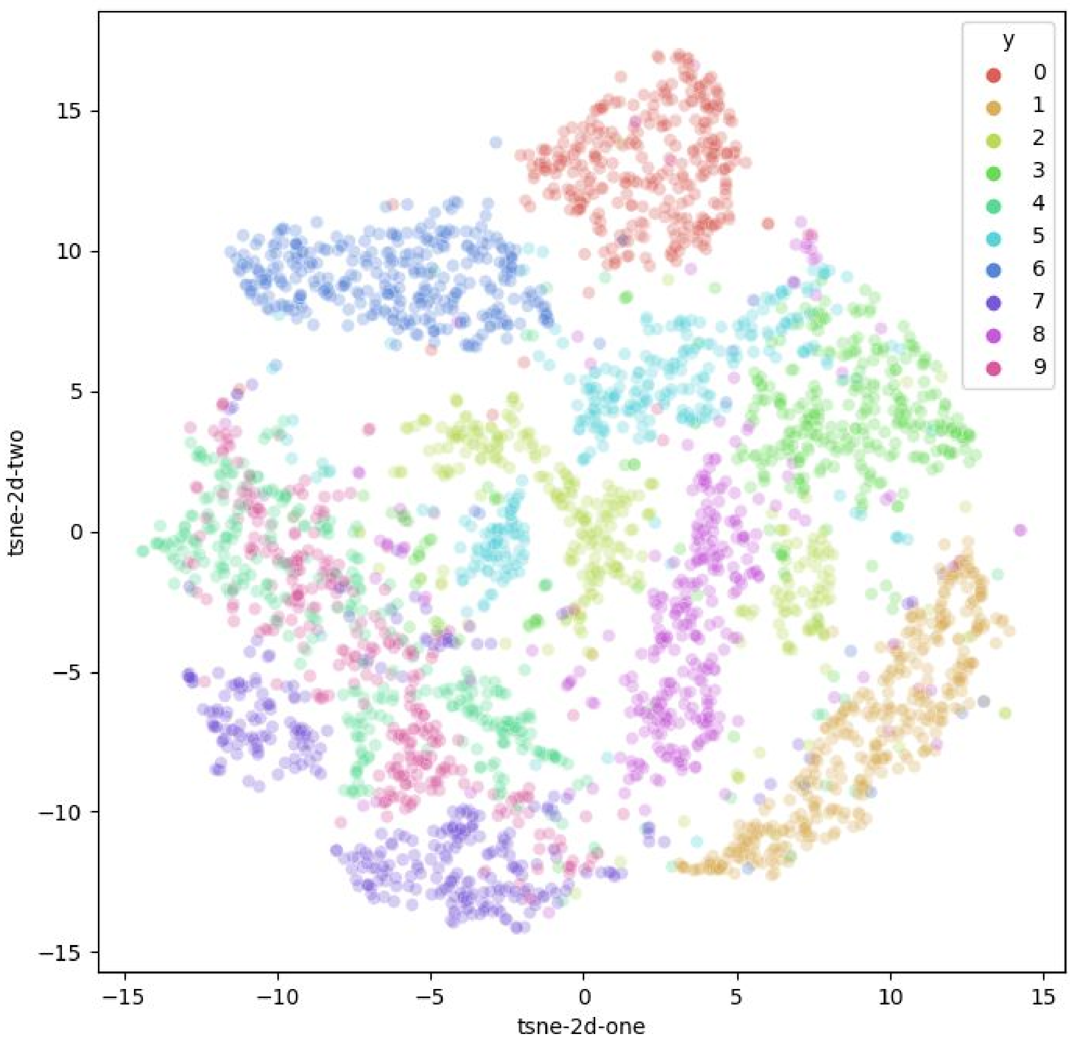
c) t-SNE is used to represent a high dimensional dataset into a low dimensional graph. This is done in a way such that the clustering in the high dimensional plot is preserved. It preserves local similarities or small pairwise distance. A probability distribution using Gaussian distribution is created to capture the relationship in the high dimensional space. Student-t distribution along with gradient descent using a non convex function is used to recreate the relationship in a low dimensional space.

d) Stratified Sampling

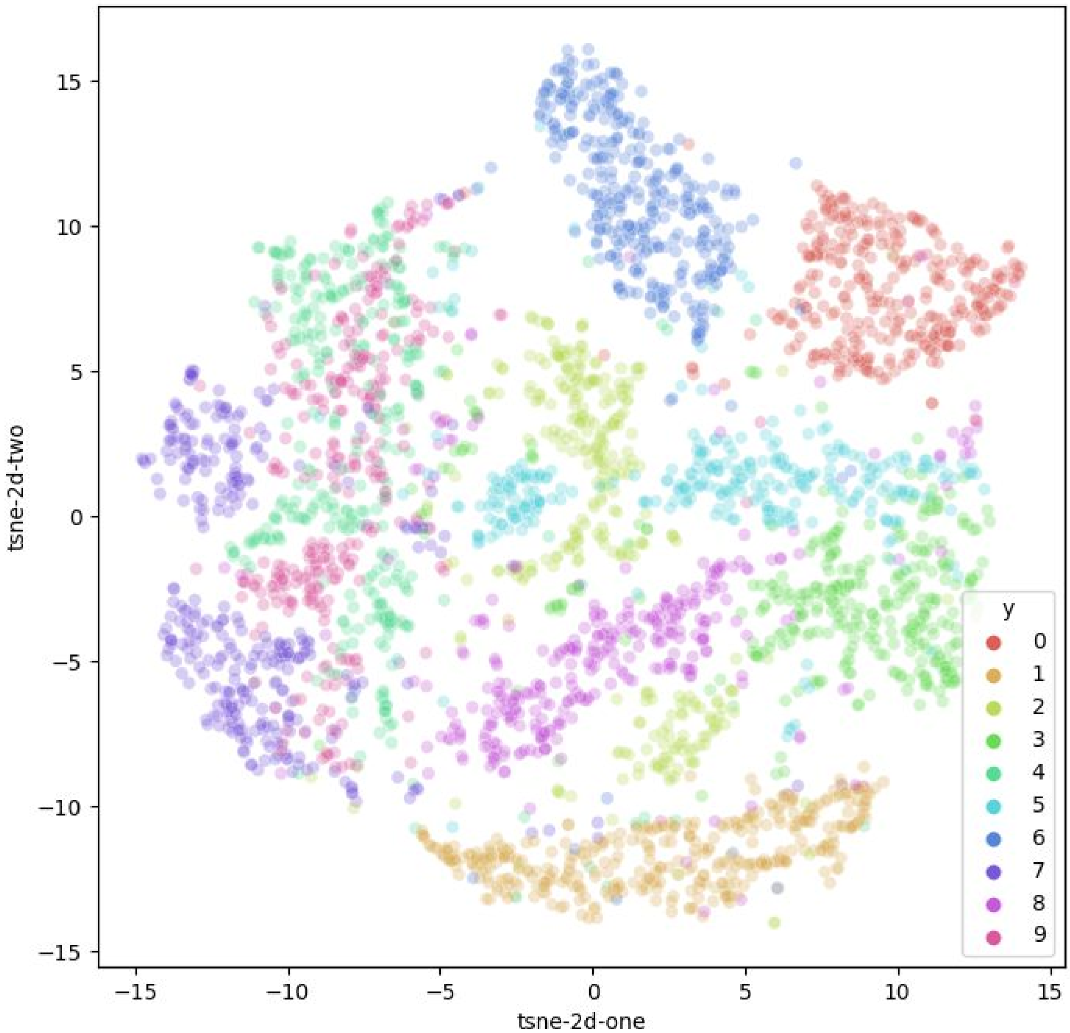
Class	0	1	2	3	4	5	6	7	8	9
Train	0.095238	0.117560	0.093452	0.100893	0.099107	0.094940	0.105060	0.102679	0.097619	0.093452
Test	0.095238	0.117857	0.094048	0.101190	0.098810	0.094048	0.104762	0.102381	0.097619	0.094048

As it can be seen from the table above, The percentage of occurrence of each class in train is almost equal to the percentage of occurrence of the corresponding class in test.

e) Testing accuracy: 85.595%



f) Testing accuracy: 85.238%



g) Both PCA and SVD give similar results. Depending upon the train test split there is only a slight difference between the accuracies obtained using PCA and SVD. The difference between PCA and SVD is that the data first gets centered in PCA and then SVD is used. So normalisation of data is being done in PCA and not in SVD.

2. a) Bias = 9.691681418761275

   Variance = 0.035229276761847714

   b) MSE = 149.03513807206136

   MSE − Bias^2 − Variance = 55.07122007

   MSE = Noise + Variance + Bias^2

   Therefore, Noise = MSE − Bias^2 − Variance

   A value of 55.07122007 for MSE − Bias^2 − Variance, indicates that there is a lot of noise in the data.

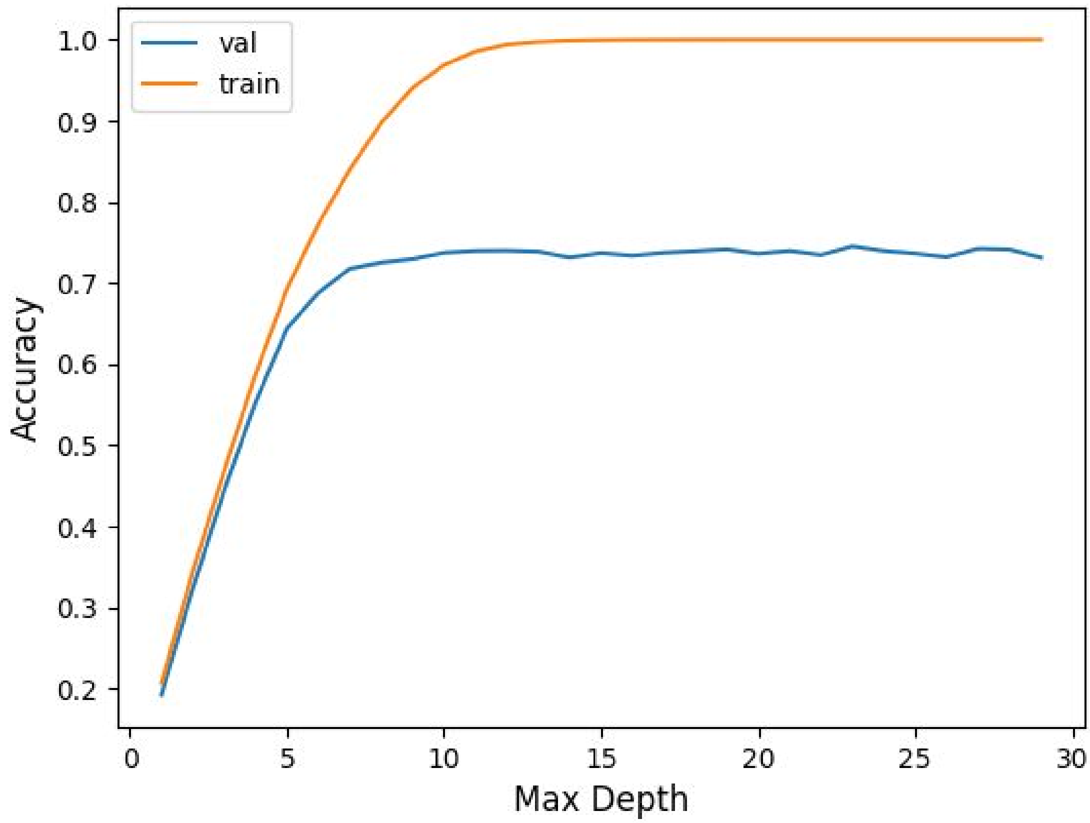
3. a) Decision Tree

Dataset	Optimal Depth	Val accuracy	Train accuracy
A	23	0.745238	1.000000
B	5	0.585417	0.710565

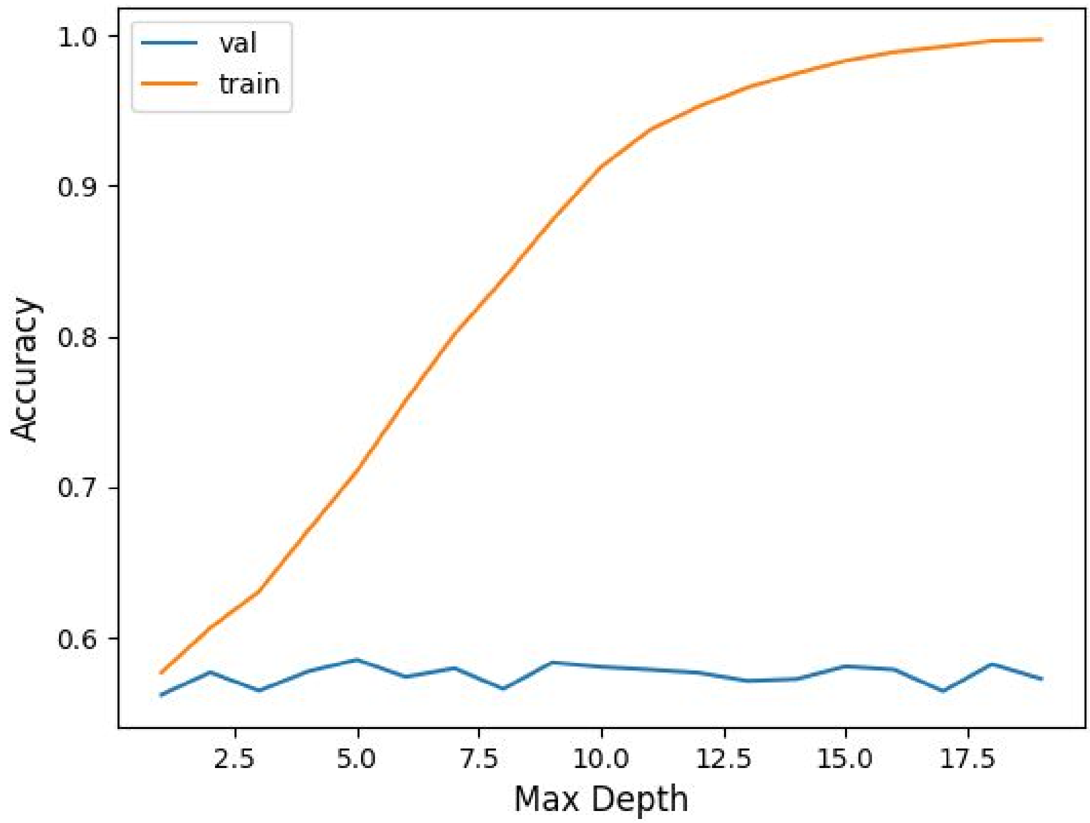
   Gaussian Naive Bayes

Dataset	Val accuracy	Train accuracy
A	0.608035714285714	0.64975198412698
B	0.571428571428571	0.580109126984127

b) <u>Dataset A</u>



Dataset B



For dataset A after tree depth 10 there is not much change in the training and validation accuracies. For dataset B, Validation accuracy remains in the range of 56-60% but the training accuracy increases as the depth increases.

c)

Dataset	Optimal Depth	Test accuracy
A	23	0.735714285714285
B	6	0.573809523809523

d) **Dataset A**

Decision Tree

Accuracy: 0.7202380952380952

Confusion Matrix

	0	1	2	3	4	5	6	7	8	9
0	64	0	2	0	0	1	3	0	1	4
1	0	87	1	0	2	2	3	1	2	1
2	2	1	58	6	1	1	6	1	9	1
3	2	1	5	58	3	6	1	2	7	0
4	1	5	2	5	61	6	3	3	5	4
5	3	1	2	7	1	47	7	0	2	4
6	2	0	1	0	5	3	60	1	2	1
7	2	1	6	2	0	0	2	67	2	3
8	2	3	1	3	2	11	1	0	47	4
9	2	0	1	4	8	3	2	11	5	56

Macro average

Macro Precision: 0.7169765193106444

Macro Recall: 0.7198199237668571

Macro F1 score: 0.7166970125412556

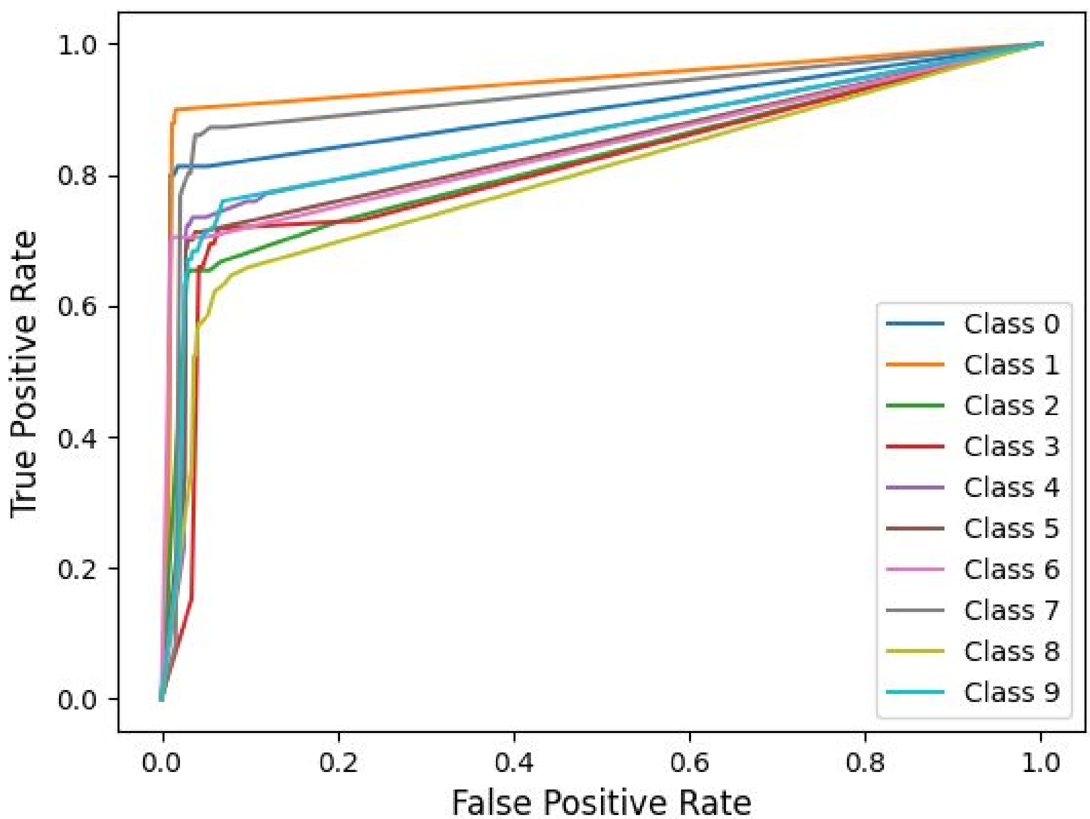
Micro Average

Micro Precision: 0.7202380952380952

Micro Recall: 0.7202380952380952

Micro F1 score: 0.7202380952380952

ROC Curve



Gaussian Naive Bayes

Accuracy: 0.5630952380952381

Confusion Matrix

	0	1	2	3	4	5	6	7	8	9
0	68	0	15	6	1	6	5	1	3	0
1	0	95	2	10	2	4	3	0	22	1
2	1	0	21	0	1	0	2	0	2	0
3	1	1	6	30	0	4	0	1	0	0
4	0	0	0	1	17	0	0	1	0	1
5	0	1	1	0	2	13	1	0	2	0
6	2	0	14	3	11	3	76	0	1	0
7	0	0	1	3	1	2	0	46	0	3
8	2	1	16	19	10	34	1	0	33	0
9	6	1	2	13	38	14	0	37	19	74

Macro average

Macro Precision: 0.5536755152122084

Macro Recall: 0.6466090000409116

Macro F1 score: 0.526416717307849

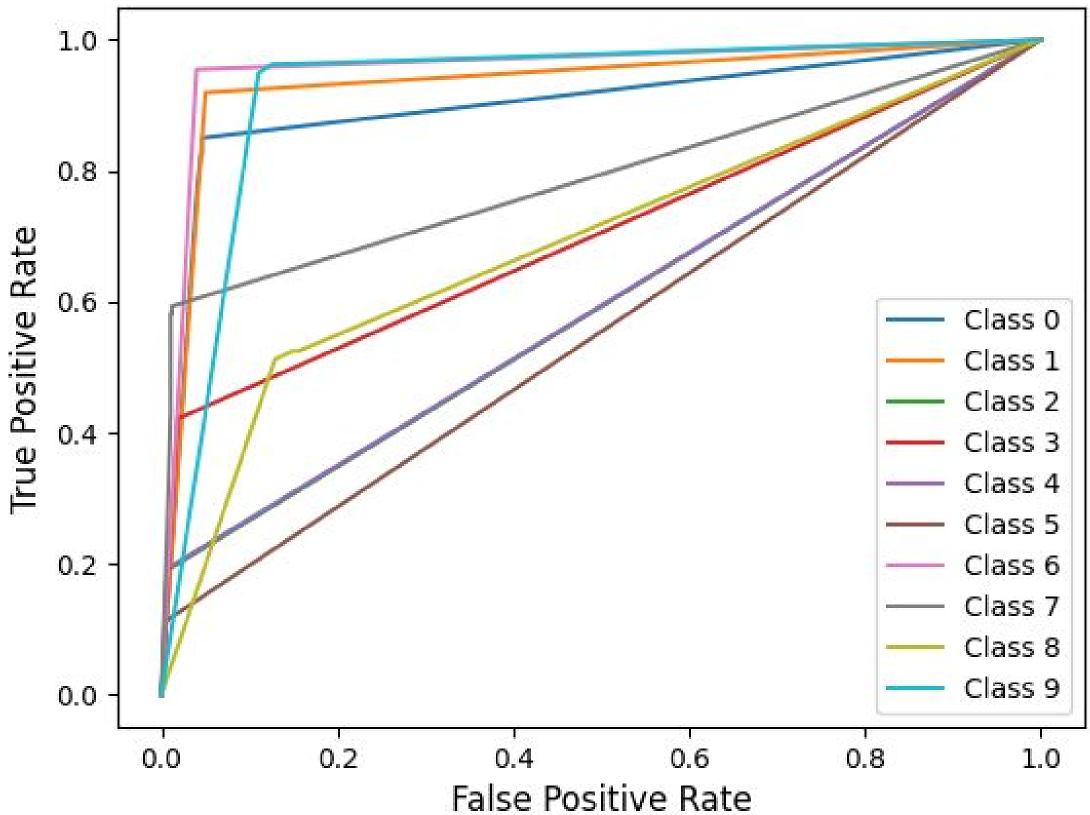
Micro Average

Micro Precision: 0.5630952380952381

Micro Recall: 0.5630952380952381

Micro F1 score: 0.5630952380952381

ROC Curve



Dataset B

Decision Tree

Accuracy: 0.5773809523809523

Confusion Matrix

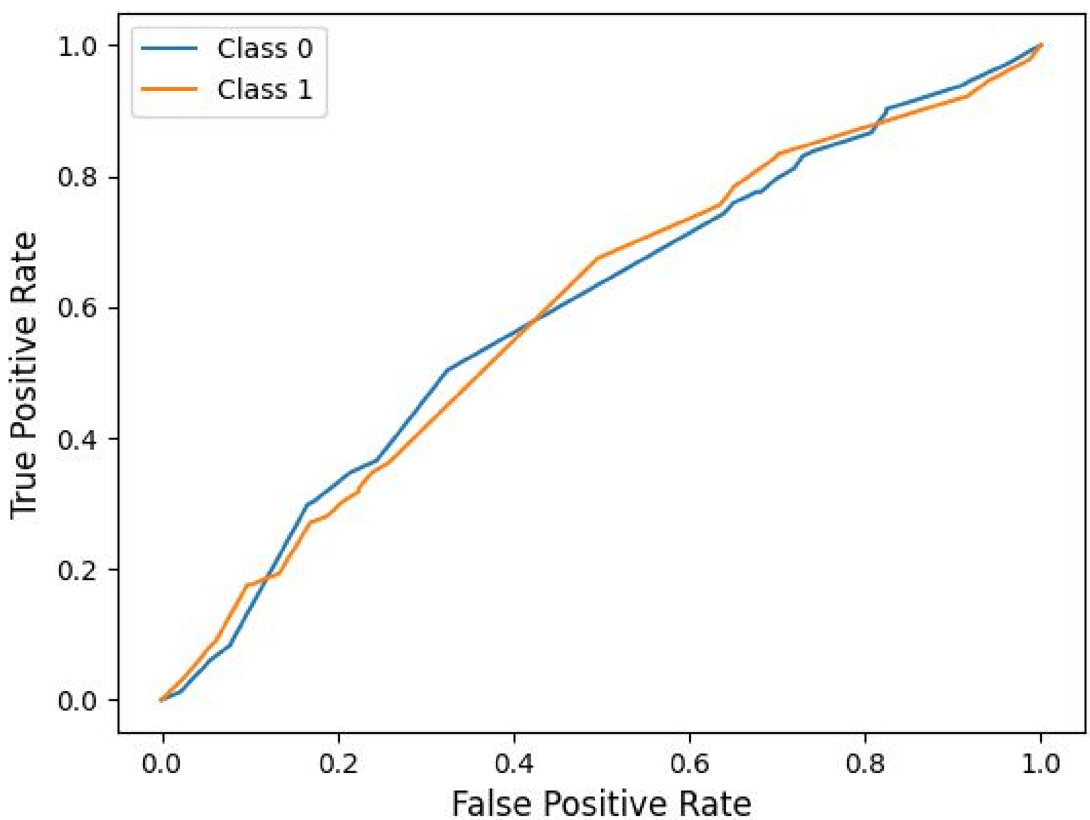
	Actual 0	Actual 1
Predicted 0	188	124
Predicted 1	231	297

Precision: 0.5625

Recall: 0.7054631828978623

F1 score: 0.6259220231822972

ROC Curve



Gaussian Naive Bayes

Accuracy: 0.5571428571428572

Confusion Matrix

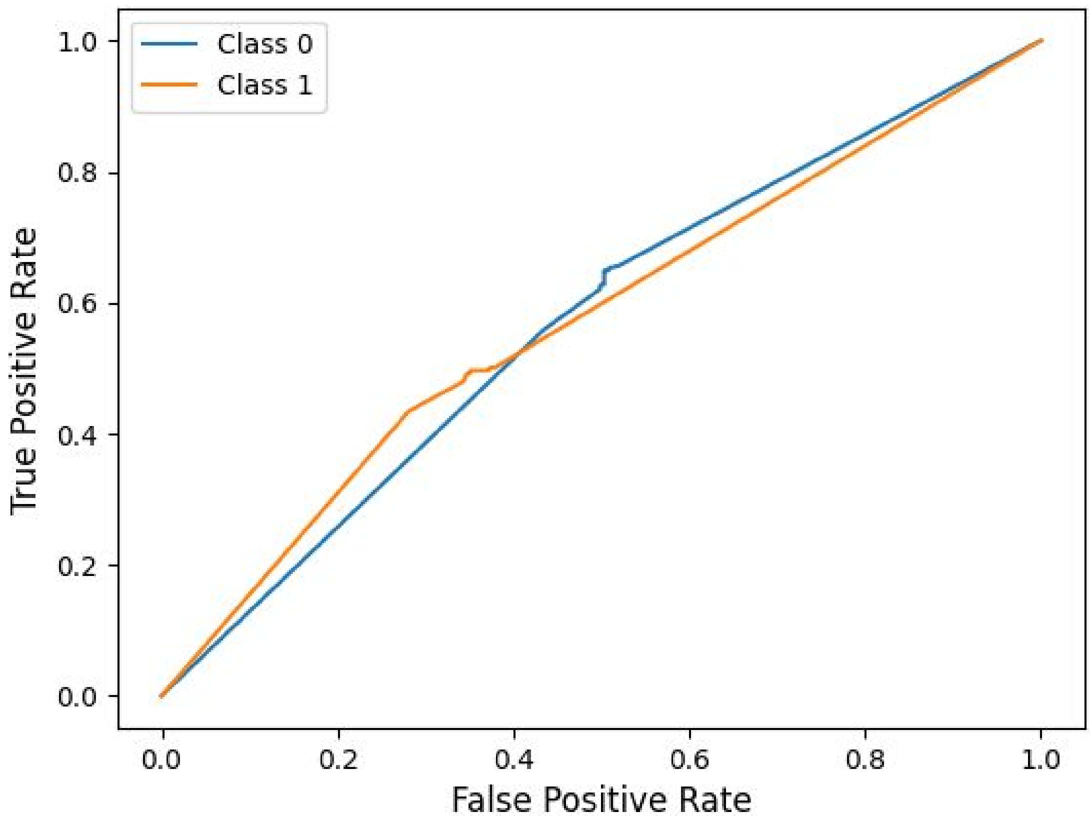
	Actual 0	Actual 1
Predicted 0	264	217
Predicted 1	155	204

Precision: 0.5682451253481894

Recall: 0.4845605700712589

F1 score: 0.5230769230769231

ROC curve



4. Dataset A

   My Gaussian Naive Bayes: 0.49604761904761905

   Sklearn: 0.5642857142857143

   Dataset B:

   My Gaussian Naive Bayes: 0.5452380952380952

   Sklearn: 0.5452380952380952

5. a)

5.a) $H(x) = -\sum_{i=1}^{n} P(x=i) \log_2 P(X=i)$

$N(\text{Playmatch} = \text{yes}) = 9$
$N(\text{Playmatch} = \text{No}) = 5$

$H(\text{Playmatch}) = -\left(\frac{9}{14}\right) \log\left(\frac{9}{14}\right) - \frac{5}{14} \cdot \log\left(\frac{5}{14}\right) = 0.94$

Outlook

Playmatch	Sunny	Overcast	Rain	Total
yes	2	4	3	
No	3	0	2	
	5	4	5	14

H(Playmatch | outlook) = − P(Sunny) [P(Playmatch | sunny) * log P(Playmatch | sunny) + P(~Playmatch | sunny) * log P(~Playmatch | sunny)] − P(overcast) [P(Playmatch | overcast) log P(Playmatch | overcast) + P(~Playmatch | overcast)] − P(rain) [P(Playmatch | rain) log P(Playmatch | rain) + P(~Playmatch | rain) log P(~Playmatch | rain)]

$$= -\frac{5}{14}\left[\frac{2}{5}\log\frac{2}{5} + \frac{3}{5}\log\frac{3}{5}\right] - \frac{4}{14}\left[\frac{4}{4}\log\frac{4}{4} + \frac{0}{4}\log\frac{0}{4}\right]$$
$$- \frac{5}{14}\left[\frac{3}{5}\log\frac{3}{5} + \frac{2}{5}\log\frac{2}{5}\right] = 0.48$$

Climate

Playmatch	Hot	Mild	Cold
Yes	2	4	2
No	2	2	2
	4	6	4

$H(\text{Playmatch}|\text{climate}) = -P(\text{hot})\,[P(\text{Playmatch}|\text{hot}) \log P(\text{Playmatch}|\text{hot}) + P(\sim\text{Playmatch}|\text{hot}) \log P(\sim\text{Playmatch}|\text{hot})] - P(\text{mild})\,[P(\text{Playmatch}|\text{mild}) \log P(\text{Playmatch}|\text{mild})] + P(\sim\text{Playmatch}|\text{mild}) \log P(\sim\text{Playmatch}|\text{mild})] - P(\text{cool})\,[P(\text{Playmatch}|\text{cool}) \log P(\text{playmatch}|\text{cool})] + P(\sim\text{Playmatch}|\text{cool}) \log P(\sim\text{Playmatch}|\text{cool})]$

$$= \frac{-4}{14}\left[\frac{2}{4}\log\frac{2}{4} + \frac{2}{4}\log\frac{2}{4}\right] - \frac{6}{14}\left[\frac{4}{6}\log\frac{4}{6} + \frac{2}{6}\log\frac{2}{6}\right] - \frac{4}{14}\left[\frac{2}{4}\log\frac{2}{4} + \frac{2}{4}\log\frac{2}{4}\right] = 0.668$$

Humidity

Playmatch	High	Normal
Yes	3	6
No	4	1
	7	7

$P(\text{Playmatch}|\text{humidity}) = -P(\text{high})\,[P(\text{Playmatch}|\text{high}) \log P(\text{Playmatch}|\text{high}) + P(\sim\text{Playmatch}|\text{high}) \log P(\sim\text{Playmatch}|\text{high})] - P(\text{Normal})\,[P(\text{Playmatch}|\text{normal}) \log P(\text{Playmatch}|\text{Normal}) + P(\sim\text{Playmatch}|\text{normal}) \log P(\sim\text{Playmatch}|\text{Normal})]$

$$= -\frac{7}{14}\left[\frac{3}{7}\log\frac{3}{7} + \frac{4}{7}\log\frac{4}{7}\right] - \frac{7}{14}\left[\frac{6}{7}\log\frac{6}{7} + \frac{1}{7}\log\frac{1}{7}\right]$$

$= 0.547$

Wind

Playmatch	Weak	strong
Yes	6	3
No	2	3
	8	6

$P(\text{Playmatch}|\text{wind}) = -P(\text{weak})\,[P(\text{Playmatch}|\text{weak}) \log P(\text{Playmatch}|\text{weak})] + P(\sim\text{Playmatch}|\text{weak})$

$\log P(\neg Playmatch \mid weak)] - P(strong)[P(Playmatch \mid strong) \log P(Playmatch \mid strong) + P(\neg Playmatch \mid strong) \log P(\neg Playmatch \mid strong)]$

$$= -\frac{8}{14}\left[\frac{6}{8}\log\frac{6}{8} + \frac{2}{8}\log\frac{2}{8}\right] - \frac{6}{14}\left[\frac{3}{6}\log\frac{3}{6} + \frac{3}{6}\log\frac{3}{6}\right]$$

$= 0.618$

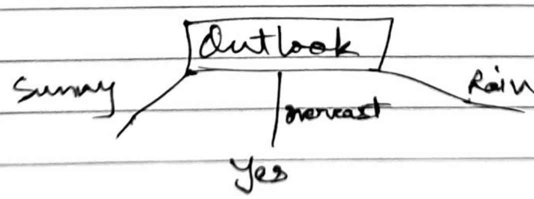
IG(Playmatch, outlook) = H(Playmatch) − H(Playmatch | outlook)
= 0.46

IG(Playmatch, climate) = H(Playmatch) − H(Playmatch | climate)
= 0.272

IG(Playmatch, humidity) = H(Playmatch) − H(Playmatch | humidity)
= 0.393

IG(Playmatch, wind) = H(Playmatch) − H(Playmatch | wind)
= 0.322

Largest information gain is for Outlook



Outlook: Sunny

Climate	Humidity	Wind	Playmatch
Hot	High	Weak	No
Hot	High	Strong	No
Mild	High	Weak	No
Cool	Normal	Weak	Yes
Mild	Normal	Strong	Yes

Outlook: Overcast

Climate	Humidity	Wind	Playmatch
Hot	High	Weak	Yes
Cool	Normal	Strong	Yes
Mild	High	Strong	Yes
Hot	Normal	Weak	Yes

Outlook: Rain

Climate	Humidity	Wind	Playmatch
Mild	High	Weak	No
Cool	Normal	Weak	Yes
Cool	Normal	Strong	No
Mild	Normal	Weak	Yes
Mild	High	Strong	No

$$H(\text{sunny}) = -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5} = \cancel{0.92}\ 0.673$$

Climate

Playmatch	Hot	Mild	Cool
yes	0	1	1
No	2	1	0
	2	7	1

$$H(\text{sunny}|\text{climate}) = -\frac{2}{5}\left(\frac{2}{2}\log\frac{2}{2} + \frac{0}{2}\log\frac{0}{2}\right) - \frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) - \frac{1}{5}\left(1\log 1 + 0\log 1\right)$$

$$= 0.12$$

~~Play~~ Humidity

Playmatch	High	Normal
Yes	0	2
No	3	0
	3	2

$$H(\text{sunny}|\text{humidity}) = -\frac{3}{5}\log\left(\frac{3}{3}\log\frac{3}{3}+\frac{0}{3}\log\frac{0}{3}\right)-\frac{2}{5}\left(\frac{2}{2}\log\frac{2}{2}+\frac{0}{2}\log\frac{0}{2}\right)$$

$$\approx 0$$

Wind

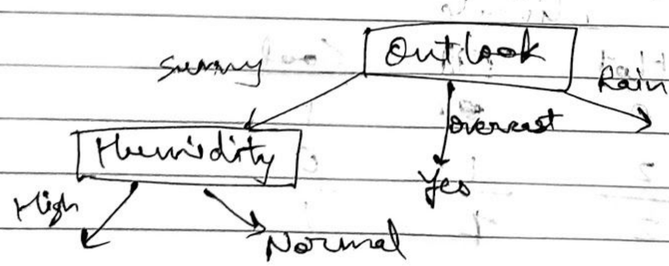
Playmatch	Weak	strong
Yes	1	1
No	2	1
	3	2

$$H(\text{sunny}|\text{wind}) = -\frac{3}{5}\log\left(\frac{2}{3}\log\frac{2}{3}+\frac{1}{3}\log\frac{1}{3}\right)$$

$$-\frac{2}{5}\log\left(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right)$$

$$\approx 0.659$$

IG (Sunny, climate) = 0.553
IG (Sunny, humidity) = 0.673
IG (sunny, wind) = 0.014
Largest Information gain is for humidity.



$H(\text{rain}) = -\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5} = 0.673$

Climate

Playmatch	Hot	Mild	Cool
Yes	0	1	1
No	$\frac{0}{0}$	$\frac{2}{3}$	$\frac{1}{2}$

$$H(\text{rain}|\text{climate}) = -\frac{0}{5}\left(\frac{0}{0}\log\frac{0}{0} + \frac{0}{0}\log\frac{0}{0}\right) - \frac{3}{5}\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right)$$

$$-\frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right)$$

$= 0.659$

Humidity

Playmatch	High	Normal
Yes	0	2
No	$\frac{2}{2}$	$\frac{1}{3}$

$$H(\text{rain}|\text{humidity}) = -\frac{2}{5}\left(\frac{0}{2}\log\frac{0}{2} + \frac{2}{2}\log\frac{2}{2}\right) - \frac{3}{5}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)$$

$= 0.382$

Wind

Playmatch	Weak	Strong
Yes	2	0
No	$\frac{1}{3}$	$\frac{2}{2}$

$$H(\text{rain}|\text{wind}) = -\frac{3}{5}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) - \frac{2}{5}\left(\frac{0}{2}\log\frac{0}{2} + \frac{2}{2}\log\frac{2}{2}\right)$$

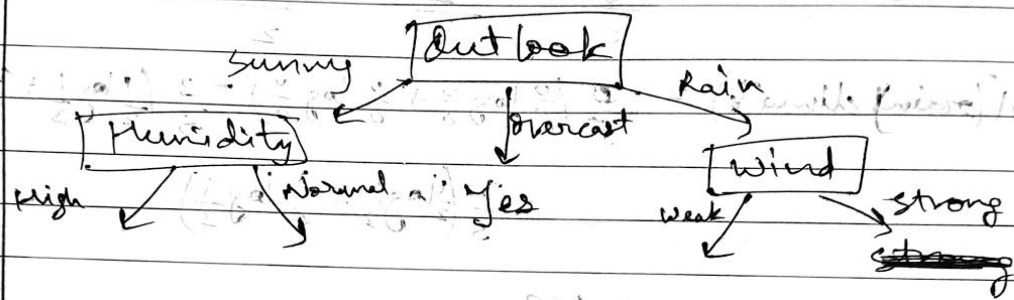
$= 0.382$

IG(rain, climate) = ~~0.014~~ 0.14

IG(rain, humidity) = 0.291

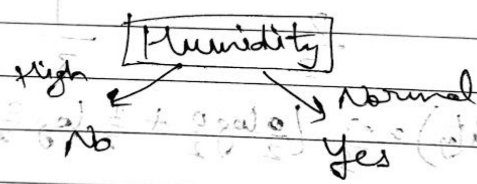
IG(rain, wind) = 0.291

Largest IG is for humidity and wind but humidity has already been used.



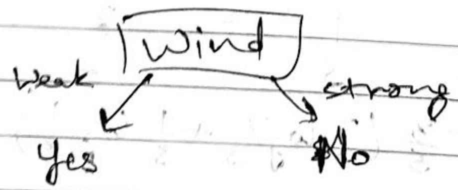
Sunny, Humidity

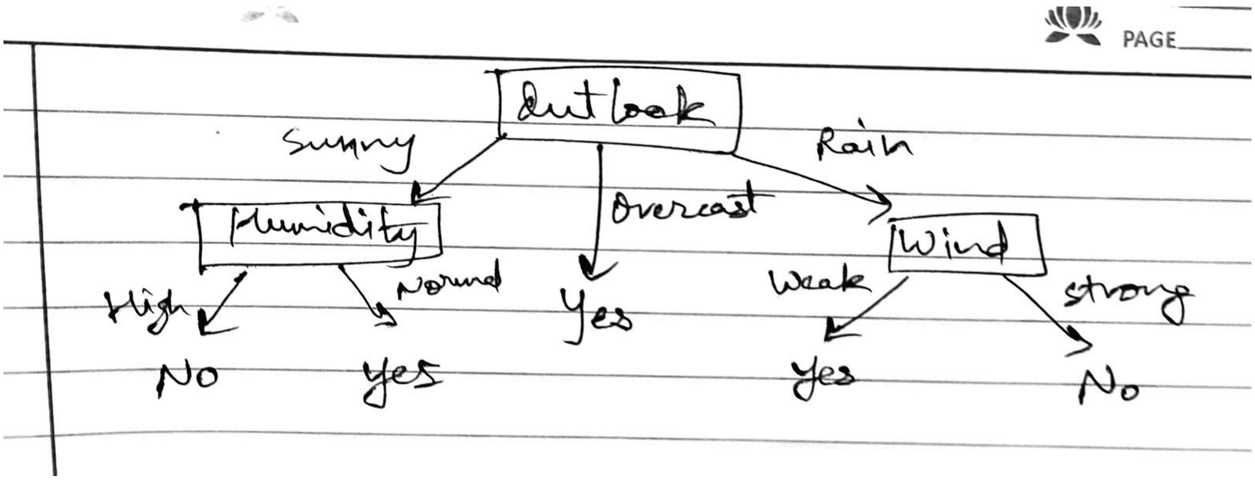
Playmatch	High	Normal
Yes	0	2
No	3	0



Rain, wind

Playmatch	Weak	strong
Yes	2	0
No	1	2
	3	2





Outlook	Climate	Humidity	Wind	Playmatch	Predicted
Sunny	Hot	High	Weak	No	No
Sunny	Hot	High	Strong	No	No
Overcast	Hot	High	Weak	Yes	Yes
Rain	Mild	High	Weak	Yes	Yes
Rain	Cool	Normal	Weak	Yes	Yes
Rain	Cool	Normal	Strong	No	No
Overcast	Cool	Normal	Strong	Yes	Yes
Sunny	Mild	High	Weak	No	No
Sunny	Cool	Normal	Weak	Yes	Yes
Rain	Mild	Normal	Weak	Yes	No
Sunny	Mild	Normal	Strong	Yes	Yes
Overcast	Mild	High	Strong	Yes	Yes
Overcast	Hot	Normal	Weak	Yes	Yes
Rain	Mild	High	Strong	No	No

Accuracy = (13/14) * 100 = 92.86

b)

Day	Outlook	Climate	Humidity	Wind	PlayMatch
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D5	Rain	Cool	Normal	Weak	Yes
D7	Overcast	Cool	Normal	Strong	Yes
D9	Sunny	Cool	Normal	Weak	Yes

Yes,the attribute Climate can be included in the decision tree. If the above subset from the given table is taken then Climate can be included in the learned decision tree.

The calculations and the tree are given below.

b) $H(\text{playmatch}) = -\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5} = 0.673$

$$H(\text{Playmatch} \mid \text{outlook}) = -\frac{1}{5}\left(\frac{1}{1}\log\frac{1}{1} + \frac{0}{1}\log\frac{0}{1}\right) - \frac{3}{5}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)$$
$$- \frac{1}{5}\left(\frac{1}{1}\log\frac{1}{1} + \frac{0}{1}\log\frac{0}{1}\right) = 0.3819$$

$$H(\text{Playmatch} \mid \text{climate}) = -\frac{2}{5}\left(\frac{2}{2}\log\frac{2}{2} + \frac{0}{2}\log\frac{0}{2}\right) - \frac{3}{5}\left(\frac{3}{3}\log\frac{3}{3} + \frac{0}{3}\log\frac{0}{3}\right)$$
$$= 0$$

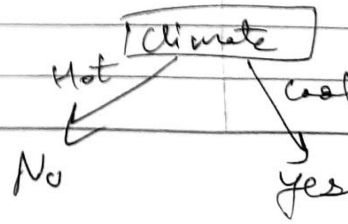
$$H(\text{Playmatch} \mid \text{Humidity}) = -\frac{2}{5}\left(\frac{2}{2}\log\frac{2}{2} + \frac{0}{2}\log\frac{0}{2}\right) - \frac{3}{5}\left(\frac{3}{3}\log\frac{3}{3} + \frac{0}{3}\log\frac{0}{3}\right)$$
$$= 0$$

$$H(\text{Playmatch} \mid \text{wind}) = -\frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) - \frac{3}{5}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)$$
$$= 0.5$$

$IG(\text{Playmatch, outlook}) = 0.291$
$IG(\text{Playmatch, climate}) = 0.673$
$IG(\text{Playmatch, humidity}) = 0.673$
$IG(\text{Playmatch, wind}) = 0.273$

IG is highest for climate and humidity. If we pick climate then,

	Climate	
Playmatch	Hot	Cool
Yes	0	3
No	2	0



c) $H(\text{playmatch}) = -\frac{4}{7}\log\frac{4}{7} - \frac{3}{7}\log\frac{3}{7} = 0.683$

$H(\text{playmatch}|\text{outlook}) = -\frac{2}{7}\left(\frac{2}{2}\log\frac{2}{2} + \frac{0}{2}\log\frac{0}{2}\right) - \frac{2}{7}\left(\frac{2}{2}\log\frac{2}{2} + \frac{0}{2}\log\frac{0}{2}\right)$

$- \frac{3}{7}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) = 0.27$

$H(\text{playmatch}|\text{climate}) = -\frac{3}{7}\log\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\left(\log\frac{1}{3}\right)\right) - \frac{1}{7}\left(\frac{1}{1}\log\frac{1}{1} + \frac{0}{1}\log\frac{0}{1}\right)$

$- \frac{3}{7}\log\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) = 0.54$

$H(\text{playmatch}|\cancel{\text{humidity}}\ \text{wind}) = -\frac{4}{7}\left(\frac{1}{4}\log\frac{1}{4} + \frac{3}{4}\log\frac{3}{4}\right) - \frac{3}{7}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)$

$= 0.59$

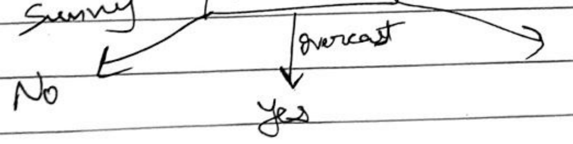
$H(\text{playmatch}|\cancel{\text{wind}}\ \text{humidity}) = -\frac{4}{7}\left(\frac{2}{4}\log\frac{2}{4} + \frac{2}{4}\log\frac{2}{4}\right) - \frac{3}{7}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)$

$= 0.67$

IG(playmatch, outlook) = 0.413
IG(playmatch, climate) = 0.143
IG(playmatch, wind) = 0.093
IG(playmatch, humidity) = 0.013
IG is the highest for outlook

Outlook

Playmatch	Sunny	Overcast	Rain
Yes	0	2	2
No	2	0	1



$$H(rain) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.636$$

$$H(rain, climate) = -\frac{1}{3}\cancel{\log}\left(\frac{1}{1}\log\frac{1}{1} + \frac{0}{1}\log\frac{0}{1}\right) - \frac{2}{3}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right)$$

$$= 0.462$$

$$H(rain, humidity) = -\frac{1}{3}\cancel{\log}\left(\frac{1}{1}\log\frac{1}{1} + \frac{0}{1}\log\frac{0}{1}\right) - \frac{2}{3}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right)$$

$$= 0.462$$

$$H(rain, wind) = -\frac{2}{3}\left(\frac{2}{2}\log\frac{2}{2} + \frac{0}{2}\log\frac{0}{2}\right) - \frac{1}{3}\left(\frac{1}{1}\log\frac{1}{1} + \frac{0}{1}\log\frac{0}{1}\right)$$

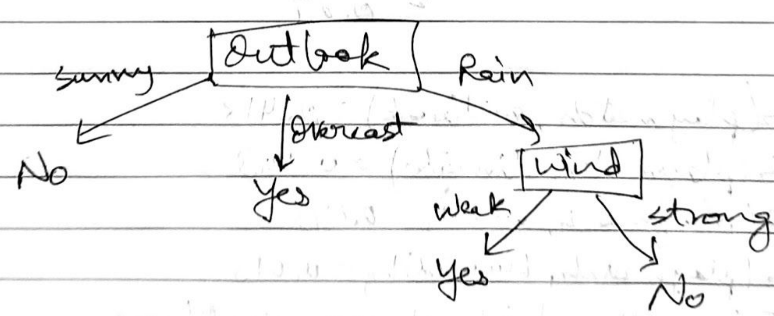
$$= 0$$

IG(rain, climate) = 0.174

IG(rain, humidity) = 0.174

IG(rain, wind) = 0.636

IG is the highest for wind.



Training set

Outlook	Climate	Humidity	Wind	Playmatch	Predicted
Sunny	Hot	High	Weak	No	No
Sunny	Hot	High	Strong	No	No
Overcast	Hot	High	Weak	Yes	Yes
Rain	Mild	High	Weak	Yes	Yes
Rain	Cool	Normal	Weak	Yes	Yes
Rain	Cool	Normal	Strong	No	No
Overcast	Cool	Normal	Strong	Yes	Yes

Training accuracy = 100%

Testing set

Outlook	Climate	Humidity	Wind	Playmatch	Predicted
Sunny	Mild	High	Weak	No	No
Sunny	Cool	Normal	Weak	Yes	No
Rain	Mild	Normal	Weak	Yes	Yes
Sunny	Mild	Normal	Strong	Yes	No
Overcast	Mild	High	Strong	Yes	Yes
Overcast	Hot	Normal	Weak	Yes	Yes
Rain	Mild	High	Strong	No	No

Testing accuracy = 71.43%

The above accuracies are obtained because the model is overfitting on the training data.

d) Pre-pruning can be used to avoid the model from overfitting. Pre-pruning means that we stop the tree from growing before all of the training set is perfectly classified. Error estimation can be used to find whether pruning or expanding a particular node would be better in a generalised sense. For this Chi-squared significance tests are used as a stopping criteria.

6. $p(\text{tough} \mid \text{tough}) = 0.7$

$p(\text{course} \mid \text{tough}) = 0.3$

$p(\text{tough} \mid \text{course}) = 0.5$

$p(\text{course} \mid \text{course}) = 0.5$

$(w_1 = \text{tough},\ w_2 = \text{course},\ w_3 = ?,\ w_4 = \text{course})$

This is a first-order Markov Model, therefore the probability of occurrence of a word depends on its previous word. So w_3 is affected by w_2 and w_4 is affected by w_3.

$p(w_3 = \text{course} \mid w_2 = \text{course})\, p(w_4 = \text{course} \mid w_3 = \text{course}) = 0.5\,(0.5)$
$= 0.25$

$p(w_3 = \text{tough} \mid w_2 = \text{course})\, p(w_4 = \text{course} \mid w_3 = \text{tough}) = 0.5\,(0.3)$
$= 0.15$

Since the only possibilities are $(w_3 = \text{tough},\ w_3 = \text{course})$

$p(w_3 = \text{course}) = \dfrac{0.25}{0.25+0.15} = \dfrac{25}{40} = 0.625$

$p(w_3 = \text{tough}) = \dfrac{0.15}{0.25+0.15} = \dfrac{15}{40} = 0.375$

7. a) The biggest advantage of decision trees over logistic regression is that they are more interpretable. Decision trees can be used to comment upon the significance of features. In logistic regression only a single linear decision boundary is possible whereas in decision tree the whole region can be divided into smaller subspaces.

b) The biggest weakness of decision trees over logistic regression is that they are prone to overfitting. Decision trees are highly affected by outliers whereas logistic regression is robust in terms of noise.

c) Yes, a decision tree can classify these vectors. Since the n vectors are linearly separable, at each node the decision tree would try to do a 50:50 split. So at each node the data is divided into half.

If h is depth of tree then,

$n = 2^h \rightarrow h = \log n$

Therefore the upper bound on the depth of the tree would be logn.

d) Now the n two dimensional vectors are not linearly separable. Still the decision tree can correctly classify these vectors. A binary search would be done on the two dimensions x1 and x2. Therefore the upper bound on the depth of the tree would be 2logn.

DATE__________
PAGE__________

8. $X = \langle X_1, X_2 \ldots X_n \rangle$

$P(Y=1) = \pi$

$\rightarrow P(Y=0) = 1-\pi$

To show: $P(Y=1|X) = \dfrac{1}{(1+\exp(w_0 + \sum_{i=1}^{n} w_i X_i))}$

For ~~Gaussian~~ Naive Bayes,

$$P(Y=1|X) = \frac{P(Y=1)P(X|Y=1)}{P(Y=1)P(X|Y=1)+P(Y=0)P(X|Y=0)}$$

$$= \frac{1}{1+\dfrac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}}$$

$$P(Y=1|X) = \frac{1}{1+\exp\left(\ln\left(\dfrac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}\right)\right)} \quad -①$$

$$P(X|Y=0) = P(X_1|Y=0)P(X_2|Y=0)\ldots\ldots P(X_n|Y=0) = \sum_{i=1}^{n} P(X_i|Y=0) \quad -②$$

$$P(X|Y=1) = P(X_1|Y=0)P(X_2|Y=0)\ldots\ldots P(X_n|Y=0) = \sum_{i=1}^{n} P(X_i|Y=1) \quad -③$$

Using ②, ③ in ①

$$P(Y=1|X) = \frac{1}{1+\exp\left[\ln\left(\dfrac{P(Y=0)}{P(Y=1)}\right)+\ln\left(\dfrac{\sum_{i=1}^{n}P(X_i|Y=0)}{\sum_{i=1}^{n}P(X_i|Y=1)}\right)\right]}$$

$$= \frac{1}{1+\exp\left[\ln\left(\dfrac{1-\pi}{\pi}\right)+\sum_{i=1}^{n}\ln\dfrac{P(X_i|Y=0)}{P(X_i|Y=1)}\right]}$$

Using Gaussian Naive Bayes,

$$P(X_i|Y=0) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(\frac{-(X_i-\mu_{i0})^2}{2\sigma_i^2}\right)$$

$$P(X_i|Y=1) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(\frac{-(X_i-\mu_{i1})^2}{2\sigma_i^2}\right)$$

$P(X_i=1|Y=1) = \theta_{i1}$ $\qquad$ $P(X_i=1|Y=0) = \theta_{i0}$

$P(X_i=0|Y=1) = 1-\theta_{i1}$ $\qquad$ $P(X_i=0|Y=0) = 1-\theta_{i0}$

$$\ln \frac{P(X_i|Y=0)}{P(X_i|Y=1)} = \ln(P(X_i|Y=0)) - \ln(P(X_i|Y=1))$$

$$= \ln\left[X_i\theta_{i0} + (1-X_i)(1-\theta_{i0})\right] - \ln\left[X_i\theta_{i1} + (1-X_i)(1-\theta_{i1})\right]$$

Since x can only take discrete values 0 and 1, we can write the above expression as follows:

$$\ln \frac{P(X_i|Y=0)}{P(X_i|Y=1)} = X_i \ln\theta_{i0} + (1-X_i)\ln(1-\theta_{i0}) - X_i\ln\theta_{i1} - (1-X_i)\ln(1-\theta_{i1})$$

$$= X_i\ln\theta_{i0} + \ln(1-\theta_{i0}) - X_i\ln(1-\theta_{i0}) - X_i\ln\theta_{i1} - \ln(1-\theta_{i1}) + X_i\ln(1-\theta_{i1})$$

$$= X_i\left[\ln\theta_{i0} - \ln(1-\theta_{i0}) - \ln\theta_{i1} + \ln(1-\theta_{i1})\right] + \ln(1-\theta_{i0}) - \ln(1-\theta_{i1})$$

$$= X_i\left[\ln\frac{\theta_{i0}(1-\theta_{i1})}{\theta_{i1}(1-\theta_{i0})}\right] + \ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right) \quad — (4)$$

$$P(Y=1|X) = \frac{1}{1+\exp\left[\ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n}\ln\frac{P(X_i|Y=0)}{P(X_i|Y=1)}\right]}$$

using (4)

$$= \frac{1}{1+\exp\left[\ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n}\left[X_i\ln\left(\frac{(1-\theta_{i1})\theta_{i0}}{\theta_{i1}(1-\theta_{i0})}\right) + \ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right)\right]\right]}$$

$$= \frac{1}{1+\exp\left[\ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n}\ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right) + \sum_{i=1}^{n}X_i\left(\ln\frac{\theta_{i0}(1-\theta_{i1})}{\theta_{i1}(1-\theta_{i0})}\right)\right]}$$

$$= \frac{1}{1+\exp\left(w_0 + \sum_{i=1}^{n} w_i X_i\right)}$$

where $w_0 = \ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=1}^{n}\ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right)$

$$w_i = \ln\left[\frac{\theta_{i0}(1-\theta_{i1})}{\theta_{i1}(1-\theta_{i0})}\right]$$