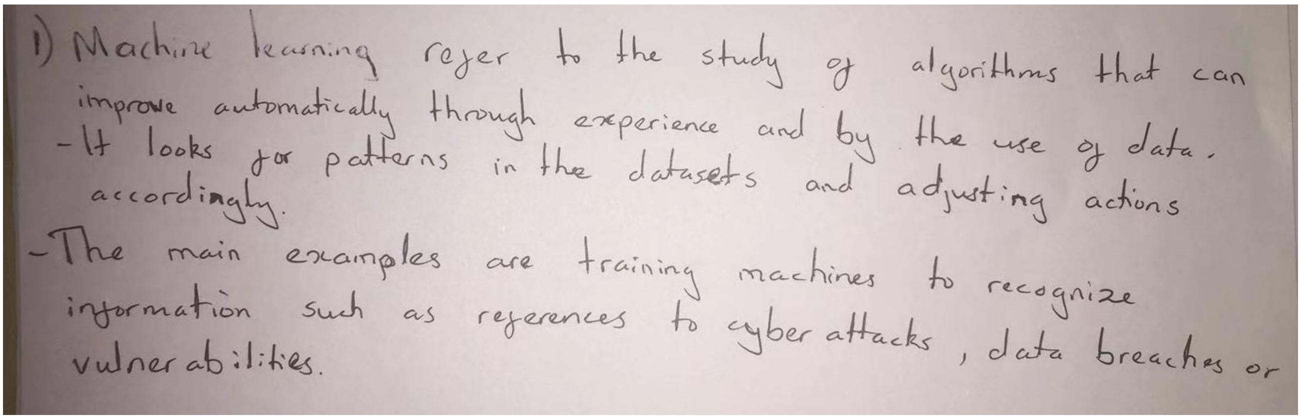
2021290010 - Mahadi Sajjad Neloy

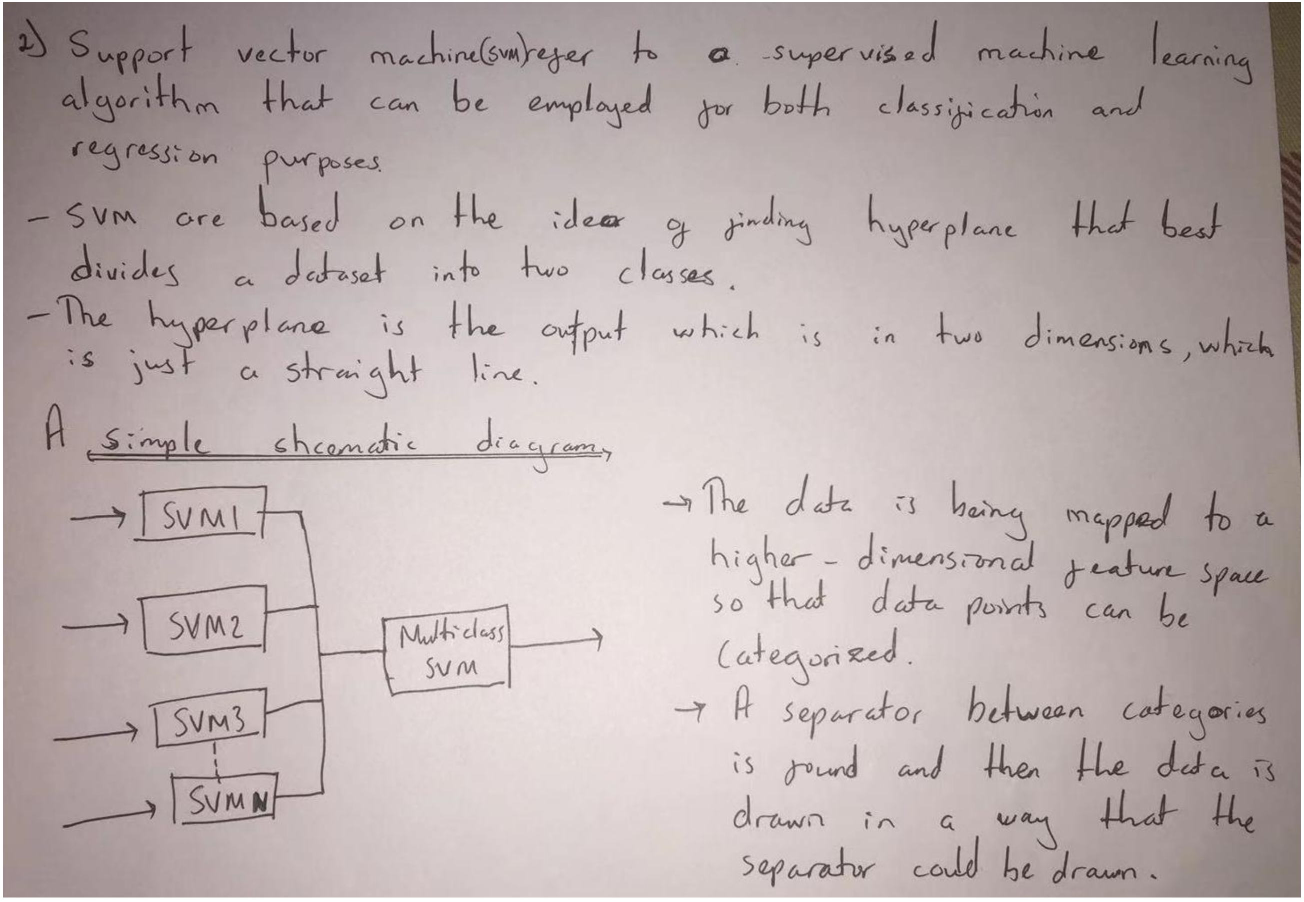
COMPUTER SCIENCE DEPARTMENT

MACHINE LEARNING - EXAM



1) Machine learning refer to the study of algorithms that can improve automatically through experience and by the use of data.

- It looks for patterns in the datasets and adjusting actions accordingly.
- The main examples are training machines to recognize information such as references to cyber attacks, data breaches or vulnerabilities.



2) Support vector machine(svm)refer to a supervised machine learning algorithm that can be employed for both classification and regression purposes.

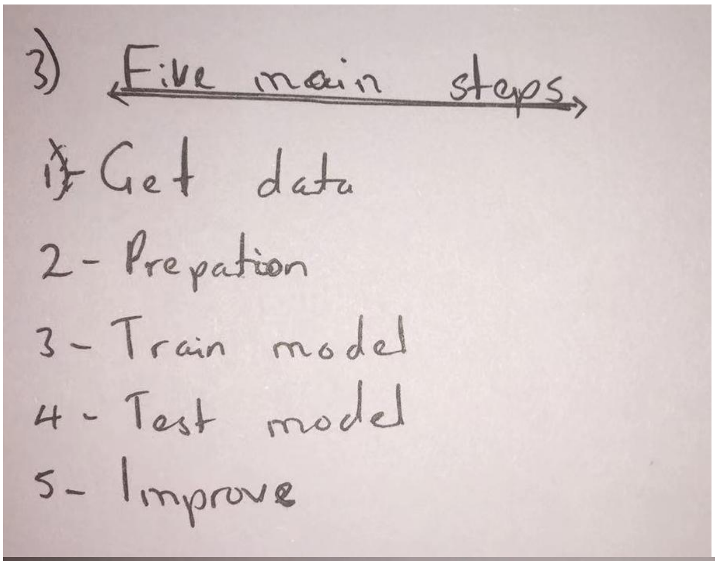
- SVM are based on the idea of finding hyperplane that best divides a dataset into two classes.
- The hyperplane is the output which is in two dimensions, which is just a straight line.

A simple shcematic diagram,

SVM1
SVM2
SVM3
SVMN
Multiclass SVM

→ The data is being mapped to a higher-dimensional feature space so that data points can be categorized.

→ A separator between categories is found and then the data is drawn in a way that the separator could be drawn.



3) Five main steps

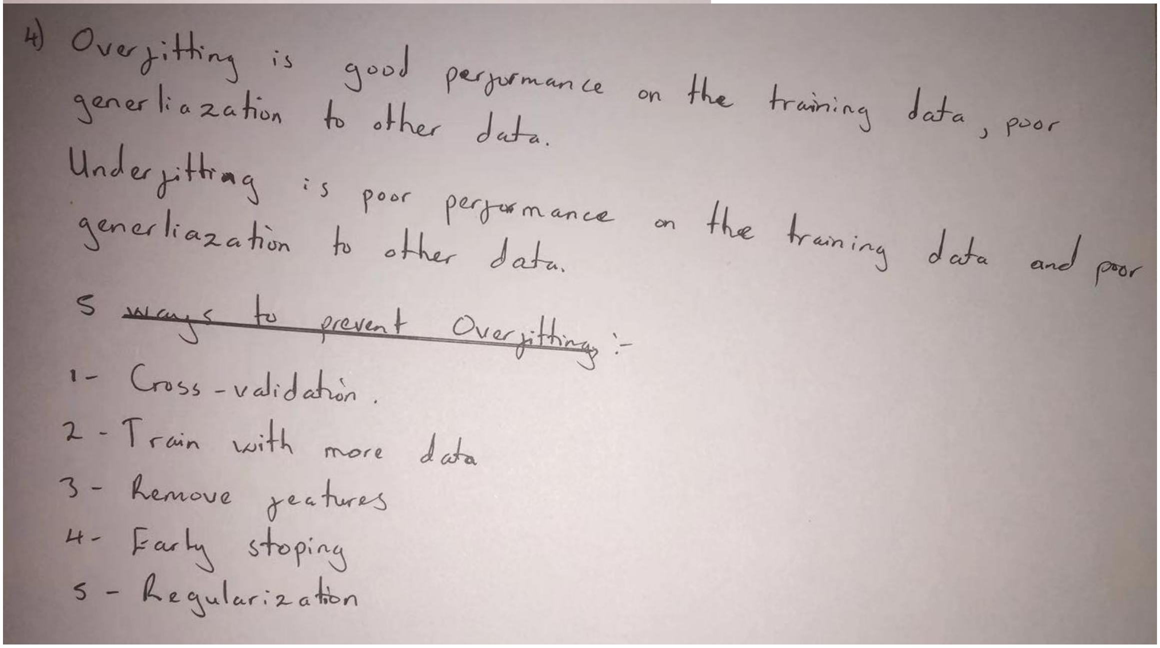
1- Get data

2- Prepation

3- Train model

4- Test model

5- Improve



4) Overfitting is good performance on the training data, poor generliazation to other data.

Underfitting is poor performance on the training data and poor generliazation to other data.

5 ways to prevent Overfitting :-

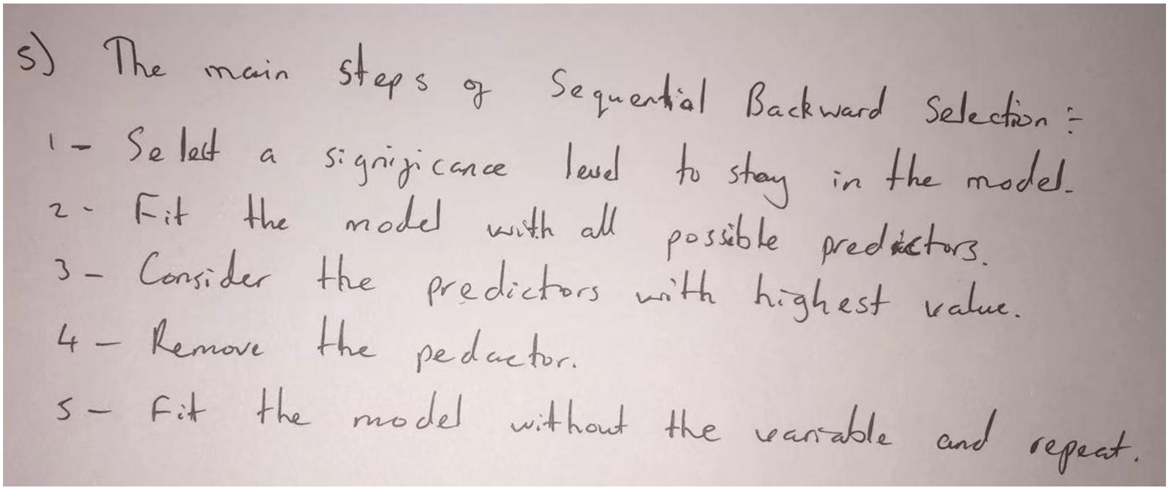
1- Cross-validation.

2- Train with more data

3- Remove features

4- Early stoping

5- Regularization



5) The main steps of Sequential Backward Selection :-

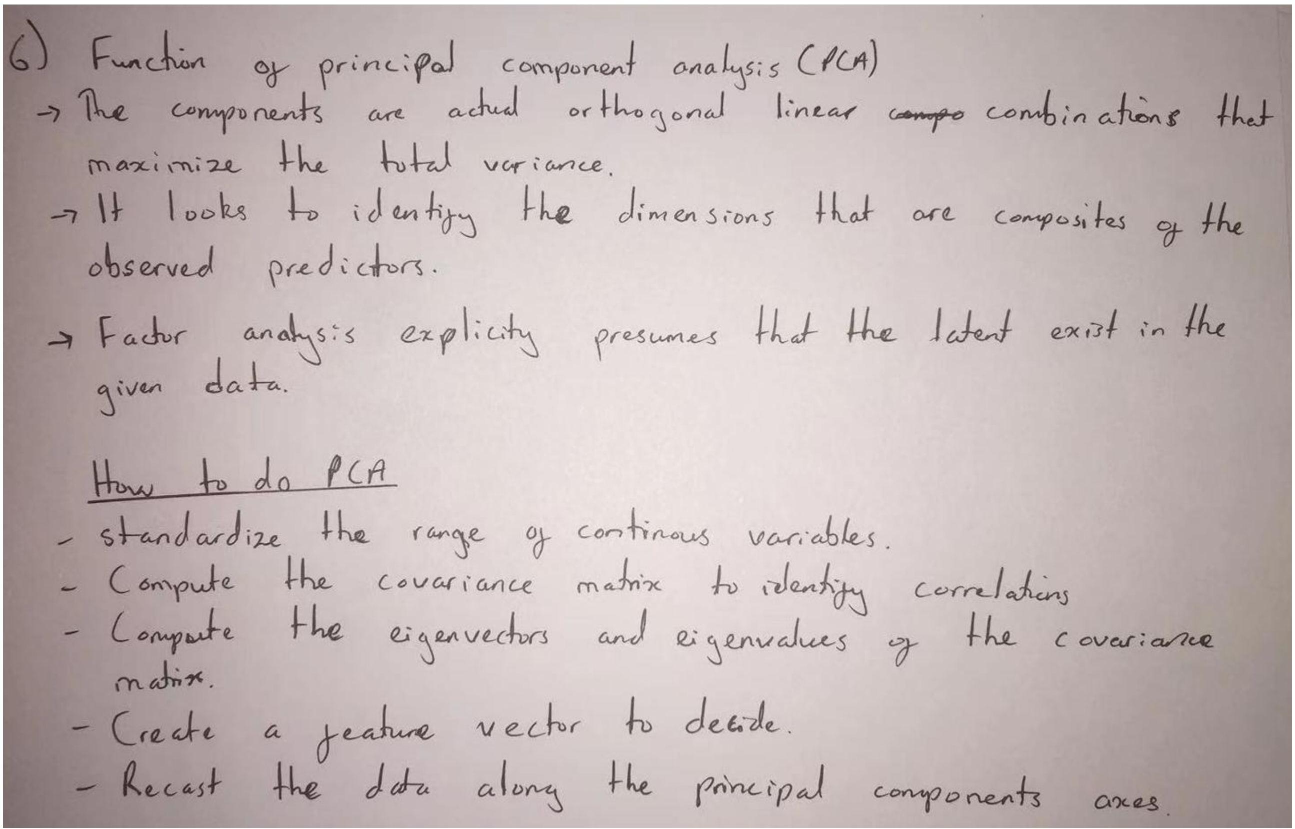
1- Select a significance level to stay in the model.

2- Fit the model with all possible predictors.

3- Consider the predictors with highest value.

4- Remove the pedactor.

5- Fit the model without the variable and repeat.



6) Function of principal component analysis (PCA)

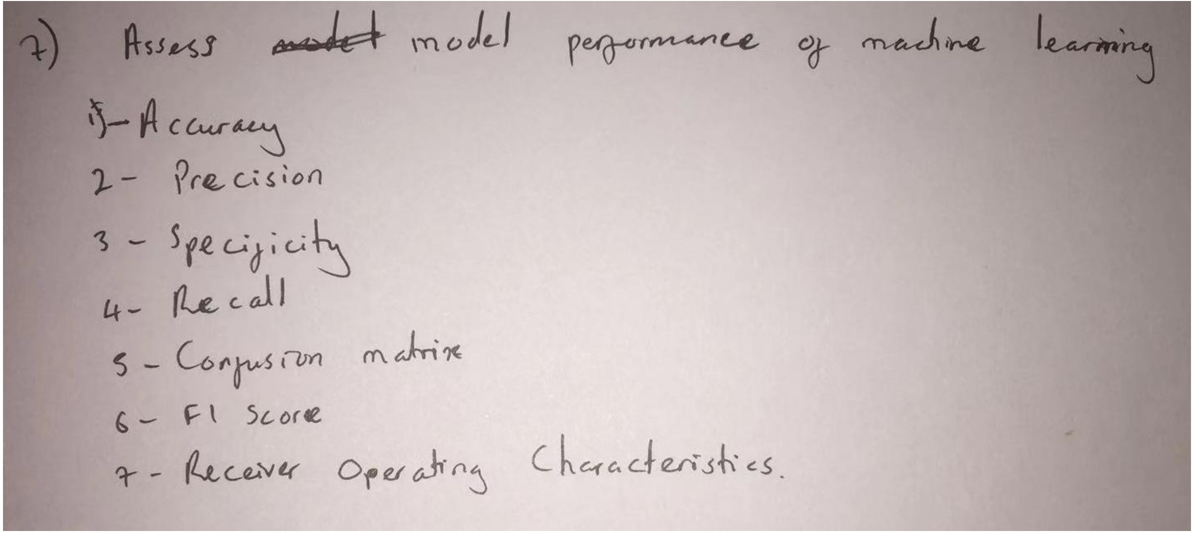
→ The components are actual orthogonal linear ~~compo~~ combinations that maximize the total variance.

→ It looks to identify the dimensions that are composites of the observed predictors.

→ Factor analysis explicity presumes that the latent exist in the given data.

How to do PCA

- standardize the range of continous variables.
- Compute the covariance matrix to identify correlations
- Compute the eigenvectors and eigenvalues of the covariance matrix.
- Create a feature vector to decide.
- Recast the data along the principal components axes.



7) Assess ~~model~~ model performance of machine learning

1- Accuracy

2- Precision

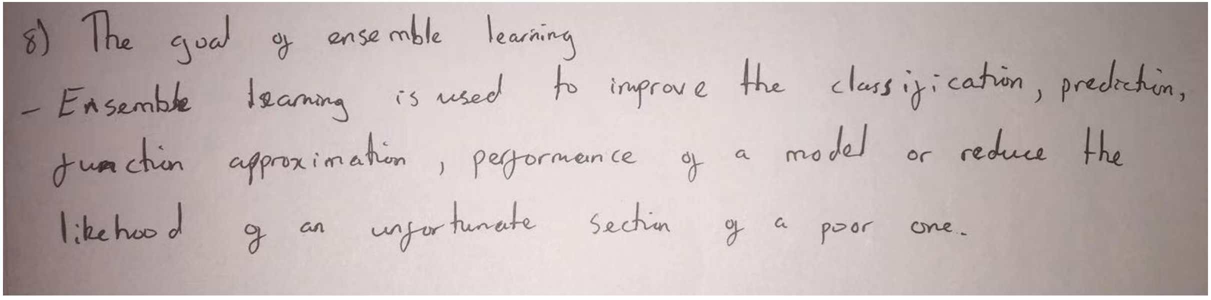
3 - Specificity

4- Recall

5 - Confusion matrix

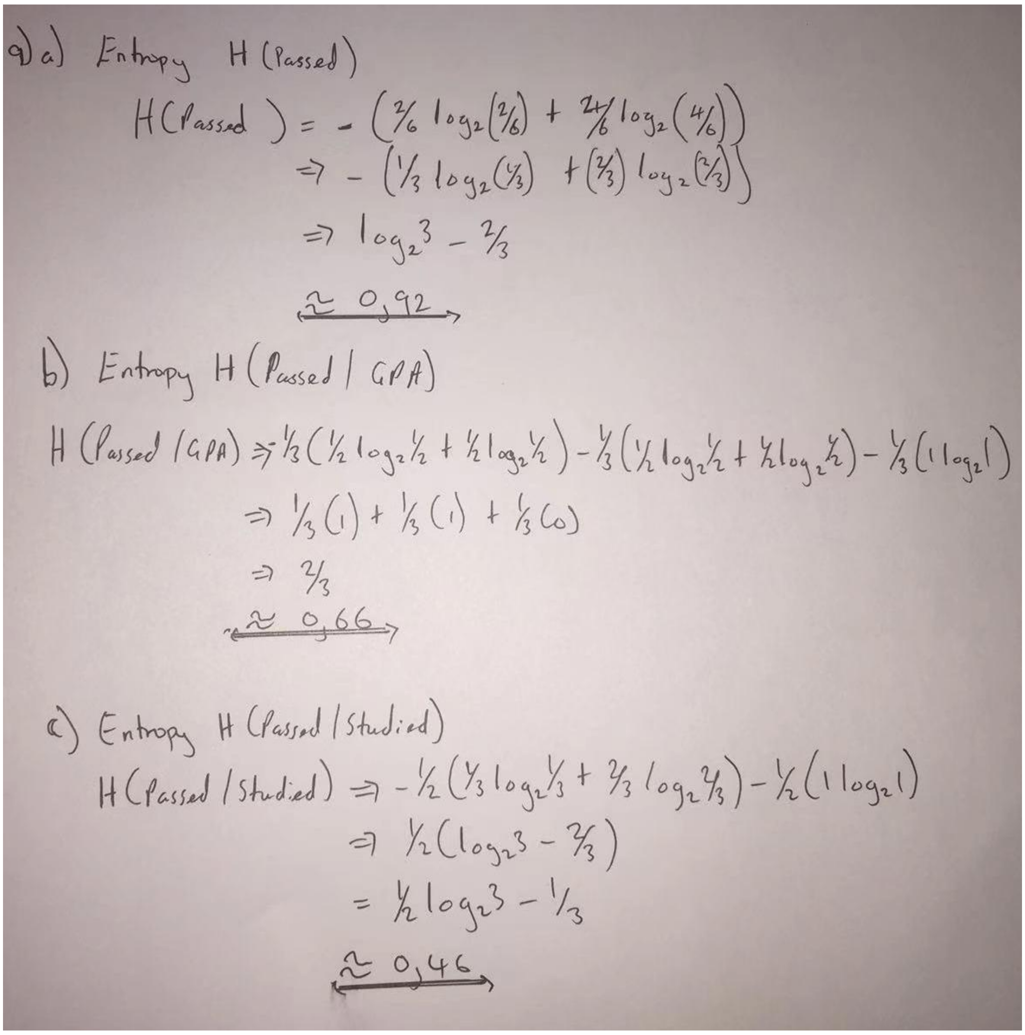
6- F1 Score

7 - Receiver Operating Characteristics.



8) The goal of ensemble learning

- Ensemble learning is used to improve the classification, prediction, function approximation, performance of a model or reduce the likehood of an unfortunate section of a poor one.



9) a) Entropy H (Passed)

$$H(\text{Passed}) = -\left(\tfrac{2}{6}\log_2\left(\tfrac{2}{6}\right) + \tfrac{4}{6}\log_2\left(\tfrac{4}{6}\right)\right)$$

$$\Rightarrow -\left(\tfrac{1}{3}\log_2\left(\tfrac{1}{3}\right) + \left(\tfrac{2}{3}\right)\log_2\left(\tfrac{2}{3}\right)\right)$$

$$\Rightarrow \log_2 3 - \tfrac{2}{3}$$

$$\underline{\approx 0{,}92}$$

b) Entropy H (Passed | GPA)

$$H(\text{Passed} / \text{GPA}) \Rightarrow -\tfrac{1}{3}\left(\tfrac{1}{2}\log_2\tfrac{1}{2} + \tfrac{1}{2}\log_2\tfrac{1}{2}\right) - \tfrac{1}{3}\left(\tfrac{1}{2}\log_2\tfrac{1}{2} + \tfrac{1}{2}\log_2\tfrac{1}{2}\right) - \tfrac{1}{3}\left(1\log_2 1\right)$$

$$\Rightarrow \tfrac{1}{3}(1) + \tfrac{1}{3}(1) + \tfrac{1}{3}(0)$$

$$\Rightarrow \tfrac{2}{3}$$

$$\underline{\approx 0{,}66}$$

c) Entropy H (Passed | Studied)

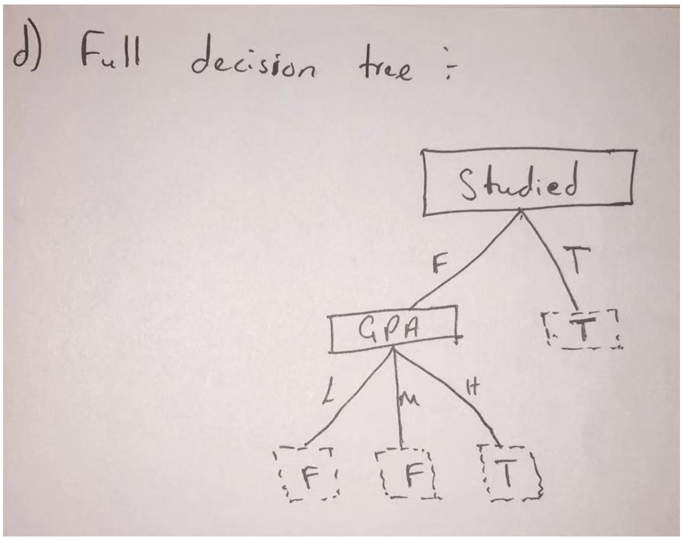
$$H(\text{Passed} / \text{Studied}) \Rightarrow -\tfrac{1}{2}\left(\tfrac{1}{3}\log_2\tfrac{1}{3} + \tfrac{2}{3}\log_2\tfrac{2}{3}\right) - \tfrac{1}{2}\left(1\log_2 1\right)$$

$$\Rightarrow \tfrac{1}{2}\left(\log_2 3 - \tfrac{2}{3}\right)$$

$$= \tfrac{1}{2}\log_2 3 - \tfrac{1}{3}$$

$$\underline{\approx 0{,}46}$$

d) Full decision tree :


Studied
F
T
GPA
T
L
M
H
F
F
T