19 essay questions
All questions were theoretical questions covering all topics with no proofs or math problems
Total marks: 120 (1 mark/min)

1) Define the Machine Learning Recipe
2) What's the key idea of the Adaboost algorithm ? Explain using a diagram :D
3) Why can't adaboost work for more than binary classification? How would you modify it to work for more?
    a) Because it may give us negative weights in some cases.
    b) We added the term log(k - 1) to the weight.
4) What's the difference between wrapper method and filter method for feature selection?
    a) Wrapper method: select features by taking into consideration the classifier you will use. for example: SFS and SBS.
    b) Filter Method: select features without taking into consideration the classifier you will use. for example: correlations of the features and the distance between class means.
5) Compare between AI, ML, & DL
    a) AI: it's categorized into 4 categories. Thinking humanly, acting humanly, thinking rationally and acting rationally. Most computer scientists consider it as acting rationally.
    b) ML: it's a subset of AI where computers are learning from examples.
    c) DL: it's a subset of MI where we use multi-layered neural networks.
6) If u train a neural network and get 54% training accuracy avbnd 51% validation accuracy, explain what u will do next
7) Explain with examples the linear perceptron update rule (Withrow Hoff wala mesh faker esmo eh)
8) Explain the naive estimator method, write the formula used and compare it to histogram analysis
    a) We perform a range analysis for each x.
    b) P = (#points falling in [x-(h/2), x+(h/2)]) / (M*h)
    c) It's the same as histogram analysis but instead of partitioning the feature space into pre-specified ranges, we take a range around each x. So the naive estimator is smoother than the histogram.
9) Explain the kernel density estimation technique
    a) We apply a bump function on each x so that we take into consideration the weight effect of each x. where the bump function should be CDF.
10) Explain why relu is used instead of sigmoid
11) Explain the least square classifier. No need for proof
12) What are the main issues in GMM?
13) types of features to be removed in feature selection method
    a) Irrelevant features
    b) Correlated features

14) Explain regularization and why it is used
15) the importance of statistical testing.
16) Explain how bayes rule can be used in classification
    a) We calculate P(x | Ci) * P(Ci) for each class then we classify x for the class that give the highest value for P(x | Ci) * P(Ci).
17) Explain the importance of deseasonalization using diagrams.
18) State the disadvantages of backpropagation.
19) Explain the three methods used to update weights