Queen Mary
University of London

Main Examination Period 2018

ECS708P Machine Learning Duration: 2 hours 30 minutes

YOU ARE NOT PERMITTED TO READ THE CONTENTS OF THIS QUESTION PAPER UNTIL INSTRUCTED TO DO SO BY AN INVIGILATOR

Answer FOUR questions

Calculators **are** permitted in this examination. Please state on your answer book the name and type of machine used.

Complete all rough workings in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

EXAM PAPERS MUST NOT BE REMOVED FROM THE EXAM ROOM

Examiners: Dr. Ioannis Patras, Dr. Anthony Constantinou

Question 1

a) Define the conditional probability $p(A|B)$ in terms of the joint probability $p(A,B)$. You may want to use a diagram/sketch.

[3 Marks]

b) This question is about random variables.

i. Give the relation between the joint probability $P(X,Y)$ and the probabilities $P(X)$ and $P(Y)$ that holds in the case that X and Y are independent random variables.

ii. Give the condition that holds when two random variables X and Y are uncorrelated.

iii. What is the relation between those these two conditions?

[6 Marks]

c) An IT worker works from home 2 days a week and at office 3 days a week. When she works from home she answers 70% of the emails within the first hour of their receipt and the remaining within the next hour. When she is at the office, she answers 50% of the emails within an hour of their receipt and the remaining within the next hour. That means that in both cases all emails will be answered within two hours from their receipt.

i. If you send her an email, what is the probability that she will answer within 1 hour?

ii. Given that she hasn't replied to your email within 1 hour, what is the probability that she is working from home? Does the information that she hasn't answered the email within 1 hour makes it more or less likely that she works from home?

[10 Marks]

d) Explain the difference between Maximum Likelihood (ML) and Maximum a Posterior (MAP) methods of learning parameters θ from data X.

[6 marks]

Question 2

a) Compare and contrast the goals in Linear Regression and Logistic Regression.

[4 marks]

b) The form of a linear regression model is $y=\mathbf{w}^T\mathbf{x}$. Assuming the mean squared error cost function, derive gradient descent updates for the weights **w**.

[9 marks]

c) What is the limitation of the networks without hidden layers, that was overcome by Multilayer Networks? Is it essential that the activation function is non-linear?

[6 marks]

d) Practical challenges with training neural networks include: (i) getting stuck in local optima, (ii) underfitting or overfitting. Explain the practical strategies you would use to overcome these issues.

[6 marks]

Question 3

(a) Describe the difference between supervised and unsupervised learning. Give an example of a real world problem that requires a supervised learning algorithm and an example of a real world problem that can be solved with an unsupervised learning algorithm. In both cases define the inputs and the outputs.

[8 marks]

(b) Describe in detail the steps of the K-means algorithm. Make sure that you define the input to the algorithm, the output, and the dimensionality of all the variables that you use.

[8 marks]

(c) Identify the two sets of variables that are estimated by the K-means algorithm. Explain what coordinate descent (or coordinate optimisation) is. Using a sketch, show that this general optimisation method is warranted to converge.

[4 marks]

(d) The K-Means algorithm converges to a local minimum. Describe a practical method to deal with this problem. Can this method be used to determine the optimal value of K?

[5 marks]

Question 4

(a) With the help of a diagram explain the main principles of the first-order Hidden Markov Model (HMM). What are the differences to a Markov model? In your answer, define the states ω_i, the symbols v_k, and the matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{jk}]$.

[6 marks]

(b) The decoding problem can be stated as follows: Given an HMM and a sequence of observation symbols $\mathbf{V}^{1:T}$ determine the most likely sequence of hidden states $\omega^{1:T}$ .What are the other two types of problems considered in the context of HMMs?

[6 marks]

(c) You are given the task to design a speech recogniser that distinguishes between the words “yes” and “no”.

i. Describe how the K-means algorithm can be used to transform a speech signal to a sequence of symbols $\{v_1, v_2, \ldots, v_K\}$. An example is given in Figure 1 below:

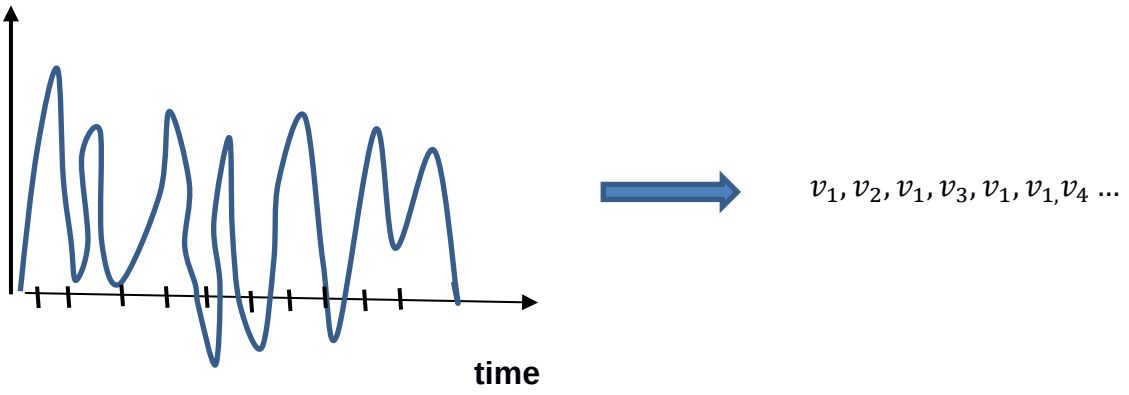

Figure 1

(Hint: As a first step divide the signal in small chunks of equal and fixed length)

ii. Describe how you would design the training process. Make reference to what kind of data you will have, which algorithm you will use (e.g. refer to your answer to part (b)) and how many HMMs you need to train. Why can’t one use a simple classifier, such as a linear logistic regression scheme?

iii. Describe how you would use the HMMs that you trained in order to make decisions at test time.

[13 marks]

End of Paper