Queen Mary
University of London

***MSC* Examination by course unit**

Wednesday 24 May 2017 10:00 am

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 ALL Four Questions.

Cross out any answers that you do not wish to be marked.

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:
Ioannis Patras
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) Give the law of total probabilities, that is express $p(A)$ using a set of events $B_1, B_2, \ldots, B_N$ and the corresponding conditional (or joint) probabilities. What are the conditions that need to hold?

[4 Marks]

c) Some emails received by users are spams containing viruses. You are building a system to detect such illicit virus emails. You start by using the feature of whether or not an email contains an executable attachment, as this an important datum indicating whether the email in fact contains a virus. Data analysis suggests that 95% of virus emails contain executable attachments, 90% of legitimate emails do not contain executable attachments, and 2% of emails overall are viruses.

i. If your classifier scans an email an executable attachment, what is the probability that the email in fact contains a virus?

ii. Comment on this value that you calculate. How does it compare with a decision based only on the frequency of the emails that contain viruses?

iii. What is the probability that your classifier makes an error?

[12 marks]

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

[6 marks]

[Q1 total 25 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 is essential that the activation function is non-linear?

[6 marks]

d) Practical pitfalls with training neural networks include: (i) getting stuck in local optima, (ii) underfitting or overfitting, (iii) bad learning rate. Explain what each of these means.

[6 marks]

[Q2 total 25 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]

[Q3 total 25 Marks]

Question 4

(a) With the help of a diagram explain the main principles of the first-order Markov Model. In your answer explain any notation that you use. Explain what is meant by the term ‘’first-order”.

[6 marks]

(b) What are the differences between a Markov-Model and a hidden Markov model (HMM)? What are the advantages of HMMs in comparison to Markov Models? Give an example of an application (a toy example will suffice) where an HMM can be used but a Markov Model cannot. 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]

(c) Given a Hidden Markov Model, its states, the observation symbols and the transition and emission probabilities,

i. Give the formula for the joint probability $p(\omega^{1:T}, \mathrm{V}^{1:T})$ of a sequence of observation symbols $V^{1:T} = (v(1), v(2), \ldots, v(t), \ldots, v(T))$ and a sequence of states $\omega^{1:T} = (\omega(1), \omega(2), \ldots, \omega(t), \ldots, \omega(T))$.

ii. Give the formula for the probability $p(\omega^{1:T})$ of a sequence of states $\omega^{1:T} = (\omega(1), \omega(2), \ldots, \omega(t), \ldots, \omega(T))$.

[6 marks]

(d) What is the evaluation problem? Using the results of (c) present a naïve algorithm that solves the evaluation problem. What is the computational complexity of that algorithm? Can this algorithm be used in practice?

[7 marks]

[Q4 total 25 Marks]

End of Paper