MSc Examination

Thursday 15th May 2014   10:00 - 12:30

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 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. Chrisantha Fernando, **Dr. Ioannis Patras,**

Question 1

a) Apply gradient descent on the least mean squared error cost function for linear regression for 2 iterations for the dataset in Table 1 below. x1 and x2 are the two input dimensions. Y is the target output for each row. Start with initial weights w0 = 0, w1 = 0, and w2 = 0. Use a learning rate of 10.

X1	X2	Y
0	0	0
0	1	1
1	0	1

Table 1

[5 Marks]

b) Plot the least mean squares cost function at the start, and after each iteration. Do you notice anything strange, and what possible changes to the previous steps would make a difference?

[5 Marks]

c) Explain how you would undertake basis function expansion of the 2 dimensional input vector to the 2nd order? Draw the resulting input vector and the values of the expanded dimensions for the inputs described in part 1a). QUESTION CONTINUES ON NEXT PAGE.

[5 Marks]

d) The following data were obtained from two kinds of fish, namely salmon (class 1) and trout (class 0). Write the equation for the output of a sigmoid logistic regression unit with an initial weight of 0.1, and a bias of 0.1, that takes the inputs from Table 2 below. In which classes are the fish classified at the start? Train the unit by gradient descent for one iteration on the dataset with a learning rate of 0.01, and say in which class each fish is classified now. Why are these wrong?

Fish Type	Length
S (1)	20
S (1)	15
T (0)	10
T (0)	5

Table 2

[5 Marks]

e) How could you use linear regression to predict the future value of a stock price, given its history, and the history of other stock prices? How would you evaluate your system before using it to invest your student loan?
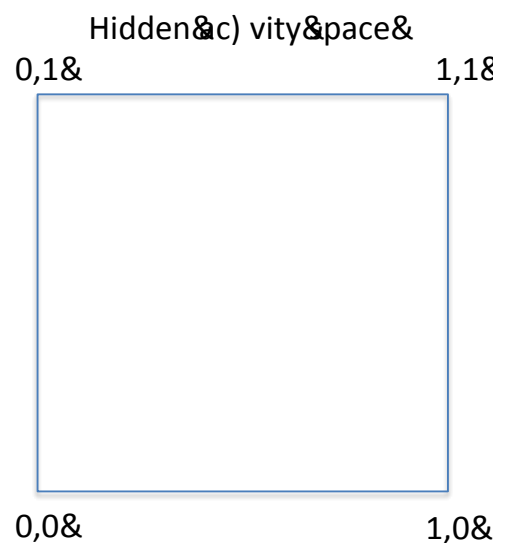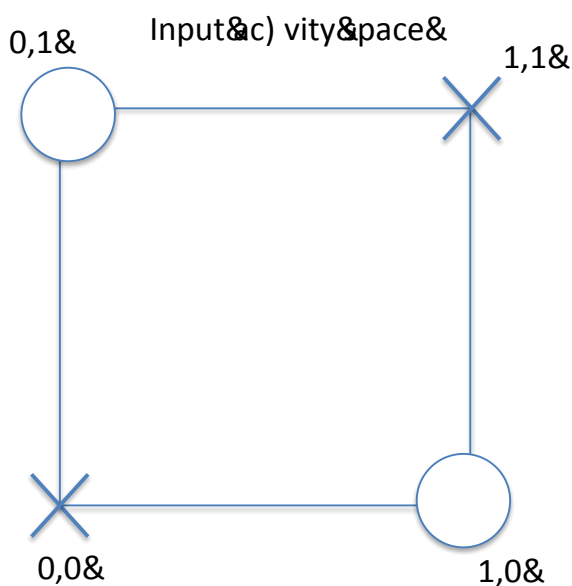
[5 Marks]
[Total 25 Marks]

Question 2

a) Consider the feedforward neural network produced by neurons with a step threshold transfer function such that when the dot product of the input vector and weight vector is less than zero, the neuron outputs 0, and when the product is greater than or equal to zero the neuron outputs 1. Draw the architecture of such a network that is capable of solving the XOR problem. Include all the weights and bias of each neuron. Show, using a table **for each neuron,** the values of the outputs of that neuron for each of the input data points (0,0), (0,1), (1,0), (1,1).

[5 marks]

b) For the network you have drawn in part 1 a) we wish to interpret what the hidden layer is doing in terms of feature mapping from the input space to a hidden layer space. The input points (0,0) and (1,1) are shown as an 'x', and input points (1,0) and (0,1) are shown a 'o' on a unit square. A unit square for the hidden state activities is also shown. Show clearly which points in hidden layer activity space each of the input points maps to. Show that a decision boundary can be drawn in this hidden layer activity space that makes the XOR problem linearly separable in this space, but not in the input space.

[5 marks]

c) Answer the following 3 parts briefly.

i). Before the backpropagation algorithm was invented, what methods were available to discover the weights of a neural network with hidden layers for regression or classification?

ii). Describe why the backpropagation algorithm is more efficient than such techniques.

iii). Why can't gradient descent, as used in logistic regression, be used to train feedforward neural networks with hidden layers?

[5 marks]

d) Describe how you would train a neural network to classify handwritten digits, including what practical methods could be used to prevent the following pathologies:

    i.   getting stuck in local optima,

    ii.   overfitting,

    iii.   underfitting.

How might you introduce the domain knowledge that the identity of the digit is invariant to translation?

[5 marks]

e) How could you tune the number of hidden neurons in the hidden layer of a neural network? Why use only one hidden layer and not 2 or 3, or 100 hidden layers? What other methods are there available for carrying out non-linear classification and regression tasks other than neural networks?

[5 marks]
[Total 25 marks]

Question 3

a) Define the joint probability $P(A, B)$ and the conditional probability $P(A \mid B)$. You may want to use a diagram/sketch. Give the formula that relates them.

[4 marks]

b) 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. Give the condition that holds when X and Y are uncorrelated. Are these conditions the same?

[4 Marks]

c) Show that the expected value of the sum of two independent random variables X and Y is equal to the sum of the expected values of X and Y. That is, show that $E(X+Y) = E(X) + E(Y)$. You may show it either for continuous or discrete variables. (Hint: You need to work with the joint probability $P(X, Y)$.)

[6 Marks]

d) An IT worker works from home 2 days a week. When she works from home there is a 30% chance she will not answer an email within an hour, 10% chance that she will not answer an email within two hours, and it is certain that she will answer all the emails within the day. When she is at office, there is a 50% chance she will not answer an email within one hour, 10% chance that she will not answer an email within the two hours, and it is certain that she will answer all the emails within the day.

   i.   If you send her an email, what is the probability that she will answer within 2 hours? Justify your answer.

   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?

[11 Marks]

[Total 25 marks]

Question 4

a) With a help of a diagram explain the main principles of the first-order Markov Model. Explain what is meant by the term ''first-order''. What are the differences with a hidden Markov model (HMM)? 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) What algorithm is used to solve the decoding problem described in 4.b). Give a brief description of the algorithm using pseudo code.

[13 marks]

[Total 25 marks]

End of Paper