COMP 24111

Two hours - online hybrid

The exam is hybrid and will be taken on line and answered on paper.

EXAM PAPER MUST NOT BE REMOVED FROM THE EXAM ROOM

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Machine Learning and Optimisation

Date: Thursday 17th January 2019

Time: 14:00 - 16:00

This is a hybrid examination with sections to be answered online and questions to be answered on paper.

Please answer All Questions in Section A online
and All Questions in Section B and Section C in separate answerbooks

---

This is a CLOSED book examination

The use of electronic calculators is permitted provided they are not programmable and do not store text

[PTO]

Section A contains Multiple Choice Questions and is restricted

Section B

Answer this section in a new answer booklet.

1. Four drivers have insured with a company based on similar auto insurance policies. The table below lists their driving experience and monthly auto insurance premiums.

Name	Driving Experience (Unit: Years)	Monthly Auto Insurance Premiums (Unit: 10 Dollars)
Tom	2	5
Mary	6	3
Kate	3	4
Bob	1	6

The task is to build a linear model to predict the monthly auto insurance premiums for new drivers according to their driving experiences. The model is to be trained using the training data as provided in the table and using the regularised least squares approach.

In general, a linear model computes an output from d inputs by $\hat{y} = \sum_{i=1}^{d} w_i x_i + w_0$. The regularised least squares approach trains the model by minimising the following objective function:

$$O = O_e + \frac{1}{2}\lambda \sum_{i=0}^{d} w_i^2,$$

where $\lambda \geq 0$ is a regularisation parameter, and O_e is the sum-of-squares error function.

a) Write down the formulation of your linear model for predicting insurance premiums and state what model parameters need to be optimised. (1 mark)

b) Write down your optimisation objective function O for training your model. (2 marks)

c) Derive the partial derivatives of the objective function. (2 marks)

d) One way to train the model is by letting the partial derivatives equal to zero and solving the resulting linear equations. Following this approach, compute the optimal model parameters, under the hyperparameter setting $\lambda = 1$. (2 marks)

e) Repeat the same process as above, and compute the optimal model parameters under the hyperparameter setting $\lambda = 0$. (2 marks)

f) Between the two linear models trained in (d) and (e), which one is a better model and explain your answer? (2 marks)

g) Instead of setting the partial derivatives to zero, stochastic gradient descent can be used for training. At the t-th iteration, only the driver Mary is used to estimate the gradient. Write down the model updating rule for this iteration. Set the learning rate as $\eta = 1$, and set the regularisation parameter as $\lambda = 0$. (2 marks)

2. There is a computer vision task called visual question answering. In this task, a machine learning system is built to automatically answer a question that is related to the content of an image. The system is trained by learning from many examples each containing an image, a question and an answer, and aims at building a mapping between the image-question pair and a possible answer. The figure below illustrates some examples used for training. Discuss whether this is an unsupervised or supervised learning task? (2 marks)

Section C

Answer this section in a new answer booklet.

1. Due to a lack of sufficient training data, the *naïve Bayesian* learning may encounter a "*zero conditional probability*" problem. For a discrete feature of three attribute values, $a = \{a_1, a_2, a_3\}$, the conditional probabilities with respect to a specific class, c, estimated on a training data set of n examples are $P(a = a_1|c) = \frac{1}{3}$, $P(a = a_2|c) = \frac{2}{3}$ and $P(a = a_3|c) = 0$, respectively. By introducing m virtual training examples, use the *m-estimate* method learned from this course unit to re-estimate all the conditional probabilities mentioned above in order to remedy the zero conditional probability problem. (5 marks)

2. Describe the main steps in the *Agglomerative* algorithm for clustering analysis. (5 marks)

3. An *internal* validity index may help the *K-means* determine an optimal number of clusters, K. With the notation used in the lecture note, describe the scatter-based *F-ratio* index, a commonly used internal index, and its implication in evaluating the quality of a partition resulting from the *K-means* clustering. (5 marks)