COM2028/11/Semr2 14/15

UNIVERSITY OF SURREY ©

Faculty of Engineering and Physical Sciences

Department of Computing

Undergraduate Programmes in Computing
Undergraduate Programmes in Mathematics

Module COM2028: 15 credits

Introduction to Artificial Intelligence

FHEQ Level 5 Examination

Time allowed: Two hours

Semester 2 2014/15

Answer **ALL** questions

Approved calculators ARE permitted

Each question carries 25 marks

Where appropriate the mark carried by an individual part of a question is indicated in square brackets [ ]

COM2028/11/Semr2 14/15

Solutions

1. This question is about *general understanding of the subject area.*

(a) Which one or more of the following are true about a cumulative histogram?

A. It has a steep slope at intensity levels that occur less frequently.

B. It has a steep slope at intensity levels that occur more frequently.

C. It increases the contrast in all image areas if used as a mapping function in histogram equalisation.

D. It can be used as a feature for classification.

[4 marks]

▌ B, D

(b) Consider the following situations, in which case(s) the colour histogram of the image will remain the same.

A. The image is flipped vertically.

B. The brightness of the image is enhanced by a factor of 15, *i.e.* the intensity level of each pixel increases by 15.

C. The image is rotated by 180 °

[3 marks]

▌ A, C

(c) The following are convolution kernels. Select those that can smooth an image.

Convolution kernels:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix} \; B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 1 \end{pmatrix} \; C = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{pmatrix} \; D = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

[4 marks]

SEE NEXT PAGE

B

(d) Which of the following techniques can be categorised as unsupervised learning?

A. k-nearest neighbour.

B. Back-propagation neural network.

C. Simulated annealing.

D. Hierarchical clustering.

E. k-means clustering.

F. A single perceptron.

G. Self-organising feature map (SOM).

[3 marks]

E, G

(e) If we are to detect cars in surveillance videos, what is the best feature to be included in the feature vector for classification?

A. The centre position of a detected object

B. Histogram of Oriented Gradients (HOG)

C. Colour histogram

D. Averaged intensities in the image frame

[2 marks]

B.

(f) The following statements are about image processing. Select each statement that is true.

A. When each pixel intensity multiplies a factor that is greater than 1, the contrast is decreased.

SEE NEXT PAGE

B. When each pixel intensity adds a factor that is greater than 1, the image brightness is enhanced.

C. When each pixel intensity is subtracted from 255, the image is flipped horizontally.

[3 marks]

▌ B

(g) Which of the following can describe k-means clustering algorithm?

A. There are k cluster centres for each cluster.

B. The k cluster centre is updated each time when the cluster members are changed.

C. The top k neighbours of each data sample are chosen and the distances between each two neighbours are calculated.

D. The algorithm will stop when all centres move to the same position.

[4 marks]

▌ A, B

(h) Which of the following can describe k-nearest mean algorithm?

A. The average of the training samples for each class is calculated.

B. When testing a sample, the distance between the testing sample and each class centre is calculated.

[2 marks]

▌ A, B

Total marks: 25

SEE NEXT PAGE

2. This question is about *these machine learning models: multilayer perceptron(MLP), self-organising feature map (SOM), and Support Vector Machines(SVMs).*

   The following is a list of possible processing steps for building the above machine learning algorithms.

   A. For any data sample, extract features.

   B. The data samples are mapped to another feature space so that the data of different classes are divided by a clear gap that is as wide as possible.

   C. The data samples are mapped to another feature space so that the data of different classes are divided by a clear gap that is as narrow as possible.

   D. Divide the data samples into training sets and testing sets.

   E. Calculate the value at each neuron through multi-layered feed forward networks.

   F. Based on the difference between the current output from the output layer and the target value/desired output, update the weights between output layer and middle layer.

   G. Update the weights between middle layer and input layer.

   H. The algorithm needs to find the support vectors that determine maximum-margin hyperplane.

   I. The algorithm needs to find the support vectors that determine minimum-margin hyperplane.

   J. Calculate the Euclidean distance between the training sample and each output neuron.

   K. The neuron with shortest distance to the training sample is selected and its weights are adjusted to be more similar to the training sample.

   L. The neuron with shortest distance to the training sample is selected and its weights are adjusted to be more dissimilar to the training sample.

   M. Assign random centres for each class group.

   For each machine learning model, choose the correct steps from above list and put them in the order to describe one training cycle.

   [25 marks](8 marks each for MLP and SOM, and 9 marks for SVMs)

SEE NEXT PAGE

MLP: A, D, E, F, G

SOM: A, D, J K

SVMs: A,D, B, H (A, D, H, B also acceptable)

Total marks: 25

SEE NEXT PAGE

3. This question is about *Bayes Theorem and its use.*

   A. What is the formula of Bayes Theorem? [3 marks]

   B. Like many other machine learning approaches, both Bayesian Classifier and Gaussian Classifier have training and classification/prediction phases. In which phase is the Bayes Theorem applied? [4 marks]

   C. Explain how Bayes Theorem is used in Bayesian classifier. [8 marks]

   D. Give one example in which a Bayesian classifier is applied. [10 marks]

   A.

$$P(A,B) = P(A) * P(B \mid A) = P(B) * P(A \mid B)$$
$$\Rightarrow P(B \mid A) = \frac{P(B) * P(A \mid B)}{P(A)} \quad (1)$$

   B. Classification phase.

   C.

   For Bayesian Classifier: Classification stage: After a Bayesian classifier has been trained, it can be used to automatically classify new items. The feature probabilities need to be combined into a single probability for the entire item.

$$Pr(Category \mid newitem) = Pr(newitem \mid Category) * Pr(Category) \quad (2)$$

   (4 marks)

   Where

$$Pr(newitem \mid Category) = Pr(Feature1 \mid Category) * Pr(Feature2 \mid Category) * \cdots \quad (3)$$

   Where $Pr(Category)$ is the overall frequency of the category. Calculate this for each category. Whichever category gets a higher score for $Pr(Category \mid newitem)$ is the predicted category. (4 marks)

   D. Assume we are analysing the sentiment about people's view on Microsoft. We collect many tweets on the topic and separate them into positive and negative groups. So there are two categories (classes), positive and negative.(2 marks) Training stage: Like all supervised methods, a Bayesian classifier is trained with

SEE NEXT PAGE

examples. Each example is a list of an items features and the classification for that item. The items are the tweets and the features are the words in the tweet. We can create a histogram of n-gram feature vector for each tweet. The training process takes an item (a tweet) and a classification. It then breaks the item in to its separate features based on the n-gram we defined. And counts for this classification for every feature. This is for the classifier to keep track of all the features it has seen so far, along with numerical probabilities that the features are associated with a particular classification. The classifier is trained by receiving examples one by one. After each example, the classifier updates the probabilities for the features and their classification in that example, generating the probability that a tweet about a certain category will contain a given word/phrase (in n-gram). (5 marks) Classification stage: After a Bayesian classifier has been trained, it can be used to automatically classify new items (new tweet). The feature probabilities need to be combined into a single probability for the entire item.

$$Pr(Category \mid tweet) = Pr(tweet \mid Category) * Pr(Category)/Pr(tweet) \quad (4)$$

Where

$$Pr(tweet \mid Category) = Pr(Feature1 \mid Category) * Pr(Feature2 \mid Category) * \quad (5)$$

Where Pr(Category) is the overall frequency of the category, ie, it is the probability that a randomly selected document (tweet) will be in this category (positive or negative), so its the number of documents in the category divided by the total number of documents. Pr(tweet) is a scale factor and has the same value for both categories. Whichever category gets a higher score for $Pr(Category \mid tweet)$ is the predicted category (positive, or negative).(3 marks)

Total marks: 25

SEE NEXT PAGE

4. This question is about *Machine Learning and Optimisation*. Design a system using k-Nearest Neighbour (kNN) algorithm that is able to learn to classify images obtained through a surveillance camera in an underground station, so that the system understands whether the images contain any human passenger or not. 5000 images labelled as "has passengers" or "empty station" are first collected and available as training data for designing such system. Another 5000 images from the surveillance videos, without any labels, are for testing. Answer the questions and complete the descriptions as below.

   A. Describe your choice of the features to be used in the system. [4 marks]

   B. Describe your designed kNN system. [8 marks]

   C. If you are unsure whether your choice of features and k is effective, it is possible to improve it using optimisation techniques. Describe the details of your approach for such optimisation process. This should include

      (i). your design of an objective function, [5 marks]

      (ii). and the key steps of your optimisation algorithm. [8 marks]

A. features: Histogram of Oriented Gradients (HOG) can be used as features. Suitable parameters such as the number of pixels in each cell, the number of cells in each block, and the number of orientations in each cell should be identified.

B. Describe your designed system.

The features for each training sample f_i are extracted, where f_i is the feature vector and $i = 1, 2, ..., 5000$. For each testing sample, its features f_q are also extracted in the same way. The Euclidean distance between f_q and each f_i is calculated. There should be 5000 distances, which are then ranked in order from small to large. The training samples with top k shortest distances are chosen, where k is a pre-defined value such as 7, or 15. These samples are k-nearest neighbours for the testing sample. The majority class in the k-nearest neighbours should be the class for the testing sample.

C. Describe the objective function and the optimisation algorithm.

(i) Objective function can be measured using a set of training samples with ground truth. Run through kNN to check the classification results. The objective function is to calculate the number of time that kNN correctly classifies the images. The score indicates the quality/cost of the kNN with its current features and the choice of k.

SEE NEXT PAGE

(ii) Genetic Algorithm (or other optimization algorithm)can be used. The chromosome for each solution can be the parameter setup of the HOG features, e, g, the number of cells in each block, the cell size and the number of orientations in each cell, and k value for kNN. (3 marks) A number of solutions can form the initial population, then for each solution, the set of testing samples are used to get scores through the objective function. The solutions are then ranked. Top solutions are selected based on predefined percentage of such selection or a subset of solutions are selected based on Roulette wheel selection process. Such subset of solutions will be used in the next generation.(3 marks) Some new solutions are generated through crossover and mutation of the top solutions, as well as those from random generation, and these make up the full set of solutions for the next generation. The objective function will apply to the next generation again and the whole process repeats until a satisfactory the top score cannot improved any more. (2 marks)

Total marks: 25

END OF PAPER

INTERNAL EXAMINER: DR. H.L.TANG
EXTERNAL EXAMINER: PROF. DAVID MARSHALL