

Name:

Student number:

COMP9417 Machine Learning and Data Mining
Mid-session Examination:
SAMPLE QUESTIONS

Your **Name** and **Student number** must appear at the head of this page.

Duration of the exam: 1 hour.

This examination has **five** questions. Answer **all** questions.

Total marks available in the exam: 40.

Multiple-choice questions may require **more than one** answer.

Show all working in your script book.

This page intentionally left blank.

Question 1 [Total marks: 10]

Supervised Learning – Regression

A) [2 marks] Variance measures the *spread* of values of some random variable X around its mean $E(X)$. We remember variance can be expressed as the “mean of the squares minus the square of the mean”, but which of these definitions of variance is the correct one?

- (1) $E(X^2 - E(X))$
- (2) $E(X - E(X))^2$
- (3) $E(X^2 - E(X))^2$
- (4) $E(E(X^2) - E(X))$
- (5) $E(E(X^2) - E(X))^2$

B) [2 marks] If, for some estimator, Mean Squared Error (MSE) approaches zero as sample size increases, then the estimator is said to be:

- (1) correct
- (2) complete
- (3) consistent
- (4) unbiased
- (5) biased

C) [2 marks] Covariance of two random variables x, y is determined in relation to their differences from their respective means \bar{x}, \bar{y} . Covariance is observed when, for all instances x_i, y_i of the random variables:

- (1) $x_i < \bar{x}, y_i > \bar{y}$ or $x_i < \bar{x}, y_i < \bar{y}$
- (2) $x_i < \bar{x}, y_i < \bar{y}$ or $x_i < \bar{x}, y_i > \bar{y}$
- (3) $x_i > \bar{x}, y_i > \bar{y}$ or $x_i > \bar{x}, y_i < \bar{y}$
- (4) $x_i < \bar{x}, y_i < \bar{y}$ or $x_i > \bar{x}, y_i > \bar{y}$
- (5) $x_i > \bar{x}, y_i < \bar{y}$ or $x_i > \bar{x}, y_i > \bar{y}$

D) [2 marks] Which of the following statements about the correlation of two random variables x, y is true?

- (1) positive correlation between x and y means x causes y
- (2) zero correlation between x and y means x has no relationship with y
- (3) negative correlation between x and y means x has no relationship with y
- (4) non-zero correlation between x and y means x and y have some relationship
- (5) correlation of r between x and y means $y = r \times x$

E) [2 marks] Which of the following do you consider to be correct statements ?

- (1) linear regression can fit non-linear dependencies of y on \mathbf{x} if the parameters \mathbf{w} are non-linear
- (2) linear regression cannot fit non-linear dependencies of y on \mathbf{x}
- (3) linear regression can fit any dependency of y on \mathbf{x} using logarithmic transformations of \mathbf{x}
- (4) linear regression can fit any dependency of y on \mathbf{x} using polynomial transformations of \mathbf{x}
- (5) linear regression can fit linear dependencies of y on non-linear transformations of \mathbf{x}

Question 2 [Total marks: 6]

Nearest neighbour classification

Under what conditions, if there are any, does the nearest neighbour algorithm do linear classification ?

HINT: suppose for a two-class problem there are exactly two exemplars, one for each class. Suppose further that you are just using the nearest neighbour classification algorithm, i.e., k -NN where $k = 1$. If this algorithm is using Euclidean distance, what will the decision boundary look like ? Is this the same if Manhattan distance is used ? Explain your answer.

Question 3 [Total marks: 10]

Decision Tree Learning

The table below contains a sample S of ten examples. Each example is described using two Boolean attributes A and B . Each is labelled (classified) by the target Boolean function.

Id	A	B	Class
1	1	0	+
2	0	1	-
3	1	1	-
4	1	0	+
5	1	1	-
6	1	1	-
7	0	0	+
8	1	1	+
9	0	0	+
10	0	0	-

A) [2 marks] What is the entropy of these examples with respect to the given classification ?

- B) [3 marks] What is the information gain of attribute A on sample S above ?
- C) [3 marks] What is the information gain of attribute B on sample S above ?
- D) [2 marks] Which would be chosen as the “best” attribute by a decision tree learner using the information gain splitting criterion ? Why ?

Question 4 [Total marks: 4]

Naive Bayes Classification

Suppose for a two-class classification problem you have m Boolean features. How many probabilities will you have to estimate from your training data ? Show your working.

Question 5 [Total marks: 10]

Perceptrons

A) Let the weights of a two-input perceptron be: $w_0 = -0.2$, $w_1 = 0.5$ and $w_2 = 0.5$. Assuming that $x_0 = 1$, what is the output of the perceptron when:

[i] [1 mark] $x_1 = 0$ and $x_2 = 0$?

[ii] [1 mark] $x_1 = 0$ and $x_2 = 1$?

Letting $w_0 = -0.7$ and keeping $x_0 = 1$, $w_1 = 0.5$ and $w_2 = 0.5$, what is the perceptron output when:

[iii] [1 mark] $x_1 = 1$ and $x_2 = 0$?

[iv] [1 mark] $x_1 = 1$ and $x_2 = 1$?

[v] [2 marks] What effect has changing the bias weight had ? Has this changed the Boolean function that the perceptron implements ?

B) [4 marks] Apply the perceptron training algorithm with $\eta = 1$ to the perceptron model following [ii]. That is, weight vector $\mathbf{w} = (-0.2 \ 0.5 \ 0.5)$ and the example $\mathbf{x} = (1 \ 0 \ 1)$. Does this give the weight vector at [iii] ? Explain what has happened.