1. What is a soft classifier?
   a) A classifier that can predict more than two classes.
   b) A classifier that outputs an estimation of reliability (or probability) for each class.
   c) A classifier that can predict numeric values and be used as a regression model.
   d) A classifier that can predict an infinite number of classes.

2. Which kind of task is this?
   "Determine the probability that a student drops off in the following year"
   a) Regression.
   b) Association
   c) Classification.
   d) Clustering

3. Which kind of task is this?
   "Determine what kinds of books I read, according to their genre, length, language and many other features"
   a) Classification.
   b) Regression.
   c) Correlation
   d) Clustering

4. Which kind of task is this?
   "Determine what combinations of pathologies old people have "
   a) Association
   b) Classification.
   c) Correlation.
   d) Clustering

5. Given the following exact relation between variables: $x_1 = 3.2x_2 - 5.2$ and $x_3 = 5.4x_4 + 2.2$.
   a) The correlation between x_1 and x_2 is higher than the correlation between x_3 and x_4.
   b) The correlation between x_1 and x_2 is lower than the correlation between x_3 and x_4.
   c) The correlation between x_1 and x_2 is equal to the correlation between x_3 and x_4.
   d) We cannot know the correlation of these variables, only the slope between them.

6. When should we use cross-validation?
   a) Always, it comes by default with many libraries.
   b) When we have a small number of examples.
   c) When we have a large number of examples.
   d) Never, it breaks the golden rule of evaluation.

7. Which kind of task is this?
   "Determine how long an offer was active according to the sales of the product"
   a) Correlation
   b) Classification.
   c) Regression.
   d) Clustering

8. What is collaborative filtering?
   a) The recommendations are produced by observing the preferences of similar users.
   b) The recommendations are produced by observing the user's most similar items.
   c) The recommendations are produced by observing the characteristics of the items (shape, price, colour, etc.).
   d) The recommendations are produced by observing the characteristics of the users (age, gender, etc.).

9. What is the k in k-nearest neighbours (kNN)?
   a) The number of groups, as in k-means.
   b) The number of layers, as in ANN.
   c) The number of nearest neighbours to compare with.
   d) The number of kernels, as in SVM.

10. Which of the following can NOT be used for regression?
    a) Linear regression.
    b) Logistic regression.
    c) Neural networks.
    d) Non-linear regression.

ASSESSMENT
Answer Sheet (MODEL B)

Surname:	Name:
Group in English: ⊠	

In the following table, circle the correct answer for each question.

Question	Answer			
1	a	**b**	c	d
2	a	b	**c**	d
3	a	b	c	**d**
4	**a**	b	c	d
5	a	b	**c**	d
6	a	**b**	c	d
7	a	b	**c**	d
8	**a**	b	c	d
9	a	b	**c**	d
10	a	**b**	c	d

The result will be calculated by the statistical correction formula:

$$(Right - Wrong/3) \times 1$$

which discounts the probability of getting a right answer by chance on a question with four possibilities.

The mark is between 0 and 10.

Remember that this assessment is just 10% of the final qualification for the course.