6CCS3ML1 (Machine Learning)

Tutorial 01

(Version 1.5)

1. Give an example of the following types of machine learning:

   (a) Supervised learning

   (b) Unsupervised learning

   For each example, give the feature set, and explain the aim of the learning.

   To explain: in the Iris example, the features are measurements that describe each flower (petal area and sepal area, for example), and the task is to identify which species a given Iris flower comes from.

   Your example should not be one that was mentioned in the lecture.

2. For each of the following machine learning tasks:

   (a) Binary classification

   (b) Multiclass classification

   (c) Clustering

   (d) Regression

   For each example, give the feature set, and for explain the aim of the learning task.

   Your example should not be one that was mentioned in the lecture.

3. (a) Explain how a k-nearest neighbour classifier is trained.

   (b) Explain how a k-nearest neighbour classifier is used to classify a new example.

   (c) For a given set of data, what difference in performance would you expect to see between a 1-nearest neighbour classifier and a 5-nearest neighbour classifier?

4. The data in the table below are the results of testing a binary classifier:

Example	Classifier output	Correct class	Example	Classifier output	Correct class
E01	1	1	E11	1	1
E02	1	1	E12	1	0
E03	1	0	E13	1	1
E04	1	1	E14	0	1
E05	0	0	E15	0	0
E06	0	0	E16	0	0
E07	0	1	E17	1	1
E08	0	0	E18	1	1
E09	1	1	E19	0	1
E10	0	0	E20	0	0

Given this data:

(a) Calculate the accuracy of the classifier.

(b) Write down the confusion matrix.

(c) What does the confusion matrix tell you about the classifier?

(d) Calculate the precision of the classifier.

(e) Calculate the recall of the classifier.

(f) Calculate the F1 score of the classifier.

For each question you should explain the reasoning behind your answer.

5. A (simple) 3-nearest neighbour classifier includes the following examples:

Instance	Features		Class
	x_1	x_2	
X_1	1	1	C_1
X_2	2	2	C_1
X_3	1	3	C_1
X_4	4	2	C_1
X_5	1	6	C_2
X_6	2	4	C_2
X_7	2	5	C_2
X_8	3	4	C_2
X_9	5	4	C_2

The distance metric used by this classifier is Manhattan distance, which, for two examples i with features (x_1^i, x_2^i) and j with features (x_1^j, x_2^j) is:

$$distance = |x_1^i - x_1^j| + |x_2^i - x_2^j|$$

where $|a|$ is the absolute value of a.

Which class would this classifier allocate to:

(a) An instance with $x_1 = 3$ and $x_2 = 1$?

(b) An instance with $x_1 = 4$ and $x_2 = 5$?

(c) An instance with $x_1 = 2$ and $x_2 = 3$?

(d) An instance with $x_1 = 4$ and $x_2 = 3$?

For each example calculate the probability of the example being classified in both classes using the "sophisticated" method from the slides.

Version list

- Version 1.0, January 19th 2019.
- Version 1.1, January 24th 2019.
  - Corrected typos
- Version 1.2, February 1st 2019.

  - Added an explanation of Manhattan distance in question 5

- Version 1.3, January 16th 2020.
  - (Updated module number for 2020 onwards).
- Version 1.4, January 8th 2021.
- Version 1.5, January 20th 2021.
  - Fixed indices under question 5.