

# 1 Introduction

## 1. Give a definition of “Pattern Recognition”.

*Pattern Recognition is concerned with creating algorithms that can assign names to observations.*

*Or,*

*Pattern Recognition is concerned with making decisions based on data.*

## 2. Which of the following problems is a suitable application for pattern recognition?

### a. Classifying numbers into primes and non-primes.

*No. Can be more easily and more accurately performed analytically.*

### b. Detecting potential fraud in credit card charges.

*Yes.*

### c. Determining where a cannon ball is likely to land given information about elevation and direction of the barrel, wind direction, etc.

*No. Better to do it analytically (e.g. using Newton's equations).*

### d. Identifying the species a newly discovered “bug” belongs to.

*Depends! Some species are defined by simple rules (female mammals produce milk, etc.), so given the right data might easily simply write down the solution. Given other data, like images of different bugs, we would need to perform pattern recognition.*

## 3. Give brief definitions of the following terms:

### a. Exemplar

*a particular datapoint which is represented by a feature vector (also called an item, sample, instance,...).*

### b. Dataset

*the collection of feature vectors for all exemplars.*

### c. Generalization

*how well a model performs on new data.*

### d. Overfitting

*making the model so specific to the training data that it fails to generalise to new data.*

### e. Decision Theory

*methods for making decisions that reduce cost rather than misclassification rate.*

### f. Feature Space

*the (multidimensional) space defined by the feature vectors in the dataset.*

**g. Linearly Separable**

*exemplars from two classes can be separated by a hyperplane in feature space.*

**h. Dichotomizer**

*a classifier that places exemplars in one of two classes (also called a binary classifier).*

**i. Hyper-Parameter**

*a value used by the learning algorithm in its search for the optimal parameters of the classifier.*

**j. Grid search**

*a method of trying to find suitable hyper-parameters that searches all possible combinations of values within defined ranges.*

**k. Training data**

*the collection of feature vectors used by the learning algorithm to tune the parameters of the classifier.*

**l. Test data**

*the collection of feature vectors used by to evaluate the performance of the trained classifier (this dataset should be distinct from the training data to ensure generalisation).*

**4. What types of learning best describe the following three scenarios, in which a coin classification system is being created for a vending machine.**

- a. Measurements of a large number of coins are taken. The algorithm finds that these measurements fall in several “bins”. It finds decision boundaries that separate these bins and uses these to classify new coins.**

*unsupervised learning (clustering)*

- b. The measurements of each coin is presented to the classifier which makes a decision. The classifier changes its decision boundaries based on whether this decision was correct or incorrect.**

*reinforcement learning*

- c. Measurements of a large number of coins are taken. The algorithm uses this data, together with known class labels, to infer decision boundaries which it then uses to classify new coins.**

*supervised learning (classification)*

**5. Briefly explain the following types of learning method:**

**a. Classification**

*A method that learns to predict a class label associated with each exemplar.*

**b. Regression**

*A method that learns to predict a continuous value for each exemplar.*

**c. Semi-supervised**

A method that learns using both labelled and unlabelled training exemplars.

**d. Transfer**

A method that pre-trains a classifier on another task before training it on the main task in the hope that the pre-training will help improve performance on the main task.

6.	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Class	Features	Class	Features	Class	Features	Class	Features
	1	5	1	5.5	1	(5,4)	1	(5.3,4)
	2	2	2	2.3	2	(2,9)	2	(2.3,9.1)
	1	4	1	4	1	(4,3)	1	(4,3)
	1	7	1	7	1	(7,4)	1	(7,4.1)
	2	1	2	1.8	2	(1,5)	2	(1.8,5.5)

Identify which of the above datasets are:

**a. univariate-continuous**

*dataset 2*

**b. multivariate-discrete**

*dataset 3*

**c. multivariate-continuous**

*dataset 4*

**d. univariate-discrete**

*dataset 1*

7. A classifier is designed to determine if a feature vector is or is not in a certain class. The output produced by the classifier is 1 if the sample is predicted to be in the class, and 0 otherwise. The following table shows the feature vectors for the samples in the test set, along with the class labels predicted by the classifier and the true class labels of each sample.

Features	Predicted Class	True Class
(5.3,4)	1	1
(2.3,9.1)	0	1
(4,3)	1	0
(7,4.1)	1	1
(1.8,5.5)	0	0
(6,3.1)	1	1
(4.5,3.5)	0	1

Draw a confusion matrix for this data.

		predicted label	
		true	false
true label	true	3	2
	false	1	1

8. For the results given in the previous question calculate the following performance metrics:

a. the error-rate

$$= \frac{FP+FN}{TP+TN+FP+FN} = \frac{1+2}{3+1+1+2} = \frac{3}{7}$$

b. the accuracy

$$= \frac{TP+TN}{TP+TN+FP+FN} = \frac{3+1}{3+1+1+2} = \frac{4}{7}$$

c. the recall

$$= \frac{TP}{TP+FN} = \frac{3}{3+2} = \frac{3}{5}$$

d. the precision

$$= \frac{TP}{TP+FP} = \frac{3}{3+1} = \frac{3}{4}$$

e. the  $f_1$ -score

$$= \frac{2 \times TP}{2 \times TP + FP + FN} = \frac{2 \times 3}{2 \times 3 + 1 + 2} = \frac{6}{9} = \frac{2}{3}$$

$$\text{alternatively: } f_1\text{-score} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} = \frac{2 \times \frac{3}{5} \times \frac{3}{4}}{\frac{3}{5} + \frac{3}{4}} = \frac{2 \times \frac{9}{20}}{\frac{12}{20} + \frac{15}{20}} = \frac{2 \times 9}{27} = \frac{2}{3}$$

9. The following table shows exemplars from two classes. Sketch the feature space, and suggest a suitable location for a decision boundary that will minimise the number of mis-classifications. If the cost of erroneously choosing class 1 is higher than the cost of erroneously choosing class two, how will this effect the location of the decision boundary?

Class	Features
1	(3,4)
1	(4,7)
1	(7,5)
2	(2,5)
2	(3,9)

