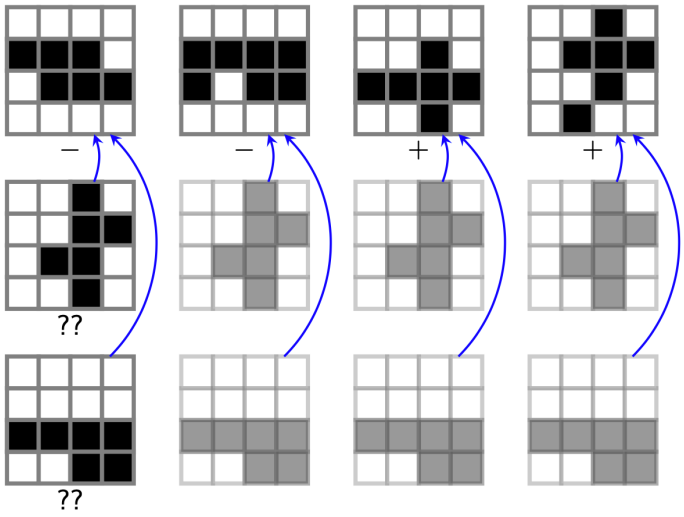
Cambridge AI+: Exercises for Supervision 1 (k-NN and Regression)

Question 1. *In this question, we will apply 2-NN to a small problem coming from image classification.*



We are given a database of 4 images, two are labelled − and two are labelled +. Classify the two new images using the 2-NN rule. Use the hamming distance (number of differently coloured cells) as the distance function.

Question 2. *Consider the following data set, where the second column is the only feature and the third column is the label:*

Student	*TV hours*	*Exam Grade*
1	2	8
2	4	5
3	3	5.5
4	5	4
5	6	4.5
6	8	2
7	3	6.5
8	5.5	5
9	7	1.5
10	9	3

Apply the 2-NN rule to this data set and sketch the function you obtain. What is your predicted exam grade for a new student with (i) 0 TV hours and (ii) 6.5 TV hours. Discuss the results.

Question 3 ((advanced, requires programming)). *The following represents a 2-dimensional data set with two numerical features and two binary classes.*

Point i	(x_1^i, x_2^i)	(y^i)
1	$(2.78, 25.5)$	-1
2	$(1.46, 23.6)$	-1
3	$(3.39, 44.4)$	-1
4	$(1.38, 18.5)$	-1
5	$(3.06, 30.1)$	-1
6	$(7.62, 27.9)$	$+1$
7	$(5.33, 20.8)$	$+1$
8	$(6.93, 17.7)$	$+1$
9	$(8.67, -2.4)$	$+1$
10	$(7.67, 35.1)$	$+1$

1. *Do you recognise any issues about the data set? If so, apply some preprocessing before proceeding to the next part of the question.*

2. *Consider now the unlabelled point*

$$\mathbf{x} = (x_1, x_2) = (4.41, 25.0),$$

which we would like to classify. Perform a k-nearest neighbour search for $k = 1, 3, 5$ and Use majority vote to determine if x should get label -1 or $+1$. If you like, you can also experiment with different distance functions.

Question 4. *Consider the following data set, which records several phone calls of clients to a company.*

Client	*Length of Call in minutes*	*Number of Purchased Items*
1	13	2
2	17	1
3	32	2
4	36	4
5	45	4
6	53	3
7	67	6
8	74	5
9	83	7
10	132	11

1. *Fit the data to a linear regression model and compute the R^2-value.*
2. *Consider a client that calls the company for 90 minutes. Can you estimate the number of purchased items?*
3. *Somebody tells you that a recent client purchased 3 items, but you do not know the length of the call. Can you estimate that number?*

Question 5. *A medical study examined 300 people with high blood pressure and 200 people with low blood pressure. During the period of the study, 30 of these people in the low-blood-pressure group and 100 in the high-blood pressure group suffered from cardiovascular disease (heart disease).*

1. *Sketch how the data set looks like (i.e., explanatory variables and outcomes).*
2. *Apply logistic regression.*
3. *Consider now the classification problem. How would you classify a person with high-blood-pressure? How could you try to refine your prediction?*