COMP7990 (Section 3 and Section 4)

Quiz

Write all your answers in your answer books

Question 1 (30 Marks)

1.1 **(10 Marks)** What are new value range after normalization by decimal scaling? Justify your answer.
Answer: (-1, 1). The values are normalized by moving the decimal point such that the maximal absolute values of the new mapped features is smaller than 1.

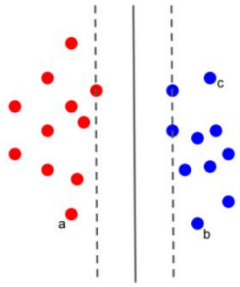
1.2 **(10 Marks)** In real-world data analytic problems, data with missing values for some attributes is often occurred. Why missing values are common in real-world application? Describe various methods for handling missing values.
Answer: Missing data may be due to: (1) Equipment malfunction; (2) Data is not entered; (3) Certain data may not be considered at the time of data collection; (4) Not applicable for some cases
Possible methods: (1) Ignore the data sample with missing values; (2) Ignore attributes with missing values; (3) Fill in missing value by a global constant or Attribute mean/Median/mode or using a model to predict the missing value (data imputation)

1.3 **(10 Marks)** Suppose we have the following values for prices (already sorted in increasing order): [5, 10, 15, 21, 21, 22, 25, 28, 30, 31, 31, 32]. Use smoothing by bin means to smooth the above data using equal-depth binning with bin depth of 4. Illustrate your steps.
Answer:
1. Partition data into bins
Bin 1: [5, 10, 15, 21]
Bin 2: [21, 22, 25, 28]
Bin 3: [30, 31, 31, 32]
2. Computing the mean of each bin:
Bin1: 12.75
Bin2: 24
Bin 3: 31
3. Smoothing by mean
[12.75, 12.75, 12.75, 12.75, 24, 24, 24, 24, 31, 31, 31, 31]

Question 2 (40 Marks)

2.1 (**10 Marks**) In *K*-means algorithm and *K*-nearest neighbor algorithm, which one can be used for classification problem? And why?
Answer: K-nearest neighbor algorithm can be used for classification problem. K-means is a clustering algorithm.

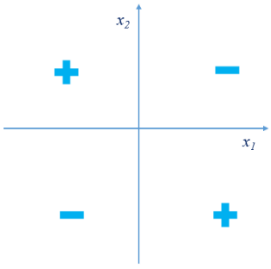
2.2 (**10 Marks**) Consider the following figure that shows the decision boundary of SVM on a dataset.



What are the α_i value for data points a, b, c? And why?

Answer: The α_i values for a, b, c are all zeros because a, b, c are not support vectors.

2.3 (**10 Marks**) Consider the following data set



What is the training error if we applied 3-nearest neighbor algorithm on this data? Justify your answer.

Answer: The training error is 100% (i.e., training accuracy is 0%). For each data point, the three nearest neighbors are itself and other two points from the other class. Therefore, the classification label for each training data point will be wrong. The training error is 100%.

2.4 **(10 Marks)** What is the purpose of minimizing $\frac{\|\mathbf{w}\|^2}{2}$ in the SVM optimization problem?

Answer: maximize the margin

Question 3 (30 Marks)

Consider the following perceptron,

$$f(\mathbf{x}) = \begin{cases} 1 \text{ if } (w_0 x_0 + w_1 x_1 + w_2 x_2) > 0 \\ -1 \text{ otherwise.} \end{cases}$$

3.1 **(10 Marks)** Suppose the current model parameters are $w_0 = 1$, $w_1 = 1$, $w_2 = -1$. Is the following data sample correctly predicted? What are the new model parameters after updating model based on this data sample?

x_0	x_1	x_2	y
1	1	1	1

Answer: Since $w_0x_0 + w_1x_1 + w_2x_2$=(1)*1+(1)*1+ (-1)*1=1>0, *f(***x***) is 1 and the true label is* 1. *Therefore, this sample is correctly predicted. We do not need to update the model. The new model parameters are still* $w_0 = 1$, $w_1 = 1$, $w_2 = -1$.

3.2 **(10 Marks)** Suppose the current model parameters are $w_0 = 1$, $w_1 = 1$, $w_2 = -1$. Is the following data sample correctly predicted? What are the new model parameters after updating model based on this data sample?

x_0	x_1	x_2	y
1	−1	1	1

answer: *Since* $w_0x_0 + w_1x_1 + w_2x_2$=(1)*1+(1)*(-1)+(-1)*1=-1<0, *f(***x***) is* -1 *but the true label is* 1. *Therefore, this sample is NOT correctly predicted. We need to update the model. Based on the updating rule of perceptron algorithm, the new model parameters are:*

$w_0 = w_0 + y*x_0 = 1+1*1 = 2$

$w_1 = w_1 + y*x_1 = 1+1*-1 = 0$

$w_2 = w_2 + y*x_2 = -1 + 1*1 = 0$

3.3 (**10 Marks**) Describe two limitations of perceptron algorithm. And how to address to these two limitations.

Answer:

(1) Do not converge if the data is linearly un-separable.

(2) Can not solve nonlinear separable problem.

(3) If the data is linear separable, it finds one of many possible hyperplanes separating the data without considering the maximum marge principle.