Quiz4 --- KNN, Logistic Regression (Logit), Association Analysis and Miscellaneous

Due Dec 2 at 11:59pm **Points** 35 **Questions** 10
Available until Dec 9 at 11:59pm **Time Limit** None
Allowed Attempts Unlimited

This quiz was locked Dec 9 at 11:59pm.

Attempt History

	Attempt	Time	Score
KEPT	Attempt 4	5 minutes	35 out of 35
LATEST	Attempt 4	5 minutes	35 out of 35
	Attempt 3	2 minutes	22 out of 35 *
	Attempt 2	less than 1 minute	7 out of 35 *
	Attempt 1	33 minutes	15 out of 35 *

* Some questions not yet graded

Score for this attempt: **35** out of 35
Submitted Nov 23 at 3:14pm
This attempt took 5 minutes.

Question 1

3 / 3 pts

Which of the following statements is true about the k-NN algorithm?

1- k-NN performs much better if all of the data have the same scale.

2-k-NN works well with a small number of features (X's) however, struggles when the number of inputs is very large.

3-k-NN makes no assumptions about the functional form of the problem solved.

1 and 2

Correct!

All of the above

1 and 3

Only 1

Question 2

3 / 3 pts

Which of the following distance measure do we use in case of categorical variables in k-NN?

1)Hamming Distance
2)Euclidean Distance
3)Manhattan Distance

2

1, 2 and 3

3

1 and 2

2 and 3

Correct!

1

Question 3

2 / 2 pts

Logistic regression assumes a:

Linear relationship between continuous predictor variables and the outcome variable.

Linear relationship between observations.

Correct!

Linear relationship between continuous predictor variables and the logit of the outcome variable.

Linear relationship between continuous predictor variables.

Question 4

3 / 3 pts

Regarding bias and variance, which of the following statements are true?

Correct!

Models which overfit have a low bias and underfit have a low variance

Models which overfit have a low bias and underfit have a high variance

Models which overfit have a high bias and underfit have a low variance

Models which overfit have a high bias and underfit have a high variance

Question 5

2 / 2 pts

Adding more basis functions in a linear model:

Decreases estimation bias

Correct!

Decreases model bias

Doesn’t affect bias and variance

Decreases variance

Question 6

2 / 2 pts

Choose which data mining task is the most suitable for the following scenario: diagnosing the level of flood severity.

Prediction

Correct!

Association rules

Classification

Anomaly detection

Question 7

5 / 5 pts

For the question given below consider the data Transactions :

I1, I2, I3, I4, I5, I6
I7, I2, I3, I4, I5, I6
I1, I8, I4, I5
I1, I9, I10, I4, I6
I10, I2, I4, I11, I5
With support as 0.6 find all frequent itemsets?

<I2>, <I4>, <I5>, <I2, I4>, <I2, I5>, <I4, I5>, <I2, I4, I5>

<I1>, <I4>, <I5>, <I6>

<I11>, <I4>, <I5>, <I6>, <I1, I4>, <I5, I4>, <I11, I5>, <I4, I6>, <I2, I4, I5>

Correct!

<I1>, <I2>, <I4>, <I5>, <I6>, <I1, I4>, <I2, I4>, <I2, I5>, <I4, I5>, <I4, I6>, <I2, I4, I5>

Question 8

8 / 8 pts

Describe how you would correct for zero probabilities or solve zero frequency problems in Naive Bayes? (Explain in detail)

Your Answer:

One thing to keep in mind is that if the conditional probability for any of the attributes is zero, then the entire expression for the class-conditional probability becomes zero. Zero probabilities can arise when the number of training instances is small and the number of possible values of an attribute is large. To address this problem, it is important to use other conditional probability estimates that are not as brittle as simply using fractions of training instances. We can achieve this by using the alternative estimates of conditional probability known as Laplace estimate and m-estimate. Note that even if the number of instances in the class equals 0 for the original estimating formula, both Laplace and m-estimate provide non-zero values of conditional probabilities.

Question 9

2 / 2 pts

The inference made by an association rule does not necessarily imply causality.

Correct!

True

False

Question 10

5 / 5 pts

In naive bayesian classification, describe the main difference between how discrete categorical and continuous numerical outcomes are handled. Only provide the main differences.

Your Answer:

For categorical features, the estimation of P(Y|X) is simply:

P(Y|X) = $\frac{P(X|Y)P(Y)}{P(X)} = \frac{P(X,Y)}{P(X)}$

However, one issue is that if some feature values never show, their likelihood will be zero, which makes the entire expression zero. You fix this with Laplace or m-estimates.

For continuous features, you can handle them with discretization,which involves partitioning the range of data into discrete bins. You could also assume a certain form of probability distribution for the continuous variable and estimate the parameters of the distribution using the training data. For instance, one can use a Gaussian distribution to represent the conditional probability of continuous attributes.

Quiz Score: **35** out of 35