Machine Learning - Lecture 6: Knowledge test

Chris Thornton

October 20, 2011

Question

Newspapers sometimes rank universities in terms of numbers of applicants. What is the explicit structure of the data? Suggest some possible forms of implicit structure.

Question

What is the difference between Euclidean and city-block distance?
How can we choose between them in a particular application?

Question

What benefits might be obtained by normalizing the values of a discrete variable? How could the normalization be accomplished?

Question

What is the difference between a condition*ing* and a condition*ed* value in a defined probability?

Question

Where we have just one class, and one attribute variable, we can work out all conditional probabilities directly from the dataset. Why is this more difficult with more than one attribute?

Question

In a visual display of k-means clustering (with data based on two numeric variables), it appears as if the centroids repel each other. Explain this effect.

Question

How can we obtain predictions from a set of centroids that have been obtained using the k-means algorithm? Specify a reasonable decision rule.

Question

Let's say we use k-means for predicting classifications, but find that with only n means in play, predictions are often wrong. Are we bound to improve prediction performance if we add one more mean (i.e., one more centroid)?

Question

The following data describe individuals in terms of occupation, symptom and ailment.

SYMPTOM	OCCUPATION	AILMENT
sneezing	nurse	flu
sneezing	farmer	hayfever
headache	builder	concussion
headache	nurse	flu
sneezing	teacher	flu
headache	teacher	concussion

Work out the probability of one of these indviduals being a nurse.

Work out the conditional probability of someone having concussion given they're a builder, and then the conditional probability of a concussed builder having a headache.

Use Bayes rule to get the probability that a sneezing builder has flu. Use the NBC to predict his/her probable ailment.