15-781 Midterm Example Questions

1 Short Answer

(a) **(True or False?)** If $P(A|B) = P(A)$ then $P(A \wedge B) = P(A)P(B)$.

(b) What is the entropy of the following probability distribution: $[0.0625, 0.0625, 0.125, 0.25, 0.5]$?

(c) **(True or False?)** Because decision trees learn to classify discrete-valued outputs instead of real-valued functions it is impossible for them to overfit.

(d) **(True or False?)** Assuming a fixed number of attributes, a Gaussian-based Bayes optimal classifier can be learned in time linear in the number of records in the dataset.

2 Decision Trees

You are stranded on a deserted island. Mushrooms of various types grow wildly all over the island, but no other food is anywhere to be found. Some of the mushrooms have been determined as poisonous and others as not (determined by your former companions' trial and error). You are the only one remaining on the island. You have the following data to consider.

Example	IsHeavy	IsSmelly	IsSpotted	IsSmooth	IsPoisonous
A	0	0	0	0	0
B	0	0	1	0	0
C	1	1	0	1	0
D	1	0	0	1	1
E	0	1	1	0	1
F	0	0	1	1	1
G	0	0	0	1	1
H	1	1	0	0	1
U	1	1	1	1	?
V	0	1	0	1	?
W	1	1	0	0	?

You know whether or not mushrooms A through H are poisonous, but you do not know about U through W. For the first couple of questions, consider only mushrooms A through H.

(a) What is the entropy of IsPoisonous?

(b) Which attribute should you choose as the root of a decision tree? Hint: You can figure this out by looking at the data without explicitly computing the information gain of all four attributes.

(c) What is the information gain of the attribute you chose in the previous question?

(d) Build a decision tree to classify mushrooms as poisonous or not.

(e) Classify mushrooms U, V, and W using this decision tree as poisonous or not poisonous.

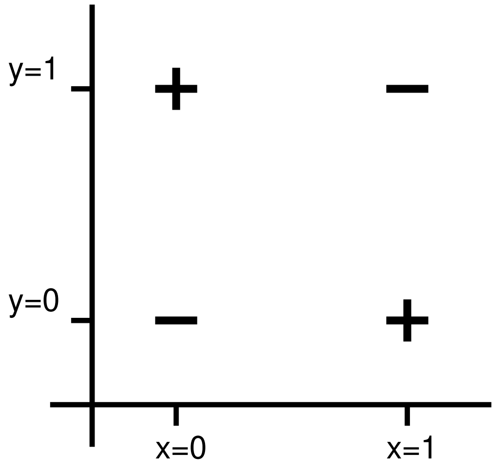
(f) If the mushrooms of A through H that you know are not poisonous suddenly became scarce, should you consider trying U, V, and W? Which one(s) and why? Or if none of them, then why not?

3 Gaussian Bayes Classifiers

1. Gaussian-based Bayes Classifiers assume that, given n classes, the kth datapoint was generated by first deciding the class of the kth datapoint according to the class prior probabilities, and then choosing the the kth input vector to be generated randomly by a Gaussian distribution with a mean and (usually) a covariance that is dependent on the choice of the class.

   Describe one or more ways in which this assumption could be wrong in practice.

2. Consider the XOR problem, in which there are two input attributes x and y which take on the values 0 and 1. The output class is positive if and only if $x \neq y$.

   What will happen if you try to train a Gaussian-based Bayes Classifier on such a dataset? Assume that the classifier is able to learn arbitrary covariance matrices.

3. Is it possible to construct a Bayes classifier for one input x so that when it is used it will predict

   - Class 1 if $x < -1$
   - Class 2 if $-1 < x < 1$
   - Class 1 if $1 < x$?

   If so, how?

3. Explain or sketch an example of a classification problem with two real-valued inputs and one discrete output in which...

3a: Gaussian Bayes Classifiers would do well on a training set but badly on a test set.

3b: Gaussian Bayes Classifiers would do well on a test set but decision trees would do badly on a test set.

3c: Gaussian Bayes Classifiers would do badly on more than one third of the test set but decision trees would do nearly perfectly on a test set.

4. PDFS: Give an example of a probability density over a single real-valued variable in which

p(x) > 0 for all x

P(X == 0) = 0

E[X] = 0

P(X == 1) > 0

5. You can expect a question like the Bayesian Gaussian MAP estimation ‘‘intellectual snobs’’ example.

6. A ''Box'' distribution of a scalar random variable is a PDF with two parameters: L and H (for LO and HIGH) in which

p(x) = 0 if x < L

p(x) = 1/(H-L) if L <= x <= H

p(x) = 0 if x > H

We'll use the notation X ~ BOX(L,H) to mean that X is a random variable drawn from a Box distribution with parameters L and H

6a: If X ~ BOX(L,H) what is E[X]?

6b: If X ~ BOX(L,H) what is Var[X]?

6c: Write P(x < q) as an if-then-else expression involving q, L and H

6d: Suppose you have data x_1 , x_2 , ... x_R i.i.d. ~ BOX(L,H) and suppose L and H are unknown. What are their MLE values? Explain. (Note this is a case where a careful few sentences explaining your answer may be better than an attempt at a proof by classic differentiation of log-likelihood)

7: Imagine you are going to learn a Naive Bayes classifier for the following data. Imagine you'll use the Box distribution described above for the real-valued parameter. Once you've learned the classifier, what is P(Happy=True | Occupation=Professor ^ Age=36) according to the classifier?

Inputs		Output
Age	Occupation	Happy
20	CTO	No
40	Prof	No
50	CTO	Yes
30	Prof	Yes
50	Prof	Yes

(Important: I have not tested this question so there may be problems I didn't expect or they may take longer than I expect. The real midterm exam *will* be tested)