COMP 5970/6970 HW 4: 5 questions 5 points 5% Credit
Due before 11:59 PM Friday March 29

Instructions:

1. This is an individual assignment. You should do your own work. Any evidence of copying will result in a zero grade and additional penalties/actions.
2. Enter your answers in this Word file. Submissions must be uploaded **as a single file** (Word or PDF preferred, but other formats acceptable as long as your work is LEGIBLE) to Canvas before the due date and time. Don't turn in photos of illegible sheets. **If an answer is unreadable, it will earn zero points.** Cleanly handwritten submissions (print out this assignment and write answers in the space provided, with additional sheets used if needed) scanned in as PDF and uploaded to Canvas are acceptable.
3. **Submissions by email or late submissions (even by minutes) will receive a zero grade.** No makeup will be offered unless prior permission to skip the assignment has been granted, or there is a valid and verifiable excuse.

Multiple Choice Questions (5 points)

In the following questions, circle the correct choice. If more than one answer is correct, circle all that apply. In those cases, partial credit will be given to partially correct answers. No explanation needed. Incorrect answers or unanswered questions are worth zero points.

1. "Naïve Bayes classifier can be used even when the class conditional distributions of the features do not follow Gaussian distributions, assuming all features are continuous-valued." The statement is:

[a] True
[b] False

Answer: [a]

2. "Naïve Bayes is the best possible classifier if the distribution of the data is known in advance." The statement is:

[a] True
[b] False

Answer: [b]

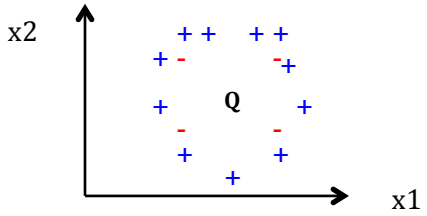
3. Suppose you are trying to approximate a function Z = X XOR Y using either Decision Tree or Naïve Bayes, and you know for a fact that X and Y are independent random variables and X, Y, Z are all Boolean. Which algorithm do you think might possibly perform better?

[a] Decision Tree
[b] Naïve Bayes
[c] Cannot say

Answer: [a]

X, Y might be independent variables, but a Naïve Bayes classifier will not model the function well since for a particular class (say, z = 0), X and Y are dependent; while a decision tree might.

4. Consider the following figure consisting of two features X1 and X2, and their corresponding class labels (+, -).



If you train a Naïve Bayes classifier on this data, what class would be predicted for the data point marked as 'Q'?

[a] +
[b] -
[c] Cannot say

Answer: [a]

Both classes have the same centroid, and + has slightly larger variance in each dimension, but the critical fact is that +'s class prior is much higher. So, we predict +.

5. Suppose we are using Gaussian Naïve Bayes for a dataset with a single attribute X and two class labels 0 and 1. Say that the parameters of the trained model are:

$x \mid y = 0 \sim N(0, 1/4)$
$x \mid y = 1 \sim N(0, 1)$
$P(y = 1) = 0.5$

What is the shape of the decision boundary of this Gaussian Naïve Bayes model?

[a] Linear
[b] Nonlinear
[c] Cannot say

Answer: [b]

The class conditional distributions have different variances. Hence the decision boundary is nonlinear.