COMPSCI 589 **Homework 1** Spring 2018

Instructions — Answer all questions below. Be concise and clear in your answers, and use technical terms and statistical concepts from class.

Deliverables — This report has only one deliverable: A single PDF file containing your name and the requested answers to all questions below. Please clearly identify each question with the corresponding number and letter. There is no need to repeat the text of the question. Handwritten diagrams are acceptable, but they must be legible or points will be deducted. Text and mathematical notation should be typewritten.

Uploading — When you complete the assignment, upload your report to the *Homework 1* assignment on Gradescope.com as a single PDF file. When you upload your report, identify the pages that correspond to each question. Failure to identify the pages with questions will result in point deductions.

Academic Honesty Statement — Your answers are required to be your own. You should cite any outside source you consult for assistance, beyond the textbook, lecture slides, and Wikipedia entries. *Copying* solutions from external sources (books, internet, etc.) or from other students is considered cheating. *Sharing* your solutions with other students is also considered cheating. *Posting* your answers to public repositories such as GitHub or Stack Overflow is also considered cheating. Any detected cheating will result in a grade of zero on the assignment for all students involved, and potentially a grade of F in the course.

1. **KNN classifiers** — In lecture, we discussed the *resubstitution estimator*, an estimator of error rate that uses *exactly the same set* of data instances to both train the classifier and to estimate its error rate. Describe concisely how you would expect the bias of a resubstitution estimate to change as the hyper-parameter K is varied in a K nearest neighbor classifier. (10 points)

2. **Naive Bayesian classifiers** — Naive Bayesian classifiers (NBCs) are directly derived from Bayes rule. In lecture, we showed the simplified form of the classification function for NBCs.

   a. Derive Bayes rule from probability axioms (5 points)

   b. Explain concisely why some terms in Bayes rule can be omitted from the classification function for NBCs. (10 points)

   c. Explain why some terms in Bayes rule can be omitted even if we want NBC to serve as a ranking classifier. (10 points)

   d. In practice, NBCs tend to push probability values to the extremes (nearly 0 or nearly 1). Use the form of the NBC classification function to explain why violations of conditional independence might produce this effect. (10 points)

3. Classification trees

a. In lecture, we described linear discriminant analysis as a *discriminative* model and methods such as naive Bayesian classifiers as an example of a *generative* model. State whether *classification trees* are discriminative or generative models, and briefly explain why using concepts from probability theory. (5 points)

b. Below is a small training set describing examples of the symptoms ($\boldsymbol{X}=\{C,R,T\}$) and diagnosis (Y) of an influenza-like disease. Based only on this training set, show the classification tree that would be constructed if we use misclassification rate as the local loss function. Assume an algorithm that constructs a classification trees using recursive partitioning and that it continues partitioning until all data at a leaf node have a single class. Draw the tree, including decision nodes, leaf nodes, and predicted class label for each leaf node. (20 points)

Diagnosis (Y)	Cough (C)	Red eyes (R)	Temperature (T)
T	F	T	98
T	F	T	103
T	T	F	105
T	T	F	98
T	F	T	98
F	F	F	98
F	F	F	98
F	F	F	98
F	F	F	100
F	F	F	103

4. **Bias and variance errors —** Assume that you learn a classification tree and that the leaf nodes estimate conditional probability distributions (rather than deterministic class labels). Consider the effects of each of the changes described below. State what effect the change would have on *both* the bias *and* variance error components. State whether the change would increase (INC), decrease (DEC), or have no effect (NO) on each of the two error components. Remember to provide an answer *both* bias and variance for each situation (e.g., INC / NO). (4 points each)

   a. Add randomly selected decision nodes to the bottom of the tree and re-estimate the probability distributions at the leaf nodes with the same data used to estimate the original probability distributions.

   b. Merge random pairs of adjoining leaf nodes and re-estimate the probability distributions at those leaf nodes with the same data used to estimate the original probability distributions.

   c. Double the amount of data for estimating the probability distributions of leaf nodes (without changing the structure of the tree).

   d. Systematically increase the estimated probability of a specific value of the class label by 10 percentage points and decrease the estimated probabilities of the remaining values of the class label by a total of 10 percentage points.

   e. Randomly change the probability distribution at each leaf node (renormalizing so they sum to one).

5. **ROC curves** — Given the results of the ranking classifier below, draw the corresponding ROC curve with points labeled by their quantitative values within ROC space. (10 points)

ID	Class	p(Class)
1	T	0.99
2	F	0.90
3	T	0.93
4	F	0.43
5	T	0.85
6	F	0.32
7	T	0.69
8	F	0.08
9	T	0.30
10	F	0.01