## STAT460 - Midterm

You may use a calculator. Remember, answer any yes/no question with both a yes/no and an explanation.

Lastly, don't interpret the amount of space given on the test as an indication of the length of the response I am expecting.

1. Look at the figures in Table 1. (Written answers should be placed in the space provided on the next page)
(a) Describe to me the concept behind the Bayes' rule. By concept I mean you don't need to use 'math' terms, but rather tell me in words what it is and why would we care
(b) For each plot, do the following: (i) draw what you think the Bayes' rule decision boundary, (ii) identify what type of classifier (LDA, QDA, or trees) would be the best choice, and (iii) draw that classifier's decision boundary. Be sure to defend your choices. Why did you pick each classifier? You won't be able to perfectly separate the data, so don't try.

(Fig. 1)

(Fig. 3)
Table 1: Figures for problem 1. The colors correspond to two classes (ie: the response has two levels). We have two covariates $X 1$ and $Y 1$.
(a)
(Fig. 1)
(Fig. 2)
(Fig. 3)
2. Let's explore a classification technique that we won't have time to talk about in class. The table below provides a training data set containing 6 observations, 3 covariates, and 1 response.

| Obs. | $X_{1}$ | $X_{2}$ | $X_{3}$ | $Y$ |
| :--- | ---: | ---: | ---: | :--- |
| 1 | 0 | 3 | 0 | Red |
| 2 | 1 | 0 | 0 | Red |
| 3 | 0 | 1 | 3 | Red |
| 4 | 0 | 1 | 2 | Green |
| 5 | -1 | 0 | 1 | Green |
| 6 | 1 | 0 | 1 | Red |

$K$-nearest neighbors (KNN) is a technique whereby we classify a point with whatever is the majority class of its $K$-nearest ${ }^{1}$ neighbors. Suppose we wish to use this data to make a prediction for $Y$ when using test data $X_{1}=X_{2}=X_{3}=0$ using KNN.
(a) Compute the distance between each observation and the test data.
(b) What is our test classification with KNN if we choose $K=1$ (that is, I want to use the nearest neighbor's label to label the test observation)? Why?

[^0](c) What is our test classification with KNN if we choose $K=3$ ? Why?
(d) What is the training error when $K=1$ ? Suggest a method for choosing $K$ given a training data set.
3. Suppose we perform all subsets, forward selection, and backward selection on a single data set. For each approach, we obtain ${ }^{2} \mathrm{p}+1$ models, containing $0,1,2, \ldots, p$ covariates. Answer the following
(a) Which of the three outputted models with k covariates has the smallest training error?
(b) Which of the three outputted models with k covariates has the smallest test error?

[^1](c) True or False:
i. The covariates in the k -variable model identified by forward selection are a subset of the covariates in the ( $\mathrm{k}+1$ ) variable model identified by forward selection.
ii. The covariates in the k -variable model identified by backward selection are a subset of the covariates in the $(\mathrm{k}+1)$ variable model identified by backward selection.
iii. The covariates in the k -variable model identified by backward selection are a subset of the covariates in the $(\mathrm{k}+1)$ variable model identified by forward selection.
iv. The covariates in the k -variable model identified by forward selection are a subset of the covariates in the $(\mathrm{k}+1)$-variable model identified by backward selection.
v. The covariates in the k-variable model identified by all subsets are a subset of the covariates in the $(\mathrm{k}+1)$ variable model identified by all subsets
4. Geometrically ${ }^{3}$, sketch the optimization problems behind ridge and lasso when $p=2$ that correspond to including some regularization. What is the tuning parameter (either label your drawing or state it qualitatively)? Draw a separate version of each optimization problem that correspond to no regularization. (Draw two figures next to each other in the space provided)
(Ridge)
(Lasso)
5. As I increase $\lambda$ in ridge (Reminder: the Lagrangian term is $\lambda\|\beta\|_{2}^{2}$ ), what happens to the overall size of the coefficient estimates? Are any coefficients set to zero?
6. As I increase $t$ in lasso (Reminder: the constraint set is $\|\beta\|_{1} \leq t$ ), what happens to the overall size of the coefficient estimates? Are any coefficients set to zero?

[^2]
[^0]:    ${ }^{1}$ As I have mentioned several times, anytime a statement about distance arises, one should ask: what notion of distance are you referring to? Answer: squared Euclidean distance

[^1]:    ${ }^{2}$ For all subsets, consider the chose size $k$ model to be the size $k$ model with the smallest mean squared error (i.e.: training error)

[^2]:    ${ }^{3}$ that is, by drawing something

