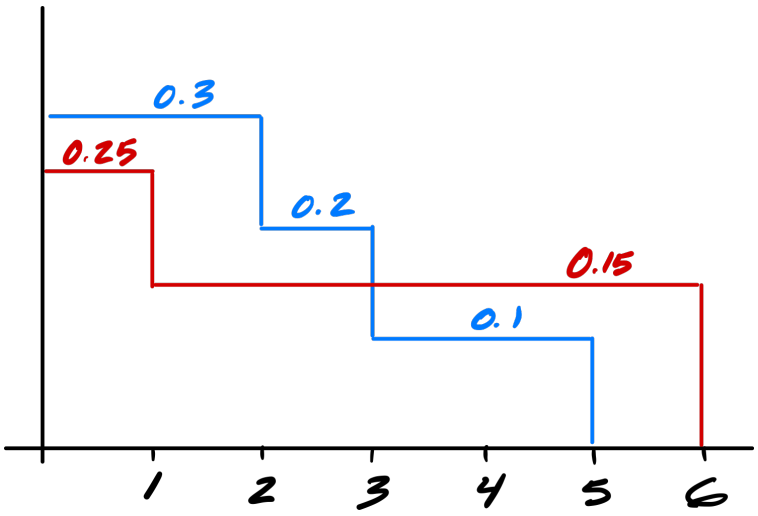
DSC 140A - Midterm 02 Review

Problem 1.

Plotted below are two conditional densities, $p_1(x \mid Y = 1)$ and $p_0(x \mid Y = 0)$, describing the distribution of a continuous random feature for two classes, $Y = 0$ (red) and $Y = 1$ (blue).



Suppose $\mathbb{P}(Y = 1) = .4$ and $\mathbb{P}(Y = 0) = 0.6$.

a) What is the prediction of the Bayes classifier at $x = 2.5$?

b) What is the Bayes error for this distribution?

Problem 2.

Suppose a data set of n data points $\{(\vec{x}^{(i)}, y_i)\}$ are drawn from a probability distribution, and that the Bayes error on this distribution is 0.21. The data set is split into a training set of n_1 points and a test set of n_2 points.

Suppose a support vector machine (SVM) is trained on the training set and tested on the test set. True or False: it is possible for the SVM to obtain an 85% accuracy on the test set.

Problem 3.

Suppose a histogram is constructed for a set of n data points $\{x^{(i)}\}$. The histogram contains a bin $[2, 5)$ and the height of the histogram within this bin is 0.25.

Suppose a point is drawn uniformly at random from the data set. What is the probability that the point is in the interval $[2, 5)$?

Problem 4.

True or False. The number of negative entries in a covariance matrix must be even.

Problem 5.

A data set contains 10 measurements of 50 different tumors. The covariance matrix containing the covariance between the features is computed. What is its shape?

Problem 6.

The procedure below is used to make a classification between two classes. What is its name?

The data from the two classes are separated and full covariance matrices for each class are computed. These separate, full covariance matrices are combined into a single covariance matrix. The class means and the shared covariance matrix are used to estimate $p(x|Y = y)$ in a Bayes classifier.

- ◯ LDA
- ◯ QDA
- ◯ Naïve Bayes
- ◯ Ridge Regression

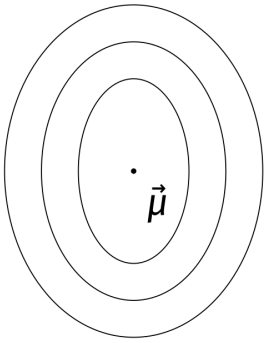
Problem 7.

The procedure below is used to make a classification. What is the procedure's name?

The data from two classes are separated and full covariance matrices for each class are computed independently (i.e., the covariance matrices are different). The class means and separate covariance matrices are used to estimate $p(x|Y = y)$ in a Bayes classifier.

- ◯ LDA
- ◯ QDA
- ◯ Naïve Bayes
- ◯ Ridge Regression

Problem 8.



The picture above shows the contour lines of a 2-dimensional Gaussian. One of the below options is the Gaussian's covariance matrix. Which is it?

- ◯ $\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$
- ◯ $\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$
- ◯ $\begin{pmatrix} 0 & 3 \\ 2 & 0 \end{pmatrix}$
- ◯ $\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$

○ $\begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}$

Problem 9.

The table below collects data on whether or not San Diego Search and Rescue needed to perform a rescue on 9 days last year. For each day, it is listed whether it was during the summer, if it was on the weekend, and whether it was raining.

Day	Is Summer?	Is Weekend?	Is Raining?	Rescue
0	False	False	True	True
1	False	True	True	False
2	False	True	False	False
3	True	True	False	True
4	True	True	False	True
5	True	False	True	False
6	False	False	True	False
7	True	False	False	True
8	True	True	False	True

Use Naive Bayes to predict whether there is a rescue on a day that 1) is not summer; 2) is not a weekend; and 3) is not raining.

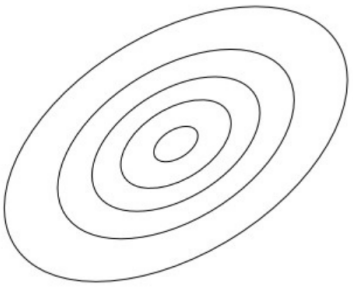
We have:

Problem 10.

Suppose that 32% of cars in a parking lot are red. Furthermore, suppose that 12% of cars in the parking lot are both red and have four doors. What is the probability that, given a randomly-selected car is red, that is has four doors?

Problem 11.

The figure below shows the contour lines of a 2-dimensional Gaussian. Which one of the following is true? Mark all that apply.



☐ It has non-zero entries off of the off-diagonal.

☐ All entries are positive.

☐ All off-diagonal entries are zero, and all diagonal entries are the same.

☐ All off-diagonal entries are zero, and all diagonal entries are different.

Problem 12.

Let X be the number of books sold at the UCSD bookstore in a given week, and let Y be the number of books sold at the UCLA bookstore in the same week. Are X and Y independent or dependent as random variables?

Problem 13.

Suppose you have two, 6-sided dice, each labeled with the numbers 1 through 6. You roll both dice; let X be the number on the first die, and let Y be the number on the second. Are X and Y independent or dependent?

Problem 14.

Suppose you have two, 6-sided dice, each labeled with the numbers 1 through 6. You roll both dice; let X be the number on the first die, and let Y be the number on the second.

Are X and Y conditionally independent given that their sum is odd?