Homework #6: Exam Problem Rework

Brandon Begaj, Casey Bruno, Kuang Jiang

**Problem No. 1**: Consider two probability distributions representing a 2-class problem:

where and .

1. (20 pts) Sketch the probability of error as a function of α. Very carefully label your graph and be as precise as possible. Hint: sketch these distributions.
2. (20 pts) For what value of α is the probability of error a maximum? Draw a sketch of the corresponding distributions to justify your choice.

**Theoretical/Analytical:**

Our analytical solution rested on our understanding of the decision rule for these classes, which have unequal prior probabilities. By using these probabilities, we were able to get the following decision rule.

Figure 1. Decision rule for the two classes

Using these equations, we identified three different cases resulting from three different ranges of α. In the middle region [1, 6], the probability of error remains a constant value. This is because the decision rule dictates that, for this range of α, we will always choose class 1 when the sample is less than 0, resulting in a 12.5% based on the prior probabilities. The other cases have more unique probabilities of error. For these ranges, (0,1) and (6,∞), both classes have the probability of error. Solving for the point where the class in error changes, we can set the inequality above into an equality and solve for x.

Figure 2. Solution for ‘threshold’ value between each class.

Using this value as our ‘threshold’ between the two classes, we can create some equations for the probability of error for each of these regions.

Figure 3. Probability of error calculations for each region.

These equations are very similar, but it is important to note that where each class is in error is different in either equation. Plotting the probability of error for each of these regions provides the following plot. Also included (in an additional file) is a gif showing how the error and error regions change based on a changing α.

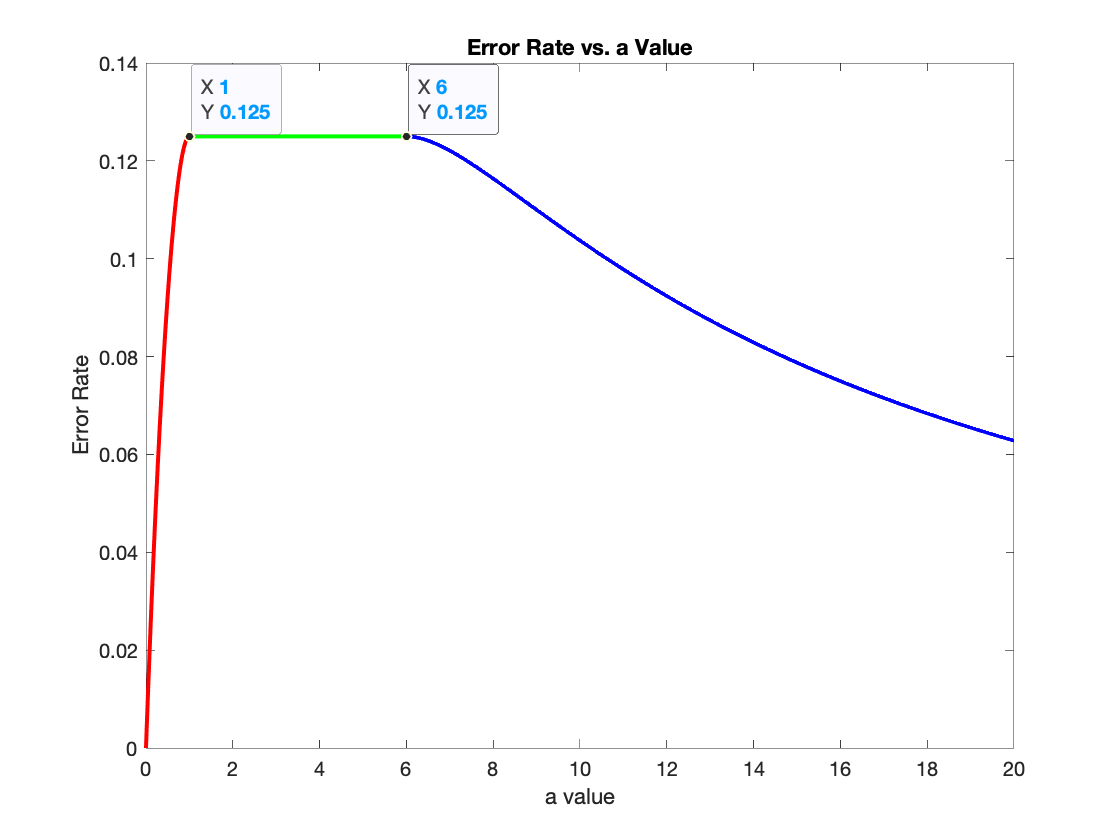


Figure 4. Plot of probability of error based on α.

**Experimental:**

In order to verify the results that we expect experimentally we must find a way to produce data which follows the given PDF’s for class 1 and class 2. A method of doing just this with only knowing the PDF’s is called Inverse Transform Sampling. The premise behind inverse transform sampling is that if you can compute the CDF of your distribution and that CDF has either a close form or nice analytical approximation, then you can generate points whose distribution is modeled by the starting PDF. The only thing that must be kept in mind is that the inverse CDF must be given uniformly distributed points between the range of the PDF in order to produce the proper output.

We know the PDF’s for both classes however, one PDF is a function of alpha as well. Since we are using software tools to simulate the error rate for the classification problem at hand, we can give alpha a constant value and vary it as we see fit to notice the behavior of the error rate.

Evaluating the CDF of these functions is straightforward as we simply integrate the PDF with respect to x (Notice that the integral under the curve from -α to 0 is always 1). In doing so we find:

Figure 5. Cumulative distribution function for Class 1.

We can compute the functional inverse of this CDF by the conventional method for finding a functional inverse. In this case we are lucky that the inverse CDF has a closed form representation given by:

Figure 6. Inverse CDF for the Class 0.

We can perform the same operation on the second class except with a special modification. Given the two PDF’s, we see that the first class only appears on the left side of the x-axis while the second class appears in both. Since we are considering classification, we know that the piece of the second class for values of x greater than or equal to 0 will never be misclassified. For this reason, we can ignore the right-hand side of the PDF for the second class. Carrying this out and remembering to normalize the PDF before continuing, we end up with the inverse CDF function:

Figure 7. Inverse CDF for Class 2.

With the inverse CDF’s computed, the only step left to generate our data is to feed uniformly distributed points in between the range of the original PDF’s. For the first class this is varied but we know that the points must be uniformly distributed between -α to 0 while the second class remains constant with values from -1 to 0. The output of the CDF’s should now follow our PDF’s and we can check that this is true by observing the shape of histograms created from the generated data. We can show this for the second class as the PDF remains constant with respect to α. When generating the data however we must also account for the priors and generate 6 times as many points from the first class as the second, since we are only generating data for half of the second class (as the right hand side of the class will always be correct).

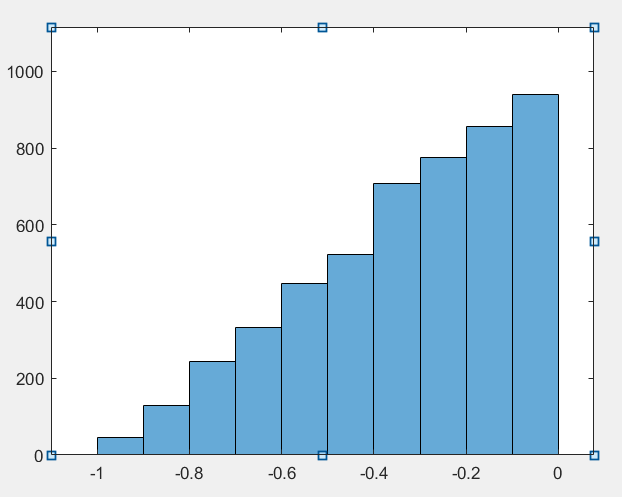


Figure 8: PDF of generated values for Class 1.

As we can see this produces the same triangle pattern that we can see if we were to plot the PDF of the second class. The last steps of the experimental results are to begin classification of the generated data. This is straightforward, as all we must do is keep track of which class the generated data came from. The decision rule takes into account the prior probabilities and we already know the PDF’s so it’s as simple as evaluating the PDF’s at the generated points and multiplying by the prior probabilities. We then assign the class who had the largest result.

Finally, we can compare the number of errors to the total number of generated points to calculate our experimental error rate. We must bring along the once assumption however that data points are uniformly distributed between -1 and 1 for the second class, so we must add however many points were generated from the second class in order to account for those points never being misclassified. This concludes the Inverse Transform Sampling method as well as classification and error rate calculations for the experimental error rate.

The following diagram shows a continuous generation for error rate with different values of . It matches the theoretical graph that we had with a similar trend of the diagram. Since results are generated randomly, the error rate is modulating around the expected theoretical value.

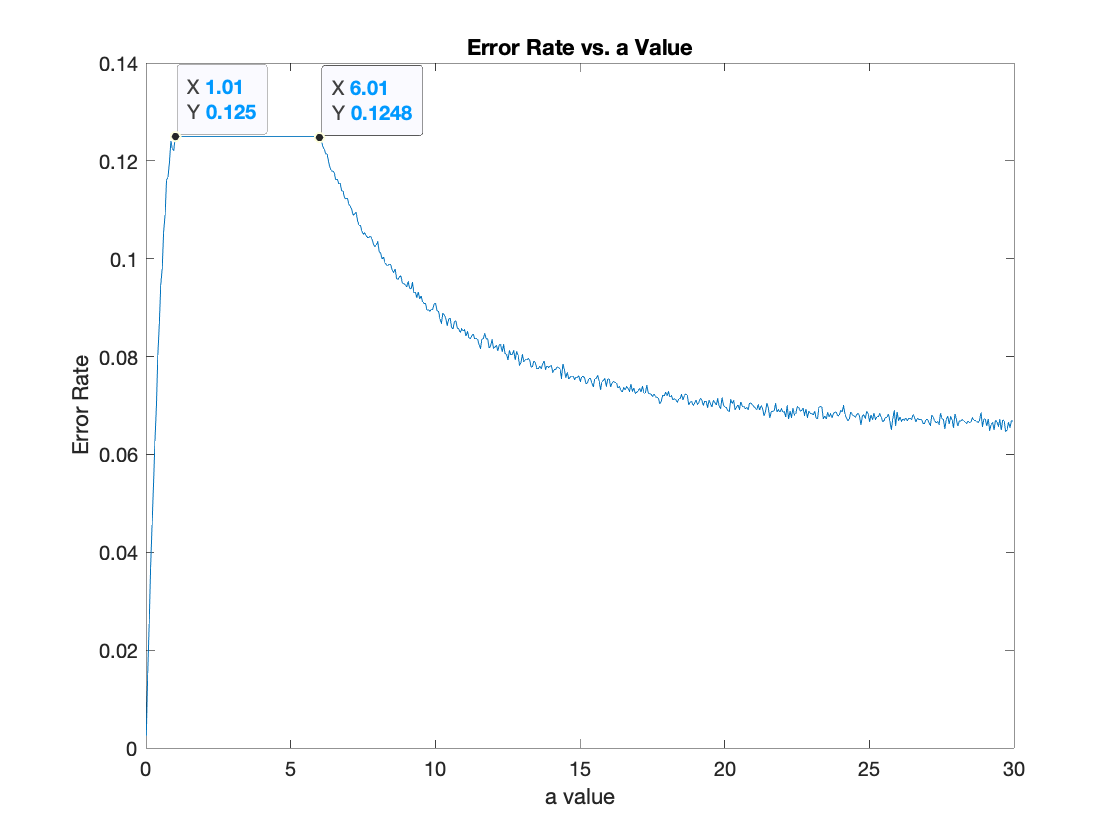


Figure 9: Plot of experimental probability of error based on α

We also generated a table that combine a random selected amount of to show the difference when generating the error rate via experimental and theoretical approach. The following table show the generated graph with corresponding label for each section. The experimental error rate that was generated fluctuate near the value of the theoretical error rate.

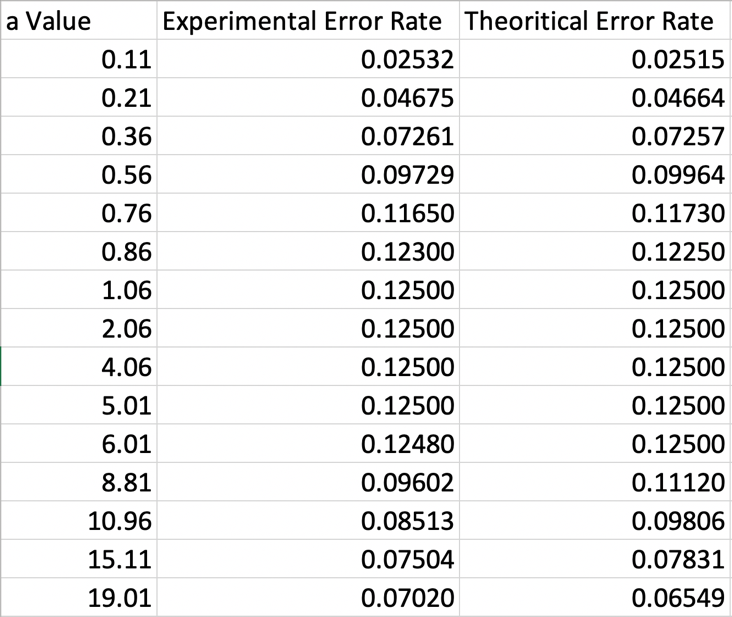


Figure 10: Table of Error Rate with Corresponding value