

# ECE 8110 Machine Learning Homework Assignment 1

January 31, 2014

Salvatore Giorgi  
*salvatore.giorgi@temple.edu*

**Exercise 1.** Generate 10,000 data points for each of two Gaussian RVs:  $\mathcal{N}(1, 1)$  and  $\mathcal{N}(-1, 1)$ . Compute the error rate using an optimal maximum likelihood decoder with a threshold set at 0. Repeat for a threshold set at  $\pm 0.5$ . Do these error rates match the predicted values?

*Solution:* First, we begin by generating the 10,000 data points from each of  $\mathcal{N}(1, 1)$  and  $\mathcal{N}(-1, 1)$ . To verify that we have done this correctly, we plot the pdfs of both distributions as well as a scaled histogram of the data points, as seen in Fig. 1 and 1.

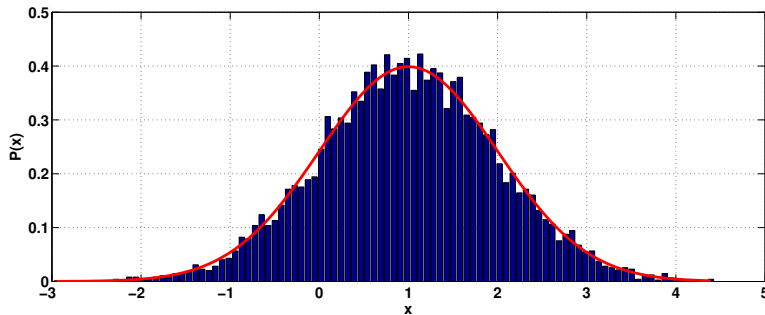


Figure 1: Plot of  $\mathcal{N}(1, 1)$  with Scaled Histogram of 10,000 Data Points

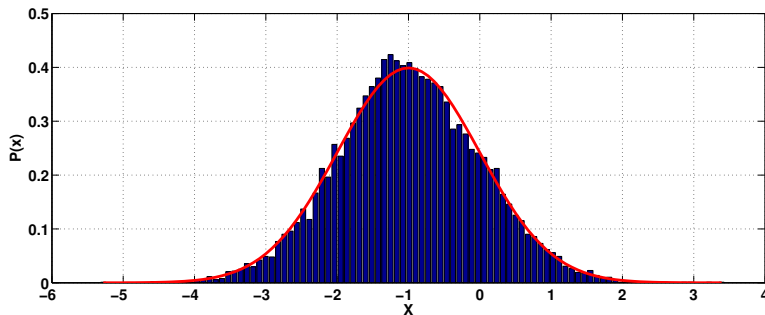


Figure 2: Plot of  $\mathcal{N}(-1, 1)$  with Scaled Histogram of 10,000 Data Points

Next, we compute the error rate using an optimal maximum likelihood decoder. We note that both distributions are the same distance from 0 and have equal variance. The decision surface is defined by [1]

$$\|x - 1\|^2 - \|x + 1\|^2 = 2 \ln \frac{P(\omega_2)}{P(\omega_1)}. \quad (1)$$

Since both distributions are the same distance from 0, the equation above shows us that the optimal decision surface is at 0. We then set the threshold equal to 0 and calculate the probability of error. To find this analytically, we integrate the pdf of  $\mathcal{N}(1, 1)$  from  $(-\infty, 0]$  and  $\mathcal{N}(-1, 1)$  from  $[0, \infty)$ . We then take these two errors, sum them, and divide by 2. Since we are finding a probability, the area under both distributions combined must equal 1, hence, we must divide the combined error probabilities by 2. The error rates can be found in Table 1.

To compute the error rate from the data, we count the number of data points from  $\mathcal{N}(-1, 1)$  greater than 0 or equal to and the number of data points from  $\mathcal{N}(1, 1)$  less than or equal to zero. We then add these points and divide by the total number of points. The error rates can be found in Table 1.

Finally, we set the threshold to  $\pm 0.5$  and calculate the error rates exactly as above. The results are shown in Table 1, along with the percent error between the two calculations.

Table 1: Predicted and Calculated Error Rates

Threshold	Predicted	Calculated	Percent Error
0	0.1587	0.1573	0.88
0.5	0.1877	0.1861	0.85
-0.5	0.1877	0.1839	2.02

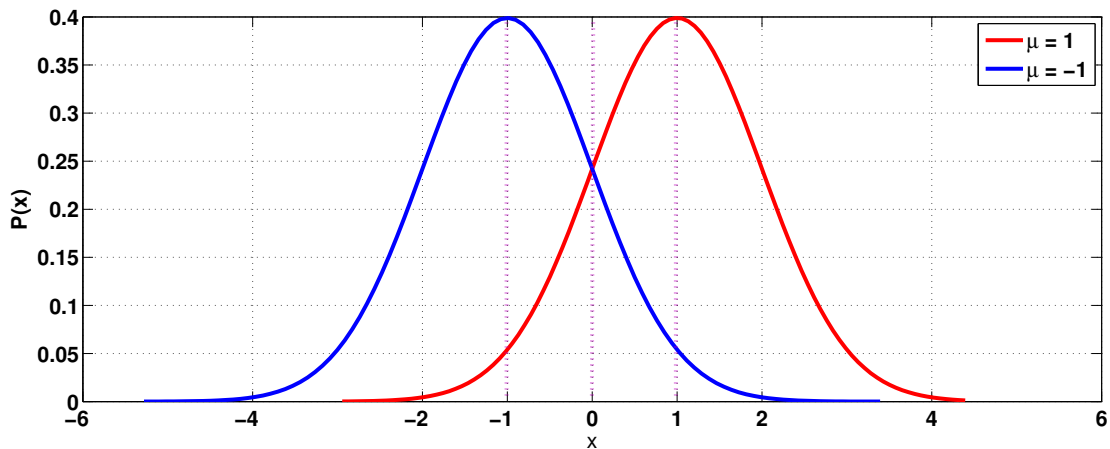


Figure 3: Plot of  $\mathcal{N}(1, 1)$  and  $\mathcal{N}(-1, 1)$  with Thresholds at  $\{-1, 0, 1\}$

To explain the error rates, we again note that both distributions have the same variance and are equal distance from 0. Thus, moving the threshold to some points  $\pm q$  will result in the same theoretical error. With reference to Fig. 3 we can see that as we move the threshold away from zero we are moving closer to the tail of one distribution, but closer to the mean of the other distribution. Therefore, as we approach the tail, the area under the tail (which is our error probability) shrinks. But at the same time, the area under the other distribution increases. We can see that this area increases faster than the area under the tail decreases. Hence, as we move away from 0, our probability of error increases.

**Exercise 2.** Generate 1000 data points for each of two Gaussian RVs:  $\mathcal{N}(1, \sigma_1^2)$  and  $\mathcal{N}(-1, \sigma_2^2)$ . Set  $\sigma_1^2 = 1$  and vary  $\sigma_2^2$  over the range  $[0.1, 2]$ . Plot the error rate as a function of  $\sigma_2^2$ .

*Solution:* We begin by plotting the pdfs of the two RVs  $\mathcal{N}(1, \sigma_1^2)$  and  $\mathcal{N}(-1, \sigma_2^2)$ , varying  $\sigma_2^2$  over the range  $[0.1, 2]$ , in Fig. 4. From this image we can see that as  $\sigma_2^2$  approaches 2, the area of the corresponding distribution under the distribution of  $\mathcal{N}(1, \sigma_1^2)$  grows. If we set the threshold at 0, we can see that we are more likely to classify a data point from  $\mathcal{N}(1, \sigma_2^2)$  as originating in  $\mathcal{N}(1, \sigma_1^2)$ . Thus, we expect our error rate to increase as  $\sigma_2^2$  approaches 2.

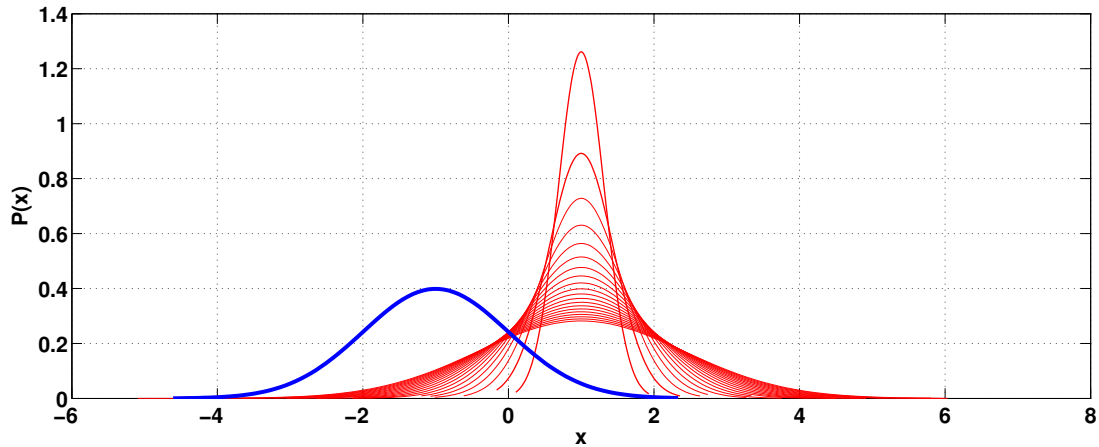


Figure 4: Plot of  $\mathcal{N}(1, 1)$  and  $\mathcal{N}(-1, \sigma_2^2)$  with  $\sigma_2^2$  over the range  $[0.1, 2]$

To find the probability of error, we used the same methods as described above. For the theoretical error rate, we integrate the pdfs on either side of the threshold corresponding to the incorrect decision. To calculate this from the data, we count the number of points from each distribution either greater than or less than the threshold and divide this number by the total number of points. Both error rates, as a function of  $\sigma_2^2$  are shown in Fig. 5. From this figure we can see that our prediction above is correct for a threshold of 0: the probability of error increases as the variance of the distribution grows.

From Fig. 5 we can see that our prediction above is correct for a threshold of 0: the probability of error increases as the variance of the distribution grows. But this is dependent on our choice of threshold. To find a lower probability of error we must also vary the threshold. Figs. 6 and 7 show the result of varying both  $\sigma_2^2$  and the threshold. We plot both the theoretical error and calculated error from the data for  $\sigma_2^2 \in [0.1, 2]$  and threshold  $\in [-0.5, 0.5]$ . From these two figures, we can see that for each given threshold it is not true that a smaller variance will result in a smaller probability of error. For example, we see that for larger negative thresholds, the probability of error graph is a parabola with a minimum around  $\sigma_2^2 = 0.5$ . Also, we see that for a threshold of approximately 0.6 and  $\sigma_2^2 = 0.1$  we have the smallest probability of error.

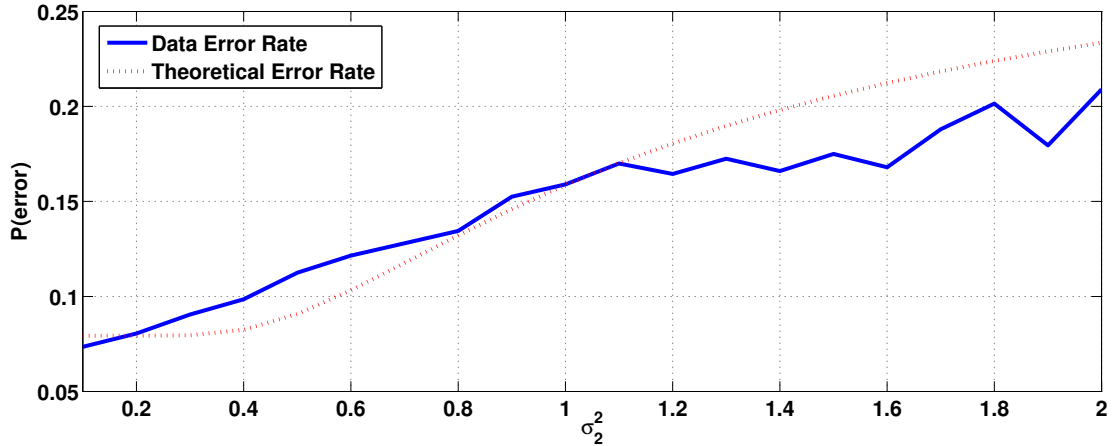


Figure 5: Probability of Error as a Function of  $\sigma_2^2$

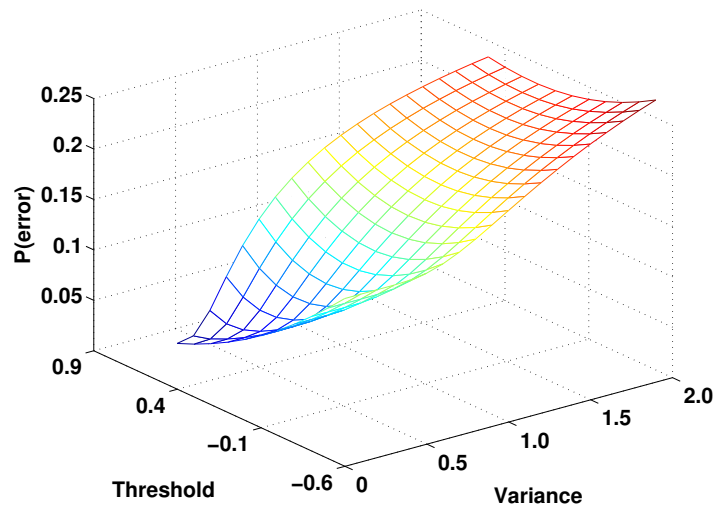


Figure 6: Theoretical Probability of Error as a Function of  $\sigma_2^2$  and the Threshold

**Exercise 3.** Generate a 2D Gaussian RV with a mean of  $[-1 \ 1]$  and an identity covariance matrix. Generate a second Gaussian RV with a mean of  $[1 \ 1]$  and an identity covariance matrix. Compute the error rate for the optimal maximum likelihood decoder. Does this match the theoretical prediction?

*Solution:* This exercise is the 2D analog to Exercise 1. We see that each distribution is the same distance from 0 and that both distributions have the same variance, which in both cases is the identity matrix. Therefore, each distribution will be symmetric about its mean. We show these two distributions and their support in Figs. 8 and 9, respectively. Putting all of this together, we get that a plane perpendicular to and intersecting the  $x$  axis will give us the decision boundary minimizing the probability of error. We show the projection of this decision plane in Fig. 9 as

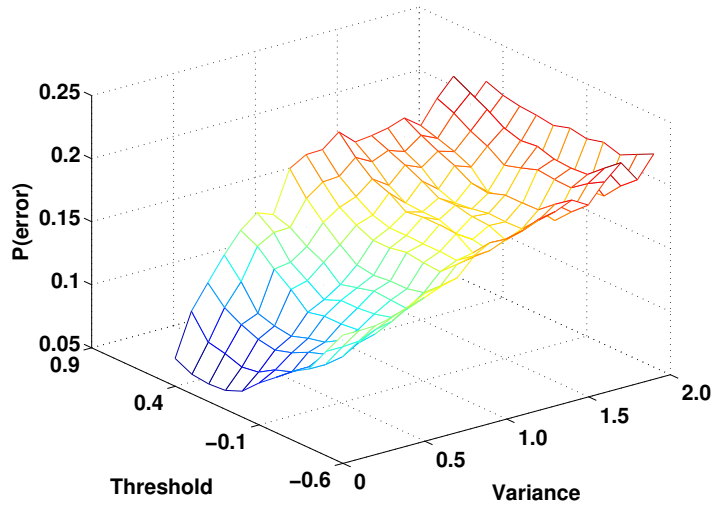


Figure 7: Calculated Probability of Error as a Function of  $\sigma_2^2$  and the Threshold

a black line at 0.

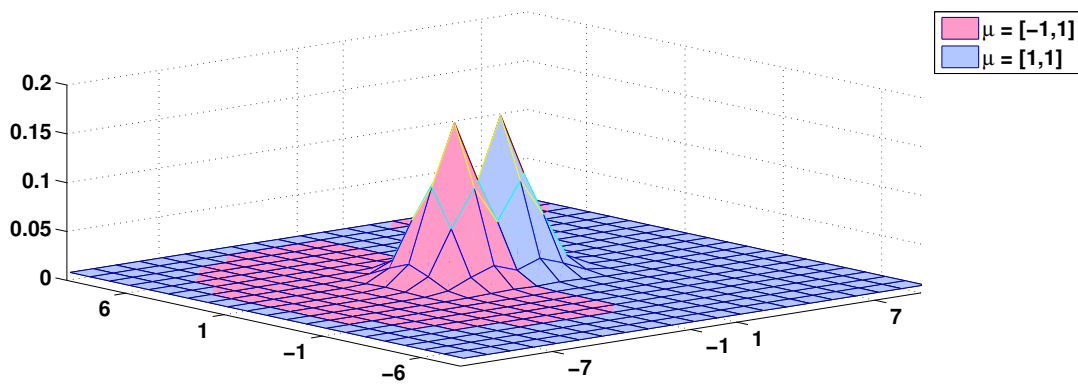


Figure 8: PDF of both Gaussian RVs

Again, we calculate the theoretical error by integrating each pdf on the opposite side of the decision boundary, adding the two values, and dividing by 2. To calculate the error from the data, we count the number of points within the incorrect decision area and divide this number by the total points (10,000 from each random variable). The error values can be found in Table 2, along with the percent error between the two calculations. From our observation that this problem is just a higher dimensional version of the first problem, we expect to get the same optimal error rate, which is indeed the case.

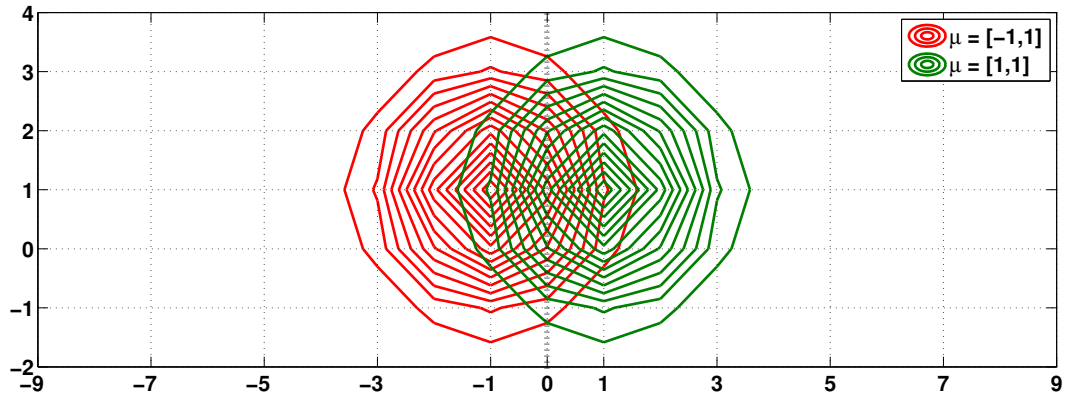


Figure 9: Support of both Gaussian RVs

Table 2: Predicted and Calculated Error Rates

Threshold	Predicted	Calculated	Percent Error
0	0.1587	0.1567	1.26

**Exercise 4.** Generate a plot of the support region for the first Gaussian RV in Exercise 3 as you vary the covariance. Demonstrate how the shape of the support region changes depending on the nature of the covariance matrix.

*Solution:* For this exercise we consider a 2D Gaussian RV with  $\mu = [1 \ 1]$  and varying covariance matrix. For each covariance matrix we plot both the pdf and the support. We consider the following covariance matrices:

$$\begin{bmatrix} 6 & 1 \\ 1 & 6n \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} 6m & 1 \\ 1 & 6 \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} 6 & l \\ l & 6 \end{bmatrix} \quad (4)$$

for  $n, m, l \in \{1, 3, 5\}$ . Note that  $\forall n, m, l$  the matrices above are all symmetric and positive semi-definite, and therefore are covariance matrices. Figs. 10 through 18 show the pdf and support of the RV with varying the covariance matrix. Since we are varying on only term each iteration or two terms in the case of (4) (because of symmetry), we expect the support to change more over one axis. This is essentially what we see in the figures. The first three (where we vary the bottom right term) grow along the vertical axis, the second three (where we vary the top left term) grow along the horizontal axis, and the last three (where we vary the diagonal terms) grow along the diagonal axis.

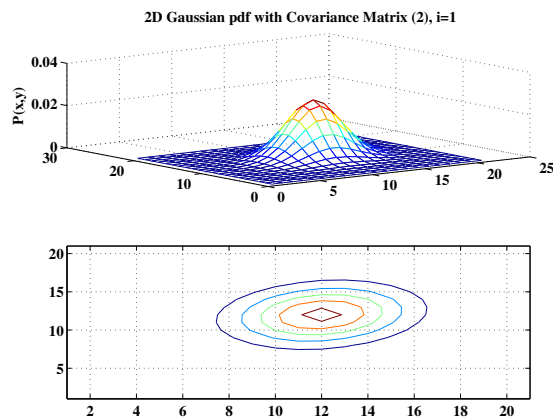


Figure 10: PDF and Support Gaussian RVs with Covariance (2) and  $i = 1$

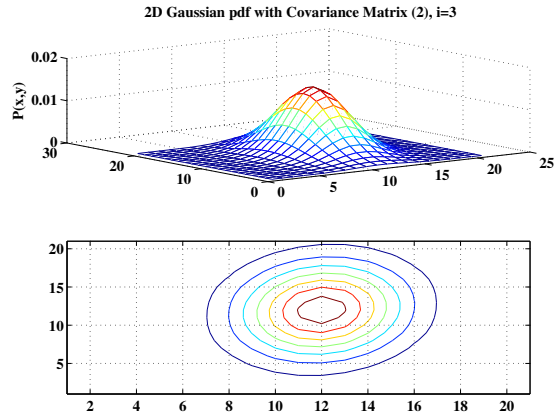


Figure 11: PDF and Support Gaussian RVs with Covariance (2) and  $i = 3$

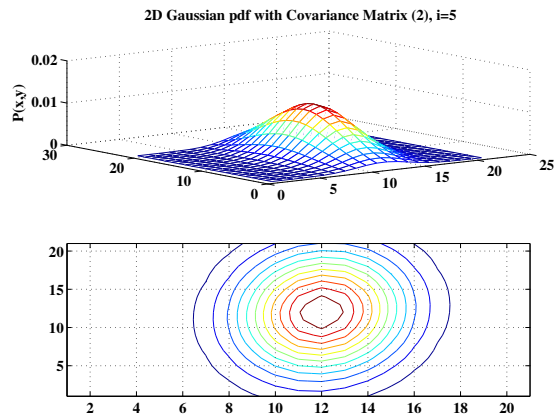


Figure 12: PDF and Support Gaussian RVs with Covariance (2) and  $i = 5$

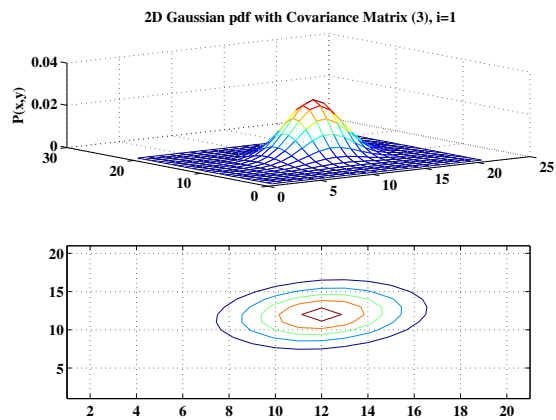


Figure 13: PDF and Support Gaussian RVs with Covariance (3) and  $i = 1$



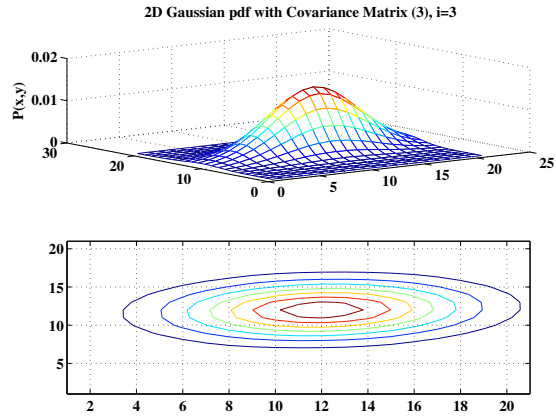


Figure 14: PDF and Support Gaussian RVs with Covariance (3) and  $i = 3$

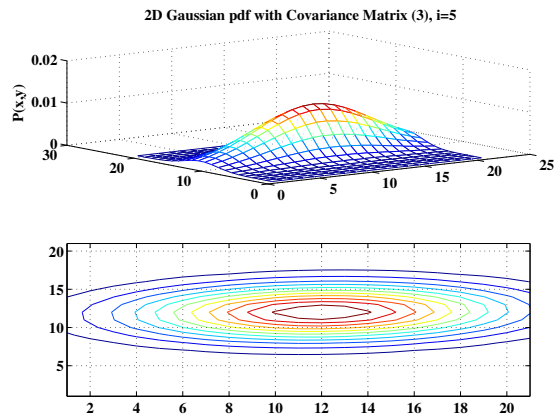


Figure 15: PDF and Support Gaussian RVs with Covariance (3) and  $i = 5$

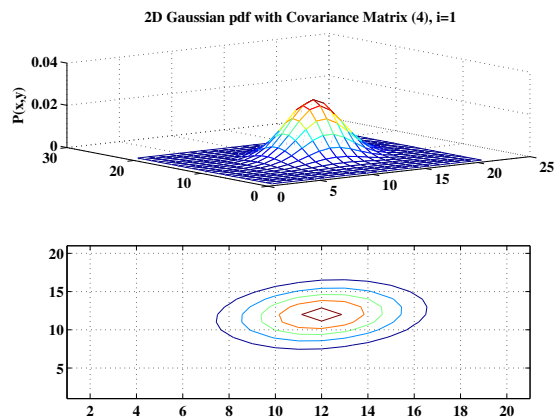


Figure 16: PDF and Support Gaussian RVs with Covariance (4) and  $i = 1$

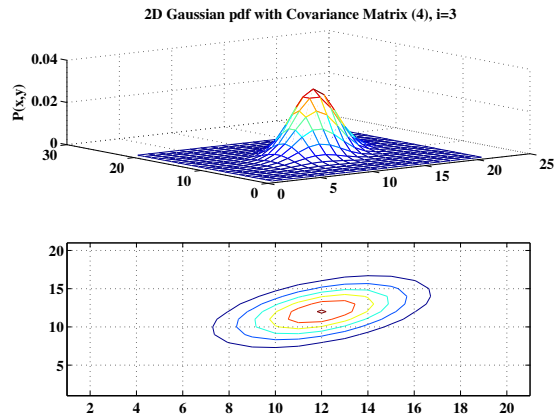


Figure 17: PDF and Support Gaussian RVs with Covariance (4) and  $i = 3$

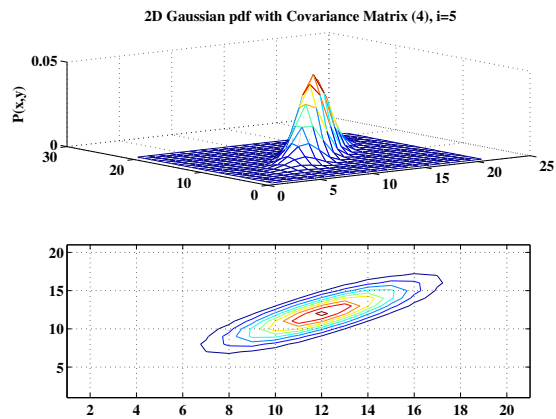


Figure 18: PDF and Support Gaussian RVs with Covariance (4) and  $i = 5$

## References

- [1] R. Duda, *Pattern Classification*, Wiley, Hoboken, NJ, 2000.