COMP6245 Foundations of Machine Learning – Lab 1

Supritha Konaje
Email: ssk1n21@soton.ac.uk
Student ID: 32864477

1. Manipulation of Vectors and Matrices

A refresher for vectors and Matrices was given in this section. To obtain the *dot product of vector x* and y, we can use an existing function of numpy that is *np.dot(x, y).* To *normalize a vector x*, we can use *np.linalg.norm(x)*. We can easily obtain the *angle between the vectors x and y* using the function, *np.arccos(np.dot(x,y) / (np.linalg.norm(x) * np.linalg.norm(y))).*

To obtain *random values*, we can use *np.random.rand(a, b)*, where a is the row size and b is column size. For a matrix A, we can find the *dimension* using *A.shape*. Same goes for vector, we can find the dimension using *x.shape*. To obtain the sum of diagonals that is a *trace of a matrix*, we can use *np.trace(A)*. To find the *determinant*, we can use *np.linalg.det(A)*. By using *np.linalg.eig(B)*, we can obtain the corresponding *eigen value and vector*.

2. Random Numbers and Uni-variant Densities

2.1 Uniform Distribution

We are considering here *1000 random numbers* that will lie between the range of 0 and 1 as per probability rule. We will be using python's numpy package function *np.random.rand(<size_of_random_number>, <dimension_of_array>).*

In Figure 1, we can see that for bin size 4, the interval range is wide so the count in each random variable in also more. Whereas, if we increase the bin size, the count of number of random numbers in each variable decreases. When we plot a histogram for uniform distribution, we expect the histogram to be flat. But as we can see in Figure 1, the histogram is not flat because the numbers are discrete. Hence, for every interval there might be higher count or lower count for the values. Every time we run the code, a new set of random numbers are generated. As and when we keep increasing the size of random numbers, we get almost a flat histogram.

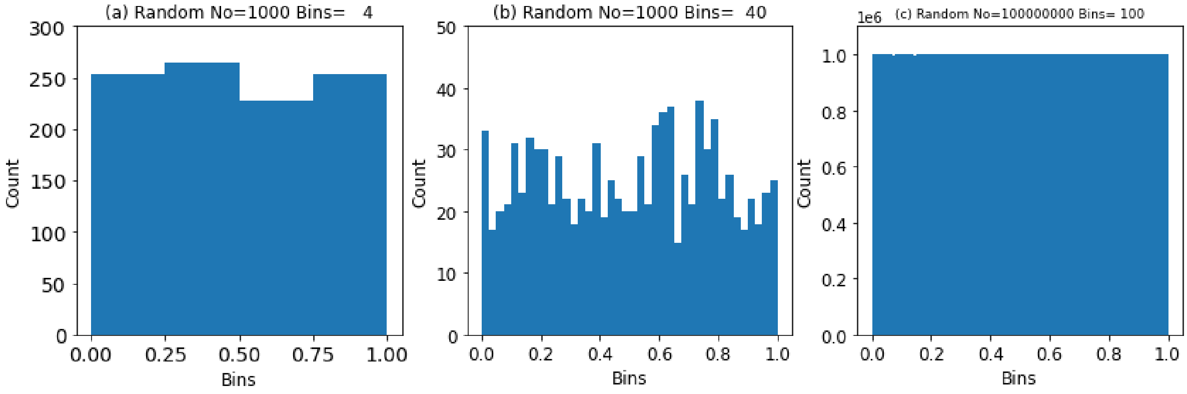

Figure 1: Histogram for Uniform (a) Random No: 1000 & Bins: 4, (b) Random No: 1000 & Bins: 40, (a) Random No: 100000000 & Bins: 100

2.21 Univariant Gaussian Distribution

For Gaussian Distribution, histogram was plotted by subtracting the sum of x random numbers from the sum of y random numbers. As we can see in Figure 2, when the value of x and y is increased, the standard distribution increases around the mean. The variance provides an idea regarding how variable the distribution is around the mean.

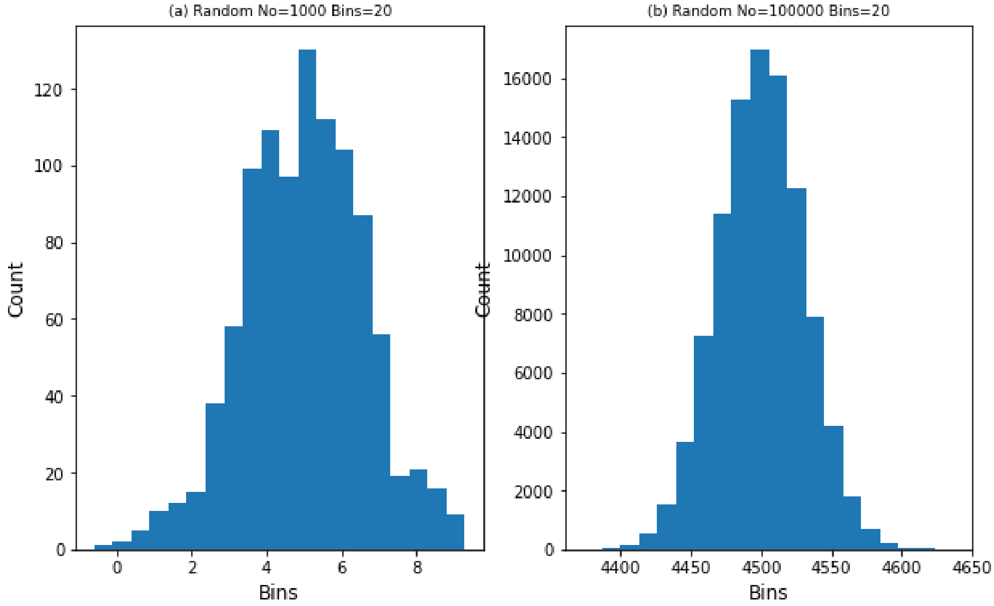

Figure 2: Histogram for Gaussian Distribution for (a) Random No: 1000 Bins: 20 x:20 y:10 (b) Random No: 100000 Bins: 20 x:10000 y:1000

3. Uncertainty in Estimation

The concept of estimation exists if we have enough sample data to compare or judge the calculated value. With estimation, there will always be uncertainty in the result that we get. More the sample data, more accurate our estimations will be. In Figure 3, we have taken one example for Sample ranging from 100 to 300 with sample size as 100. We can see that the graph is distorted as it does not have enough samples to form a stable curve. In the second graph, the sample ranges from 100 to 1500 and sample size is 1000. In that graph, the curve seems to be stable. Hence, higher the value of sample size higher the certainty in estimation. Low variance provides precise data which obeys the law of Gaussian Distribution.

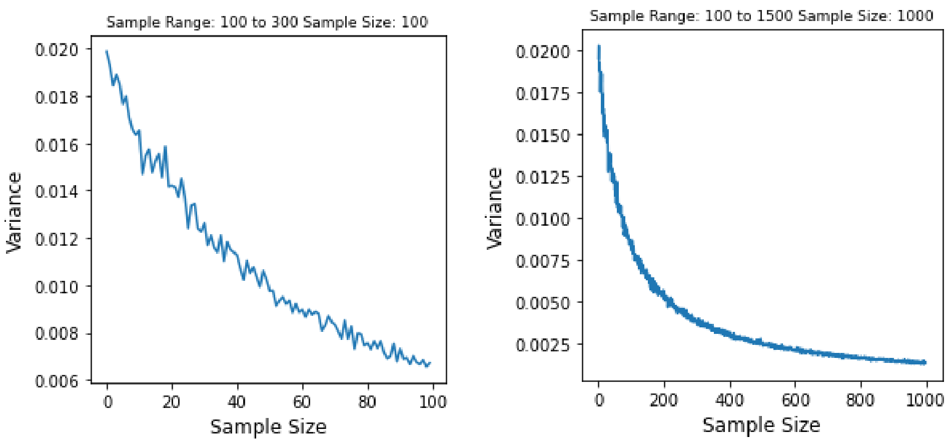

Figure 3: Graph showing the precision in data with different variance

4. Bi-variant Gaussian Distribution

Multi-Variant gaussian distribution can be obtained using the below given equation:

$$N(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

Where, we have a D-dimensional vector x, µ is a D-Dimensional mean vector, ∑ is a D*D covariance matrix and |∑| denotes the determinant of ∑.

We will be plotting a contour graph for bi-variant gaussian distribution. Following are the three covariant matrices and mean:

m1=$\begin{bmatrix}0\\2\end{bmatrix}$ C1=$\begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}$ m2= $\begin{bmatrix}2.4\\3.2\end{bmatrix}$ C2=$\begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}$ m3=$\begin{bmatrix}1.2\\0.2\end{bmatrix}$ C3=$\begin{bmatrix}2 & 0\\0 & 4\end{bmatrix}$

The result of the equation is plotted in 2D and 3D graph using the above given data. In Figure 4, for a general covariance matrix we get a shape like the purple contour. For a proportion to the identity matrix, the shape is concentric like the red colored contour. For a diagonal matrix, the contour is aligned to the coordinate axes like the green colored contour.

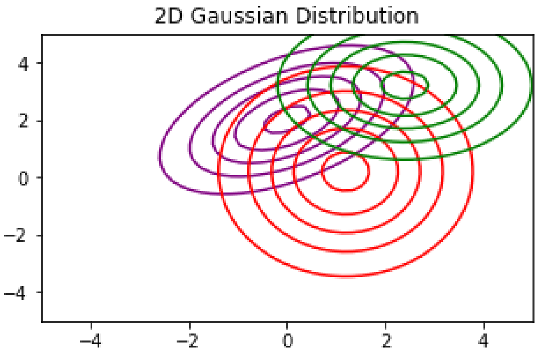

Figure 4: 2D Bivariant Gaussian Distribution

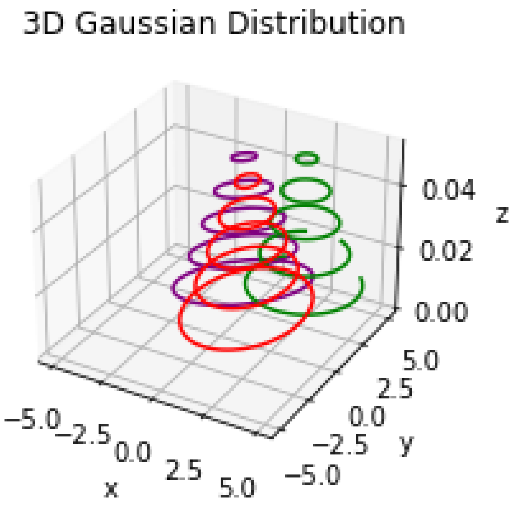

Figure 5: 3D Bivariant Gaussian Distribution

5. Sampling from a multi-variate Gaussian

For 1000 continuous (Gaussian Distribution) random numbers, we can see the scatter plot in yellow. This is representing the Isotropic Gaussian density.

We were given *Mean(µ)* =$\begin{bmatrix}0\\0\end{bmatrix}$ and $Covariance\ matrix(\Sigma)$ =$\begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}$ which has *Cholesky decomposition C* = [∑][∑]^T = $\begin{bmatrix}1.41421356 & 0\\0.70710678 & 1.22474487\end{bmatrix}$.

The red scatter plot represents the correlation between the scatter plot of random variables and the calculated product of random variables and Cholesky decomposition. ∑ can be created using Cholesky decomposition of C. Hence, we can get the correlated gaussian distribution as *Y ~ N (µ, ∑(C =* [∑][∑]^T*)).*

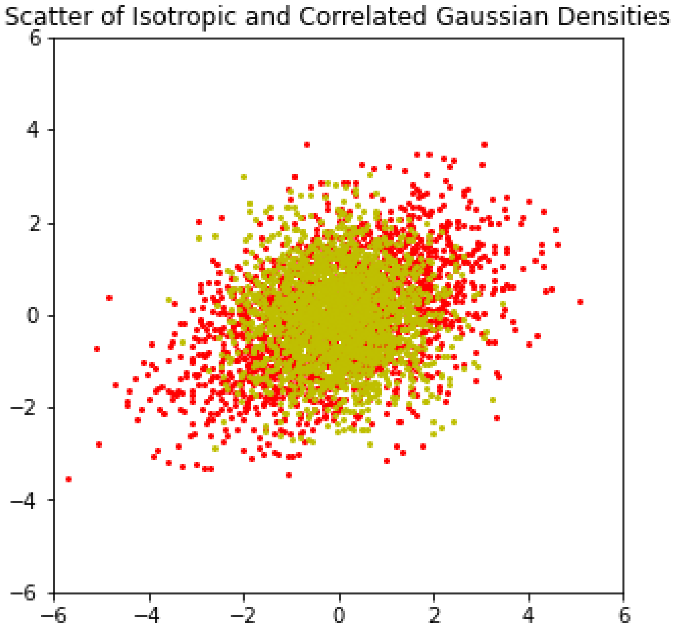

Figure 6: Scatter of Correlated Gaussian Densities

6. Distribution of Projections

We have created a graph that calculates the variance from the set $Y = \{xcos\theta + ysin\theta \mid (x, y) \in Y\}$, where Y is a set of points plotted in figure 6.

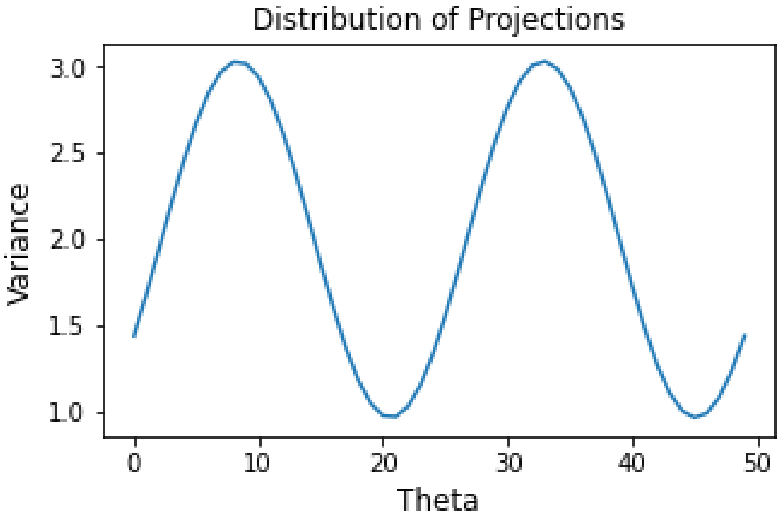

Figure 7: Graph of variance of set Y over Theta.

We can see that the variance ranges from 1 to 3. These are the eigen vector of the covariance matrix $\sum$. If we take smaller number of samples from the scatter plot, we will notice that the graph is not sinusoidal reason being that the data we have considered is discrete.