University of Southampton

Department of Electronics and Computer Science

Foundations of Machine Learning Lab 1 Report

Abhilash Pulickal Scaria

Student Id: 33124639

aps1n21@soton.ac.uk

18 October 2021

Objective

The main objective of this lab is to do sampling and projection using the properties of multivariate Gaussian densities.

Observation

1)In the first code snippet we defined a 3x3 matrix $B = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 6 & 5 \\ 1 & 5 & 9 \end{pmatrix}$. The defined matrix is a real and symmetric matrix. After finding the eigen vectors and eigen values of matrix B it is observed that the dot product of its eigen vectors are zero. This implies the vectors are orthogonal. In order to prove that this is indeed the expected result. Let us assume the two eigen vectors as u and v corresponding to eigen values α and β.

$$Bu = \alpha u \tag{1}$$
$$Bv = \beta v \tag{2}$$
$$\alpha\,(u.v) = (\alpha u)\,.v \tag{3}$$

Using eq (1) eq (3) can be written as

$$Bu.v = (Bu)^\top v = u^\top B^\top v$$

B is a symmetric matrix so

$$u^\top B^\top v = u^\top Bv$$

From eq (2)

$$u^\top Bv = u^\top \beta v = \beta\left(u^\top v\right) \tag{4}$$

Using eq(4) we get

$$\beta\left(u^\top v\right) = \beta\,(u.v)$$

Thus

$$\alpha\,(u.v) = \beta\,(u.v)$$
$$(\alpha - \beta)\,(u.v) = 0$$

Since $\alpha - \beta \neq zero$ u.v is zero. Therefore u and v are orthogonal.

2) In the second code snippet we are creating histograms using values sampled from a normal distribution. Initially we run it with 1000 values and it is observed that the histogram is not flat this is due to the fact that as the sample size is not too big some bins are less occupied than others. As we increase the data size to 100000 we see the histogram becoming flat because we have more data to plot. This can be observed from the figures 1 and 2. The reason for histogram looking different when each time it is run is because we are randomly sampling from unifrom distribution and the values we get varies during different runs.

Figure 1: Histogram of 1000 data points.

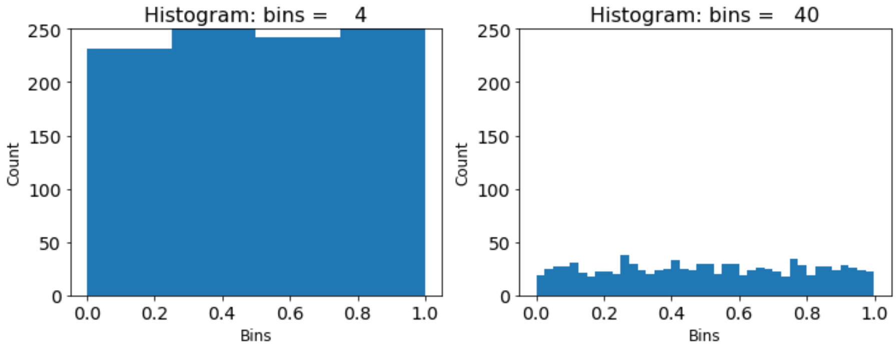
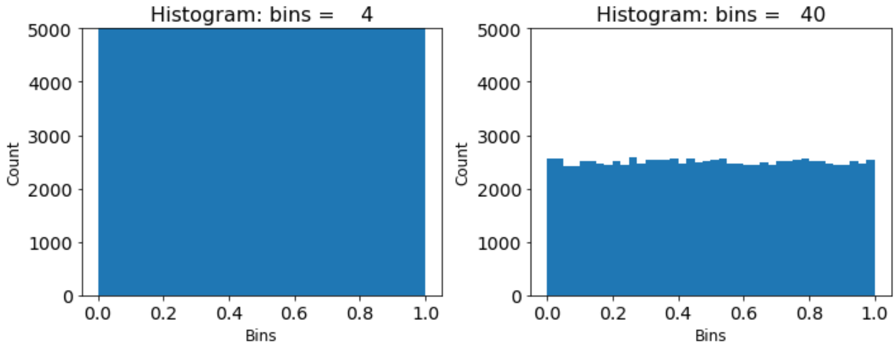
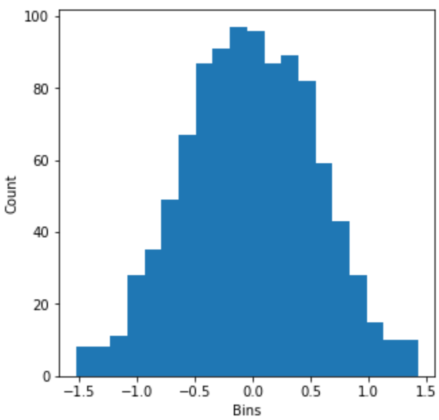


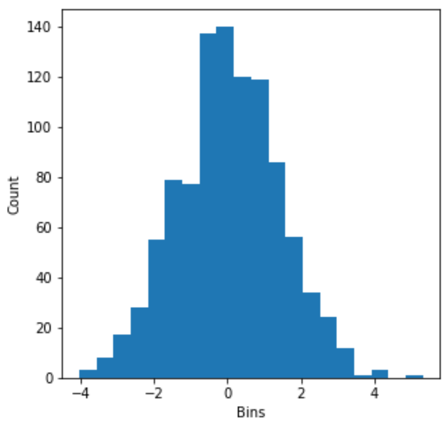
Figure 2: Histogram of 100000 data points.



3) While running the Third code snippet we observe that as the number of uniform random



(a) Histogram when 2 uniform random numbers added

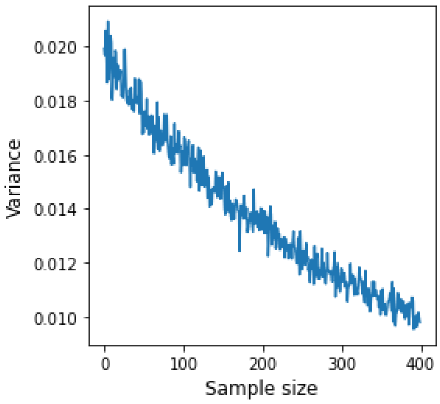


(b) Histogram when 12 uniform random numbers added

Figure 3: Histograms

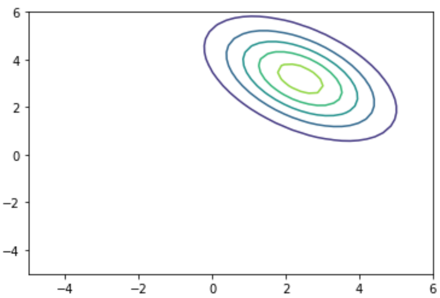
numbers added and subtracted increases from 2 to 12 the histogram changes from a triangular shape to a Gaussian distribution. This is shown in figure 3(a) and 3(b).

Figure 4: Variation of variance as sample size increases.

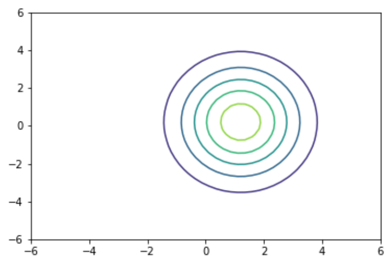


4) While calculating the variance of a univariate Gaussian density using sampling(code snippet 4) we find that as the sample size increases the variation in variance reduces. This can be observed from figure (4).

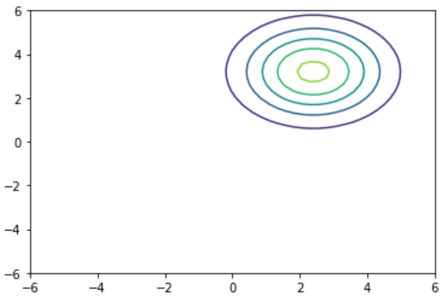
5) Using code snippet 5 we draw contours of 3 distribution(figure (5)). We can observe that the center is positioned at the mean of ditribution. From figure 5(a) it can be observed that the plot has a negative slope this is expected as the covariance is negative. The axes of ellipsoids are scaled and oriented based on the eigen values and vectors of the inverse covariance matrix.



(a) Contour plot for distribution with mean $\begin{pmatrix}2.4\\3.2\end{pmatrix}$ and covariance $\begin{pmatrix}2 & -1\\-1 & 2\end{pmatrix}$



(b) Contour plot for distribution with mean $\begin{pmatrix}1.2\\0.2\end{pmatrix}$ and covariance $\begin{pmatrix}2 & 0\\0 & 4\end{pmatrix}$



(c) Contour plot for distribution with mean $\begin{pmatrix}2.4\\3.2\end{pmatrix}$ and covariance $\begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}$

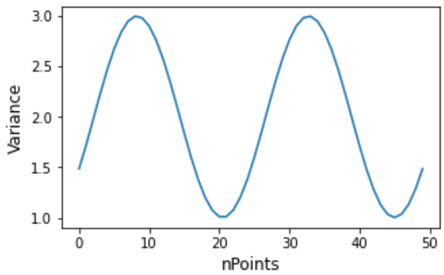
Figure 5: Contour plots

6) Using the properties learned from the above codes we sample from a multivariate distribution using code snippet 6. We use the Cholesky decomposition to obtain a lower triangular matrix. This matrix is then used to transform the bivariate gaussian data points sampled from standard

normal distribution. In this way we obtain the desired samples.

7) We project the obtained sample data from code above and calculate its variance along that direction. While plotting the variance (code snippet 7) as the theta is varied from 0 to 2π we observe the minima and maxima as 1 and 3

Figure 6: Variance as theta varied from 0 to 2π.



The eigen values and eigen vectors of the covariance matrix C are $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{pmatrix}$ respectively. It can be observed that the maximum and minimum values are same as the eigen values. The eigen vector of the maximum eigen value gives the direction in which the variance is maximum. As we plot the variance of projected data by varying the theta we obtain a sinusoidal plot. This is expected because we obtained the direction of maximum and minimum variance from the eigen vectors and when we vary the theta(direction) the variance oscilates between this minimum and maximum value.