A Report for Foundations of Machine Learning Lab 1

Name：Ziyi Guo
Email：zg2u21@soton.ac.uk
School of Electronics and Computer Science

1 Preliminary

In this part, works on basic programming in matrix manipulations was done to clarify essential matrix properties.

In the first section, for given vector *x*[1,2] and *y*[-2,1], basic manipulations to acquire the scholar product of the two vectors and the norm of the vector *x* were carried out. The results proved that the scholar product of *x* and *y* equals the multiplication of matrices x^T and *y*.

In the second section, a matrix *B*[[3,2,1],[2,6,5],[1,5,9]] was built, a 3*1 random matrix *z* [0.39864556 0.22571898 0.99052118] was generated and matrix *v*[2.6378958 7.10421088 10.44193106] was their multiplication. The outputs showed that the subtraction of *B* and B^T was zero, indicating that the matrix *B* is symmetric and for the shape of matrix *v*, the row is decided by *z* and col is decided by *B*.

In the last section, the trace, determinant, eigenvalue matrix *D* and eigenvector matrix *U* of matrix *B* was worked out. It was found that the trace of matrix *B* equals the sum of its three eigenvalues *D*[1], *D*[2], *D*[3] and the determinant of *B* equals the multiplication of the eigenvalues. Also, the multiplication of the eigenvector *U*[:,0] and the eigenvector *U*[:,1] was found to be zero, indicating that for a symmetric matrix *B*, its eigenvector matrices are orthogonal. To formally prove this conclusion, the general situation needs to be discussed:

Given x_1, x_2 are the eigenvector of symmetric matrix A, $Ax_1 = \lambda_1 x_1$, $Ax_2 = \lambda_2 x_2$. If $\lambda_1 \neq \lambda_2$, then prove $x_1^T x_2 = 0$:

A is symmetric matrix, thus

$$\lambda_1 x_1^T = (\lambda_1 x_1)^T = (Ax_1)^T = x_1^T A^T = x_1^T A$$

Then,

$$\lambda_1 x_1^T x_2 = (x_1^T A)x_2 = x_1^T(Ax_2) = x_1^T(\lambda_2 x_2) = \lambda_2 x_1^T x_2$$

Therefore,

$$(\lambda_1 - \lambda_2)x_1^T x_2 = 0$$

Given $\lambda_1 \neq \lambda_2$, then

$$x_1^T x_2 = 0$$

In conclusion, for a real symmetric matrix, its two eigenvectors corresponding to different eigenvalues are orthogonal.

2 Random Numbers and Uni-variate Density

In this part, random numbers were generated and used to plot histograms through programming.

In the first section, a uniform distributions with 1000 rows of data ranking from [0,1] was generated and used to plot two histograms, ax[0] with a bin of 4 and ax[1] with a bin of 40. The figures of several running is as below:

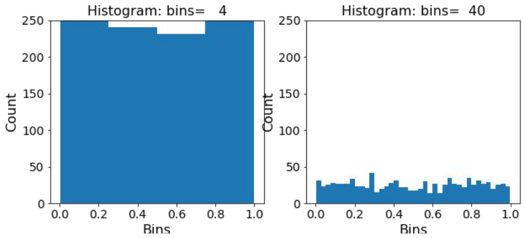

Figure 1. Histogram of the First Random Distribution

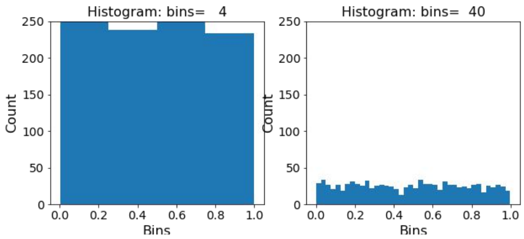

Figure 2. Histogram of the Second Random Distribution

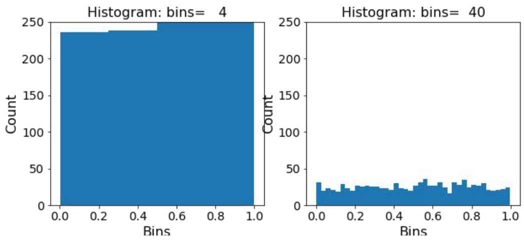

Figure 3. Histogram of the Third Random Distribution

It is observed that though the data is from a uniform distribution, the histogram does not appear flat. This is because the data of a uniform distribution has the equivalent probability to be distributed between the equal lengths, but the data is often not distributed uniformly in one observation, especially when there is no large enough amount of data. Therefore, there is actually different amount of data in each length and the histogram seems not to be flat. Also, due to the random generation algorithm, the data is different in every running and the distribution is actually changing as well. Therefore, the histogram is slightly different every time.

To start with more data, “x = np.random.rand(100000, 1)” was used to replace the original statement and a lot more data was put into the distribution. The figure of the running is as below:

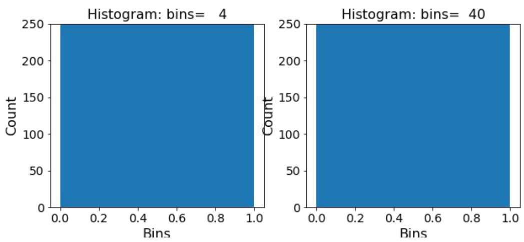

Figure 4. Histogram of the Fourth Random Distribution

It is observed that the histogram becomes much more flat and seems to be a figure of uniform distribution. This is because when there is large enough amount of data, the frequency is infinitely close to probability and the distribution becomes closer to uniform distribution.

In the second section, the adding and subtracting results of two sets of uniform random numbers were put into histogram. Different amounts of data were tested and the figures are as below:

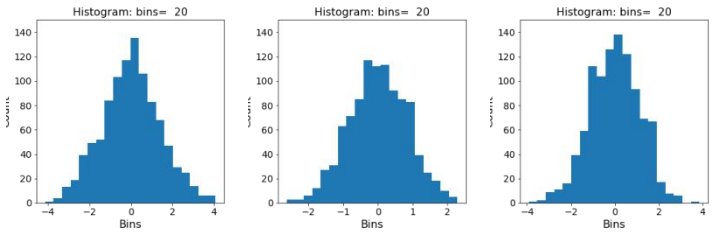

Figure 5. Histograms of 12,8,4 Random Numbers Distribution

It is observed that the more data there is in the sets, the more likely the subtraction of the numbers in the two uniform distribution is to be zero, indicating the distributions are more similar.

3 Uncertainty in Estimation

In this part, sub-graph was drawn to show the variance of a uni-variate Gaussian density. The figure below in which x axis represents the amount of samples and y axis represents the corresponding variance shows the result:

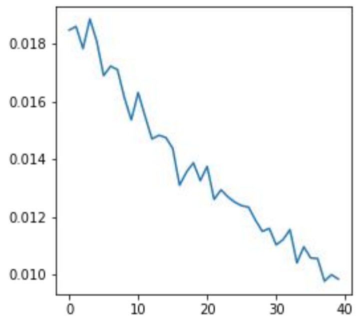

Figure 6. Sub-graph of the Variance of a Uni-variate Gaussian density

It is observed that even if there are some noises, the variance of a uni-variate Gaussian density is generally decreasing when more data is generated.

4 Bi-variate Gaussian Distribution

In this part, the contour graphs of several bi-variate Gaussian distributions were drawn. The figures of the distributions $N\left(\begin{bmatrix} 2.4 \\ 3.2 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\right)$, $N\left(\begin{bmatrix} 1.2 \\ 0.2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\right)$ and $N\left(\begin{bmatrix} 2.4 \\ 3.2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\right)$ are as below:

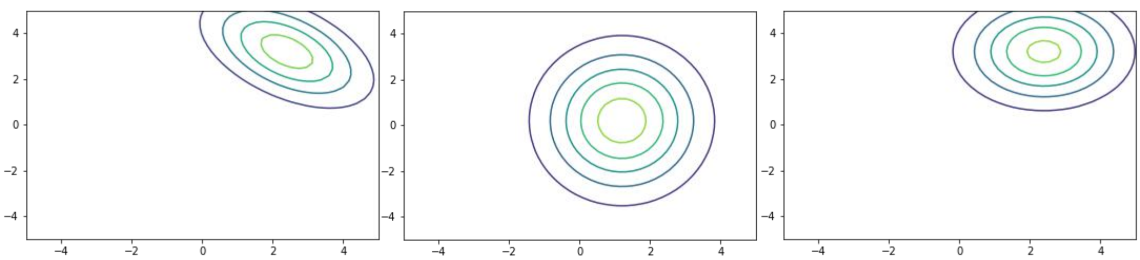

Figure 7. Contour Graphs of Three Bi-variate Gaussian Distributions

5 Sampling from a multi-variate Gaussian

In this part, samples from a multi-variate Gaussian density and a multi-variate Gaussian density with mean $m = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and covariance matrix $C = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ were taken to draw scatter plots. And the samples from the latter were used to draw a subplot of the distribution of its projection variance on vector $u = [\sin\theta \quad \cos\theta]^T$, parameterized by θ ranging from $[0,2\pi]$. The figures are as below:

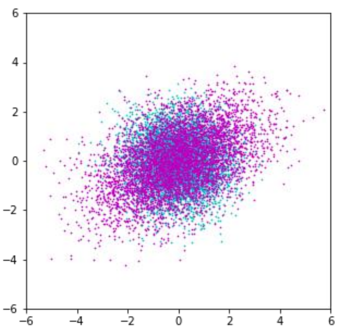

Figure 8. Scatter Plots of Multi-variate Gaussian Distributions

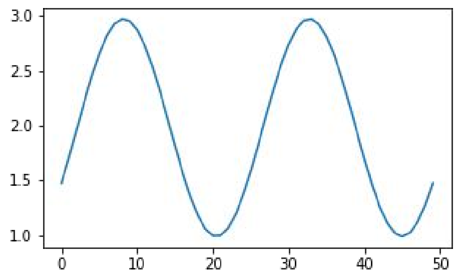

Figure 9. Subplot of the Variance of Projection Data

It is observed that the maxima and minima of the resulting plot is 3.0 and 1.0. The eigenvalues and the eigenvectors of the covariance matrix C were also found to be $\begin{bmatrix} 3.0 \\ 1.0 \end{bmatrix}$ and $\begin{bmatrix} [\,0.71 & -0.71\,], \\ [\,0.71 & 0.71\,] \end{bmatrix}$, indicating that the biggest and smallest eigenvalues of the covariance matrix C are corresponding to the maxima and minima of the projected variance and vectors $\begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix}$, $\begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix}$ are the corresponding eigenvectors. Also, the shape of the subplot seemed sinusoidal. This is because the for $Y \sim N(m,c)$, $yp = uY$, there is

$$yp \sim N(\mathrm{u}m, \mathrm{u}Cu^T)$$

Then, covariance matrix becomes

$$[sin\theta \quad cos\theta] \times \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \times \begin{bmatrix} sin\theta \\ \cos\theta \end{bmatrix} = 2\sin\theta\cos\theta + 2 = 2\sin 2\theta + 2$$

In conclusion, the variance of the projection is a sinusoidal function about variable θ.