[COMP6245 Foundation of Machine Learning]
LabOne: Multi-variate Gaussian

31240232 / Sunwung Lee

MSc Artificial Intelligence, University of Southampton

14 October 2019

1 Linear Algebra

The last line of the advanced topic, singular value decomposition(SVD) example code $\mathbf{UU^T}$, is identified an identity matrix($\mathbf{I}$). $\mathbf{U}$ is the eigenvector matrix of $\mathbf{B}$, and the result of $\mathbf{UU^T}$ must be an orthogonal matrix because it is an identity matrix. Next, SVD can be expressed by the following formula.

$$\mathbf{A} = \mathbf{U\Sigma V^T}$$
$$\mathbf{A} = m \times n \text{ Rectangular matrix}, \quad \mathbf{U} = m \times m \text{ orthogonal matrix}$$
$$\mathbf{\Sigma} = m \times n \text{ diagonal matrix}, \quad \mathbf{V^T} = n \times n \text{ orthogonal matrix}$$

The great advantage of SVD is that it can be used widely for all rectangular matrices of $m \times n$ consisting of real or complex spatial elements. In addition, SVD allows arbitrary $\mathbf{A}$ to be divided into layers based on the amount of information. Finally, the size of $\mathbf{B}$ is predictable. 1) The determinant is not 0. 2) There is an inverse matrix. 3) There are 3 eigenvalues. 4) rank = 3. Therefore, $\mathbf{B}$ is a 3×3 square matrix.

2 Random Numbers and Univariate Distributions

First of all, briefly, the definitions of the rand and randn functions.

1) rand: Create an array of the given shape it with random samples from a uniform distribution.
2) randn: Return a sample (or samples) from the "standard normal" distribution.

This can be confirmed by histogram. However, the appearance of this histogram depends on the number of bins and the number of samples. An important parameter of the hist function, bins, is defined as "int or sequence or str, optional. if an integer is given, bins + 1 bin edges are calculated and returned consistent with numpy.histogram", which provides an accurate view of the data in
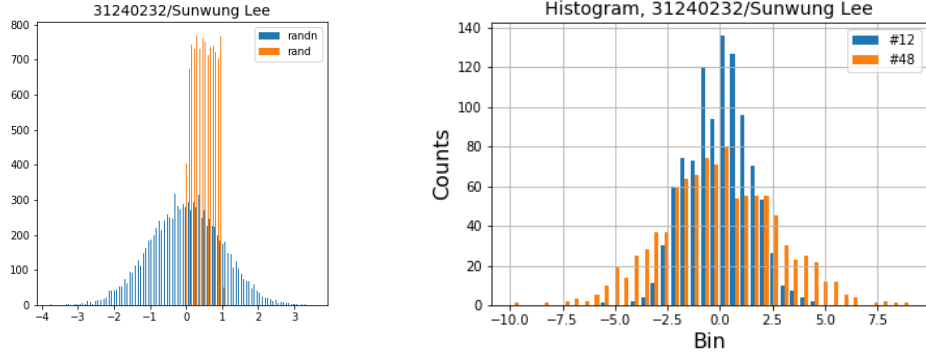
Figure 1: L: difference between rand, randn / R: change of histogram according to random numbers

histogram.

In addition, as the number of samples increases, the features of rand and randn functions become more apparent. For example, in the case of randn, the larger the number of samples, the more clearly the normal distribution with a mean of 0 and a standard deviation of 1 can be identified. Through Figure 1 it can certainly compare the features of the rand and randn functions. Histogram shows the following difference. A histogram was drawn using the difference of the sum of the random variables through the rand function. Naturally, the larger the total number of repetitions, N, the clearer the normal distribution curve. If uniform random number(URN) is 12, the mean is close to 0 and the standard deviation has $\sqrt{2}$. However, if the URN is getting larger, the standard deviation increases. Conversely, the smaller the standard deviation approaches zero.

3   Uncertainty in Estimation

The smaller the number of arbitrary variables xx, defined as xx in code, the greater the value of the variance. This is because it is not a sufficient number of variables that cannot draw a normal distribution tendency. In addition, the variance of the variances of arbitrary variable values can
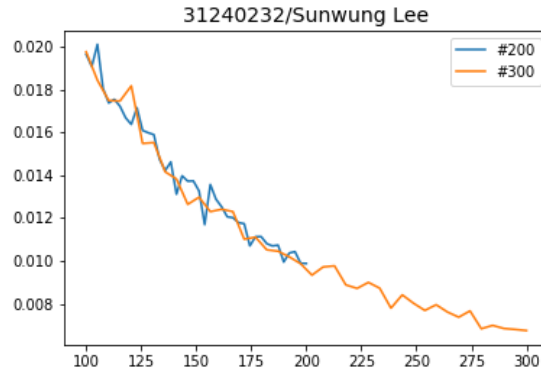

Figure 2: Uncertainty in Estimation

be analyzed. In other words, a small variance value of the variances means that when the same number of data is repeated n times, the resulting data converges further to the mean value. In the same way, this means that xx better represents a normal distribution.

This can be expected to hold the greater amount of data and to better catch the characteristics of that data.

4 Bivariate Gaussian Distribution

Why should Bivariate Gaussian Distribution use? Suppose there are two variables in the dataset (x_1, x_2), where x_1 is normally distributed and x_2 is also normally distributed. In this case, exception data may be wrongly judged. To avoid this, multivariate Gaussian distributions are used. For example, in the Bivraiate Gaussian Distribution, a researcher can check the distribution of variables for students' height (x_1), weight (x_1), and (x_1, x_2) following a normal distribution.

Let's look at the function part of the example python code.

Gauss2D(x, m, $\mathbf{C}$): The return value follows the mean m, variance $\mathbf{C}$ in Gaussian equation. Moreover, contours represent groups of the same level. Use the meshgrid function in python to draw this.
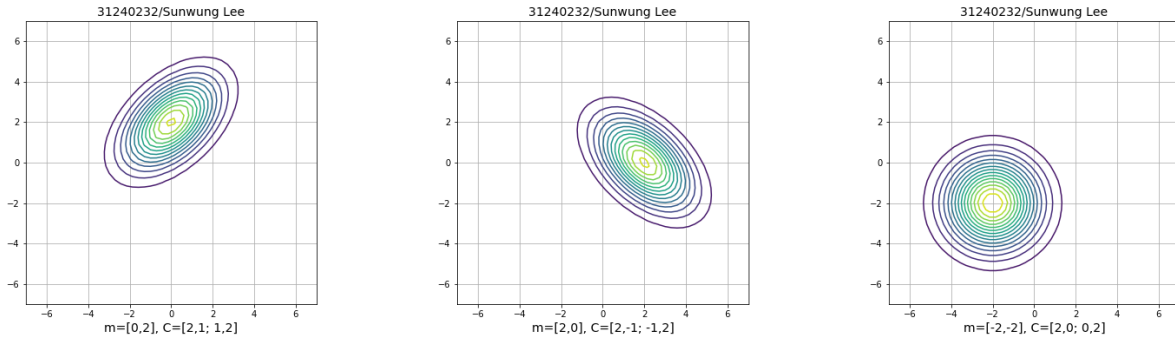

Figure 3: Examples of bivariate Gaussian distribution

The figure above shows the results of experiments using three different examples of means(m) and covariance matrices($\mathbf{C}$) using the contour function. The centre of the ellipse / circle is determined by m, and the direction of the ellipse is determined by the value of the $\mathbf{C}$.

5 Sampling from a Multivariate Gaussian Distribution

The covariance matrix($\mathbf{C}$) can be expressed as the product of the lower triangular matrix($\mathbf{A}$), by Cholesky decomposition. The Python example code generates a random number($\mathbf{X}$) that follows a normal distribution, and $\mathbf{Y}$ is expressed as the product of $\mathbf{X}$ and $\mathbf{A}$. The reason for decomposing a matrix by Cholesky decomposition is to make the calculation more convenient. This simplifies the system matrix and makes it easier to analyze.
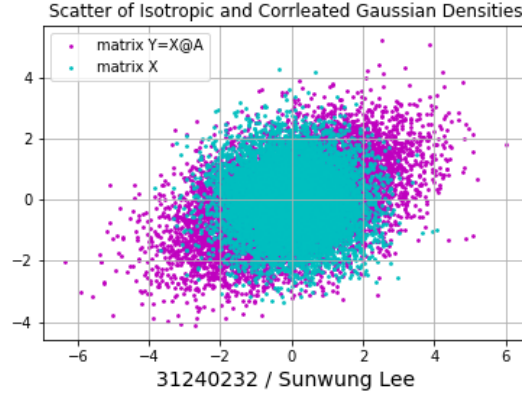
Figure 4: Scatter plot of $\mathbf{X}$ and $\mathbf{Y}(= \mathbf{XA})$

The code results show that multiplying $\mathbf{A}$ and $\mathbf{X}$ gives the effect of $\mathbf{C}$ on $\mathbf{X}$ and the shape of data changes ellipse.

6   Distribution of Projections

In the Python example code, thRange is set to 50 divided by 0 to 2π. In addition, theta is each divided by 50, which constitutes a $\mathbf{u}$. The pVars array stores the variance of the values of the projection of the $\mathbf{Y}$ onto the vector $\mathbf{u}$.
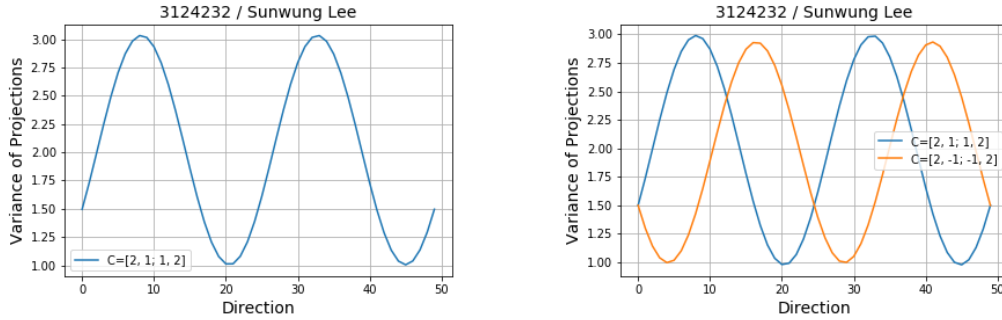

Figure 5: Examples of distribution of projection

As you can see on the left figure, the value of pVars is represented in the form of a sine function. The calculation shows that the covariance matrix has eigenvalues of 3 and 1, so eigenvectors can find two vectors, $x_2 = x_1$ and $x_2 = -x_1$.

In addition, $\mathbf{C} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$, the phase of the sine waveform is changed, and when the two graphs are drawn at the same time, it appears on the right figure.

4