CO544 Machine Learning – Lab 02

E/20/197 - Kawya A.H.D.

What do you observe for the last command above (i.e. print(np.dot(U[:,0], U[:,1])))? Can you formally prove that this is the result you would expect for the specific structure in matrix B?

B = [[3. 2. 1.]

[2. 6. 5.]

[1. 5. 9.]]

This computes the 1st two eigen vectors of B. Since Eigenvectors of a symmetric matrix are orthogonal and B is a symmetric matrix, the output is approximately zero. (-2.7755575615628914e-16)

Give an example of where Singular Value Decomposition (SVD) is used.

Data Compression: Reducing the size of data sets while preserving essential information.

Random Numbers and Uni-variate Densities

Though the data is from a uniform distribution, the histogram does not appear flat. Why?

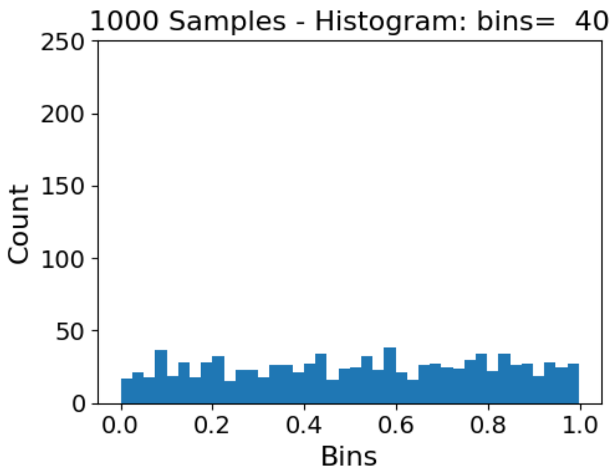
Because we are plotting a finite random sample, we are seeing random variation; called sampling noise. Although the underlying distribution is perfectly uniform, small random differences in the sample can cause some bins to be slightly higher than or lower than others. The key point to be considered is a histogram approximates the underlying distribution. The smaller the sample, the less accurate that approximation will be.

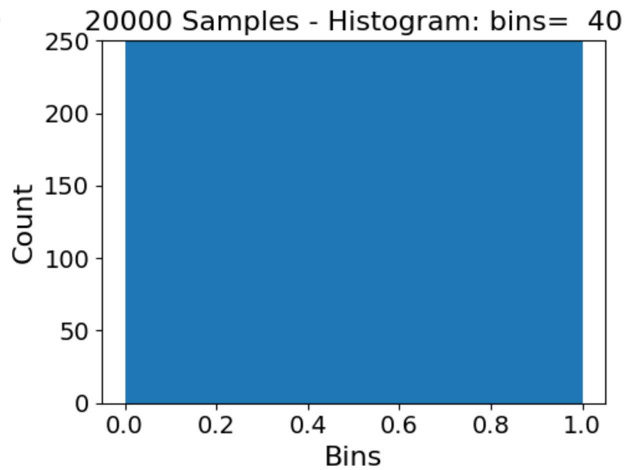
Every time you run it, the histogram looks slightly different. Why?

We are generating new random values. Even though they come from the same uniform distribution, the specific numbers differ each time. Each sample will have slightly different bin counts. Hence the histogram changes slightly on every run.

How do the above observations change (if so how) if you had started with more data?

If you increase the number of samples, Histogram becomes flatter and smoother. Then bin heights will become more equal.





What do you observe? How does the resulting histogram change when you change the number of uniform random numbers you add and subtract? Is there a theory that explains your observation?

The histogram appears bell-shaped and symmetric around 0. It looks approximately normal. Even though we are summing uniform random numbers, the distribution of their sum looks Gaussian. This is due to the Central Limit Theorem.

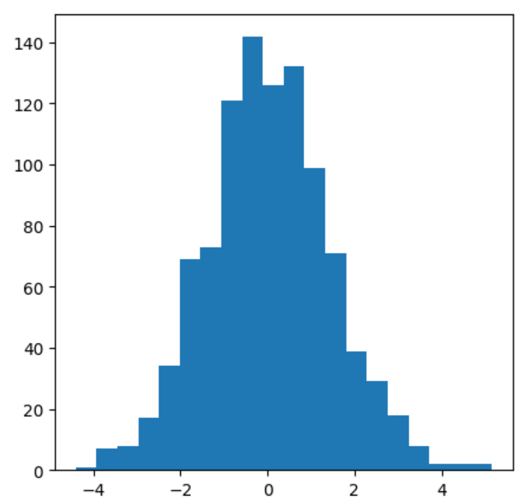


Figure 2 12 random numbers

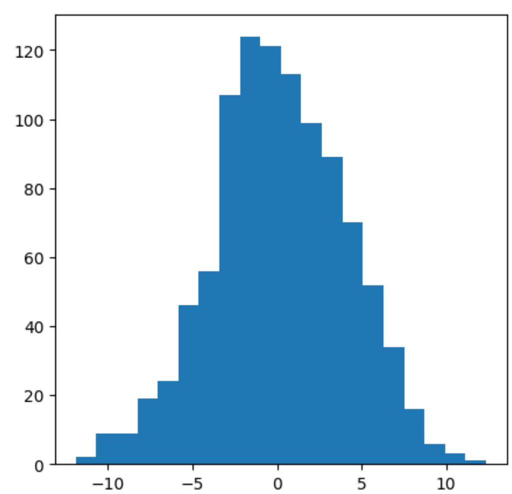
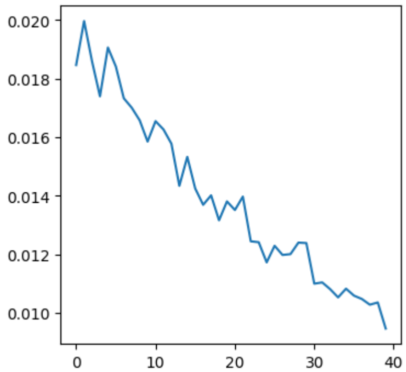


Figure 1 100 random numbers

When we increase the number of random values from 12 to 100, the resulting histogram becomes even more bell-shaped. Although the spread increases, shape gets smoother. If we decrease the number, the histogram is less smooth. Since fewer values are being summed, it is more irregular.

Uncertainty in Estimation



This plot shows how the variance of the sample variances decreases as the sample size increases.

Bi-variate Gaussian Distribution

Draw contours of the given distributions.

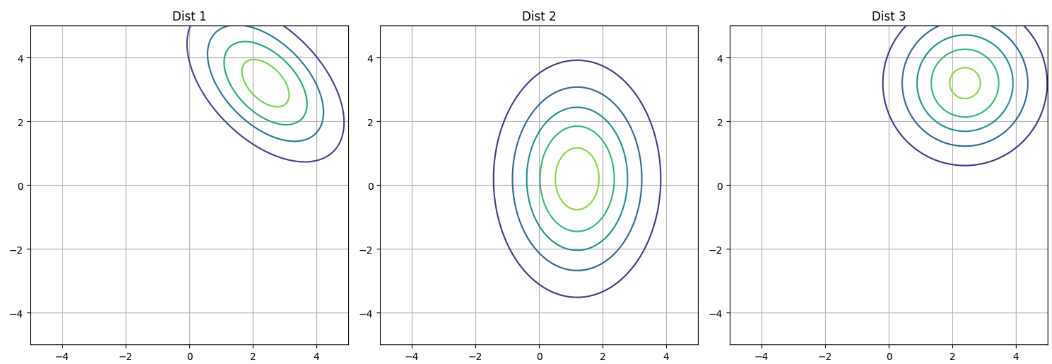
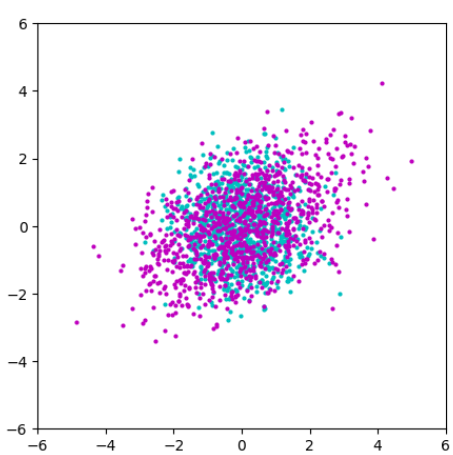


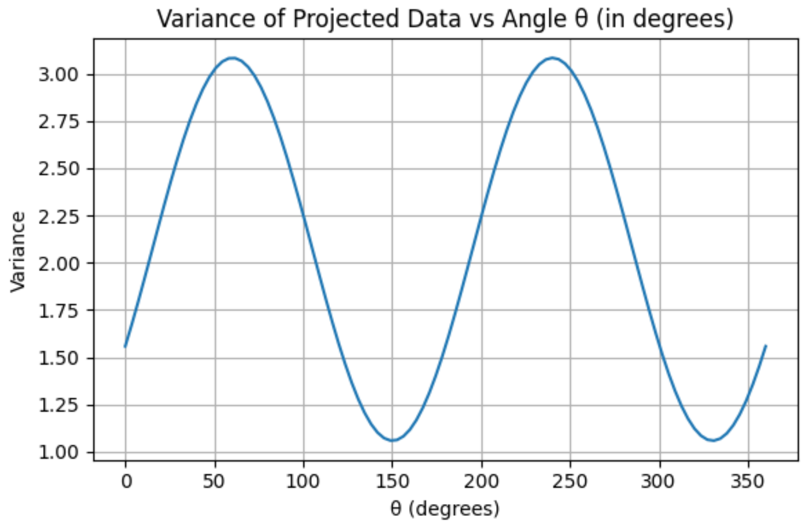
Figure	**Covariance Matrix**	**Shape of the Contours**	**Meaning**
1	[[2, -1], [-1, 2]]	Tilted ellipse (Opposite direction)	Negative Correlation
2	[[2, 0], [0, 4]]	Circle	Variables have the same variance, uncorrelated.
3	[[2, 0], [0, 2]]	Axis-aligned ellipse	Variables are uncorrelated.

Sampling from a multi-variate Gaussian



This is how multivariate data is simulated in machine learning.

Distribution of Projections



What are the maxima and minima of the resulting plot?

Maxima ≈ 3. Minima ≈ 1.

Compute the eigenvalues and eigenvectors of the covariance matrix C

Eigenvectors (Columns):

[[0.70710678 -0.70710678]

[0.70710678 0.70710678]]

Can you see a relationship between the eigenvalues and eigenvectors and the maxima and minima of the way the projected variance changes?

The maximum projected variance occurs when we project the data along the direction of the eigenvector with the largest eigenvalue. The minimum projected variance occurs along the eigenvector with the smallest eigenvalue.

The shape of the graph might have looked sinusoidal for this two-dimensional problem. Can you analytically confirm if this might be true?

Yes. For a 2D Gaussian, when projecting onto a unit vector $u(\theta) = [sin\theta \quad cos\theta]^T$, the projected variance is: $\sigma^2(\theta) = u^T C u = [sin\theta \quad cos\theta]\begin{bmatrix} a & b \\ c & d \end{bmatrix}[sin\theta \quad cos\theta]$

$\sigma^2(\theta) = a\sin 2\theta + 2b \sin\theta \cos\theta + d \cos 2\theta$

This has a form of sinusoidal.