

CO544 Machine Learning Lab 02 Report

Name: Ravindu Pathirage Index Number: E20280

Objective

The objective of this laboratory exercise was to study two uses of the properties of multivariate Gaussian densities:

1. **Sampling:** Drawing samples from a multivariate Gaussian distribution.
 2. **Projection:** Understanding how projections transform Gaussian distributions.
-

Preliminaries

Before diving into the experiments, we revisited some basic matrix and vector operations in Python, such as:

- Dot product, vector norms, matrix symmetry, matrix multiplication
- Quadratic forms and trace of matrices
- Determinant, eigenvalues, eigenvectors

We verified that the eigenvectors of a symmetric matrix form an orthogonal set using `np.linalg.eig()` and validated their orthogonality by computing $UU^T = I$.

1. Random Numbers and Uni-variate Densities

Uniform Random Numbers

We generated 1000 uniform random numbers and plotted histograms with varying bin sizes.

Observations:

- The histograms did not appear flat despite the uniform distribution, due to the limited sample size.
- Increasing the number of bins revealed finer variations in the distribution.
- More data would result in flatter, more uniform-looking histograms.

Sum of Uniform Random Numbers

We generated random variables by summing and subtracting 12 uniform random numbers:

- Resulting distribution approximated a Gaussian due to the **Central Limit Theorem (CLT)**.
 - Increasing the number of added/subtracted random numbers made the histogram appear increasingly Gaussian.
-

2. Uncertainty in Estimation

We examined the variability of variance estimates of normally distributed random data as a function of sample size.

Observations:

- As sample size increased, the variance of the estimated variances decreased.
 - Demonstrates that more data leads to more reliable statistical estimates.
-

3. Bi-variate Gaussian Distribution

We implemented a function to evaluate the probability density of a 2D Gaussian distribution and plotted contour plots for visualization.

Contour Plots for Different Parameters:

$$\begin{aligned} 1. \mu &= [2.4, 3.2], \Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \\ 2. \mu &= [1.2, 0.2], \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \\ 3. \mu &= [2.4, 3.2], \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \end{aligned}$$

Observations:

- Off-diagonal elements in Σ introduce correlation, resulting in elliptical contours with rotated axes.
-

4. Sampling from Multivariate Gaussian

Using Cholesky decomposition, we transformed standard Gaussian samples into samples from a Gaussian distribution with:

$$\mu = [0, 0], \quad \Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Scatter Plot:

- Original samples (cyan) were circular.
 - Transformed samples (magenta) formed an elongated elliptical cloud, indicating correlation.
-

5. Distribution of Projections

We projected the 2D Gaussian samples onto unit vectors parameterized by θ in $[0, 2\pi]$ and plotted the variance of the projected data.

Observations:

- The variance of projections varied sinusoidally with θ .
- **Maximum and Minimum Variances** corresponded to the **eigenvalues** of Σ .
- Projection directions at maxima/minima aligned with the **eigenvectors** of Σ .

Analytical Confirmation:

- The sinusoidal nature stems from the quadratic form of projections involving symmetric covariance matrices.
-

Conclusion

This lab demonstrated:

- Practical generation and visualization of multivariate Gaussian distributions.
- The fundamental connection between covariance matrices, eigenvalues/eigenvectors, and projections.
- How sample size affects the uncertainty of statistical estimates.

This foundational understanding is crucial for advanced topics in machine learning, such as Principal Component Analysis (PCA) and Gaussian Mixture Models (GMMs).
