Class #18: Statistical Analysis, Eigenvalues and Eigenvectors

Purpose: The objective of this experiment is to familiarize yourself with common statistical analysis terms and revisit linear curve fitting with multiple inputs.

Background: Before doing this experiment, students should be able to
- Compute determinants and perform matrix multiplication
- Use Matlab to perform compute eigenvalues and eigenvectors
- Review the background for the previous experiments.

Learning Outcomes: Students will be able to
- Understand the properties of Multi variate Gaussian distributions
- Understand the concept of Eigenvalues and Eigenvectors
- Use matrix mathematics to determine how the Eigenvalues and Eigenvectors relate to the covariance matrix

Equipment Required:

- Matlab

Keywords:

- Multivariate Gaussian
- Eigenvalues
- Eigenvectors
- Mean
- Covariance Matrix

Helpful links for this experiment can be found on the course website under Class #18.

Part A – Statistical Analysis (Multivariate Gaussian distribution)

Background

In the last experiment, we worked with the Gaussian distribution among others. In the last lab however, the probability distribution was dependent on only one variable. However, to understand more interesting phenomena, we need to look at multiple variables at the same time. For instance, if you want to predict the weather, you will be looking at different climate conditions, like the winds, the temperature, the humidity, etc. If you want to do process an image, the number of variables is the number of pixels in the image (more about this in the next experiment).

The generalization of Gaussian random variables to two dimensions is straightforward. Recall that the one-dimension distribution was given by

$$f\left(x\right)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\left(x-\overline{x}\right)^2}{2\sigma^2}}$$

where $\overline{x}$ is the **mean** (average), which we used in experiment 16 (previous class) when determining the slope and intercept of the least squares fit, and σ is the **standard deviation**.

The standard deviation is a measure of how 'spread out' the distribution appears. Considering the following two figures, Figure A-1 is an example of 'wide' Gaussian distribution with a 'larger' standard deviation and Figure A-2 is an example of a 'narrow' Gaussian distribution with a 'smaller' standard deviation. We can clearly see that a smaller standard deviation looks much 'narrower'.

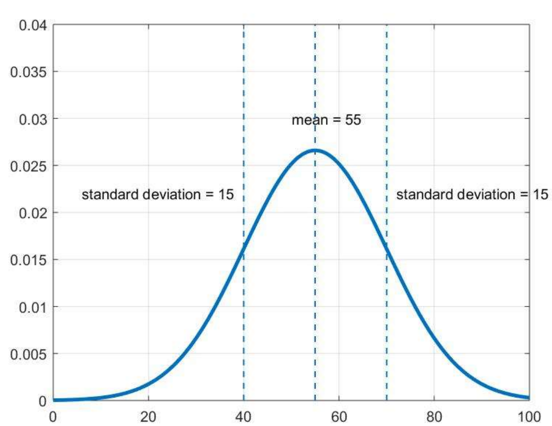


Figure A-1: Gaussian Distribution
Mean = 55, Std. Dev. = 15

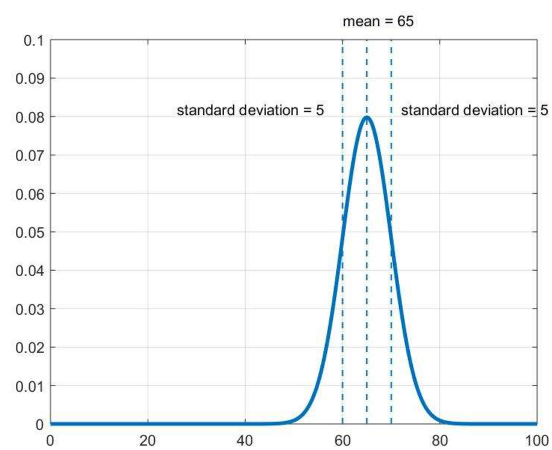


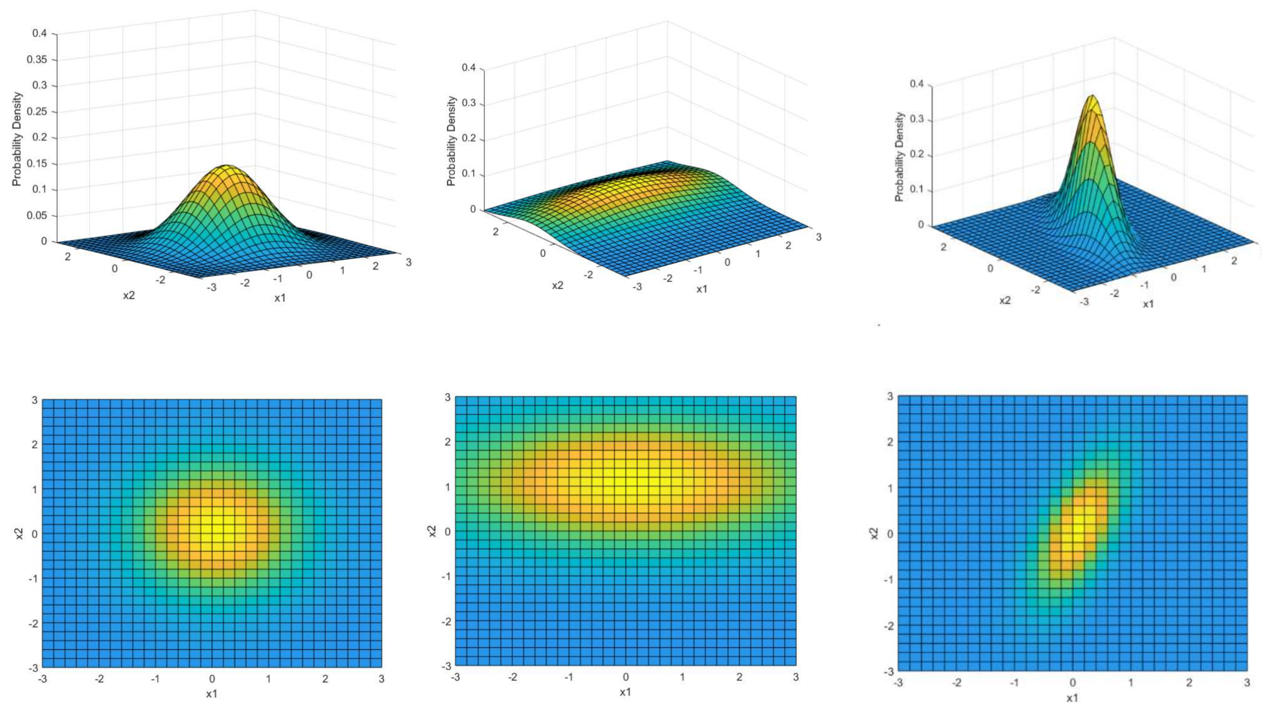
Figure A-2: Gaussian Distribution,
Mean = 65, Std. Dev. = 5

In two dimensions the Gaussian distribution can be written in the following form

$$f(x) = \frac{e^{-(x-\bar{x})^T\Sigma^{-1}(x-\bar{x})/2}}{\sqrt{det(2\pi\Sigma)}}$$

where x is now a vector, $\bar{x}$ is **the mean** which is also a vector with the same dimension that the data x and Σ is a square matrix with the dimensions of x. This matrix is called the **covariance matrix.** Notice that this is a generalization of the previous expression in one dimension. Indeed, in one dimension the covariance matrix Σ becomes the square of the standard deviation σ. (If you know your Greek letters this is even more clear: Σ is the capital version of σ). In the case of one dimension transposing a vector does not alter the number, and the determinant of a scalar is the scalar itself. **<u>Take some time to compare these two expressions and convince yourself that they are the same.</u>**

Let us looks at a couple of examples of distributions for different combinations of means and covariance matrices.



. $\bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\bar{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ \Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$ $\bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ \Sigma = \begin{bmatrix} 0.25 & 0.3 \\ 0.3 & 1 \end{bmatrix}$

<u>Important note</u>: A fair bit of mathematics is applied when studying probability distributions and you will see some of that in a future course (ECSE-2500). We don't expect you to have a deep understanding of the mathematics, but we do expect to recognize common distribution functions/histograms and concepts like mean and some intuition on covariance matrix.

We will use the three datasets in the lab website to further understand the relationship between the distribution, the mean and the covariance matrices. Since there are a large number of data points, we will use Matlab to do the heavy lifting. The commands in Matlab are straightforward. For some array called `data`,

The mean is determined using the `mean` command

```
>> mean(data)
```

The covariance matrix is determined using the `cov` command

```
>> cov(data)
```

Exercise

1) For the multivariate distributions with $\bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\bar{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$, describe how the mean and the covariance matrix get translated in the plots. For the third one it gets trickier and a full answer will be more clear at the end of this lab. However, feel free to give it a try at this point.

2) Download the Excel and load it into Matlab. Each sheet has data from a different distribution. Plot the dataset. Use the above commands to determine the mean, and the covariance matrices for each data set. Approximate your result to the first decimal. Do these match your intuition of the data distribution? Explain.

Part B – Eigenvalues and Eigenvectors

Background

The eigenvalues and eigenvectors are important properties of a matrix that help us understand its structure. These concepts are used in multiple domains of engineering as we will see in this lab and the next ones. For a square matrix A an eigenvector is a vector $v \neq 0$ is a vector such that if we multiply it by the matrix it returns a scaled version of that vector

$$Av = \lambda v$$

The scalar λ is the eigenvalue corresponding to the eigenvector v. We can rewrite the previous expression in the following form which is going to be more useful to compute the eigenvalues

$$(A - \lambda I)v = 0.$$

If $A - \lambda I$ is invertible then we have $v = (A - \lambda I)^{-1}0 = 0$. Since we are looking for vectors $v \neq 0$ we need $A - \lambda I$ to be noninvertible. If you recall experiment 15 one of the matrices was not invertible and the reason for that is that the determinant was zero. So, we can compute the eigenvalues of a matrix by solving the following equation

$$\det(A - \lambda I) = 0.$$

Let us see a couple of examples

a) $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\quad A - \lambda I = \begin{pmatrix} 1-\lambda & 0 \\ 0 & 1-\lambda \end{pmatrix}$, $\quad \det(A - \lambda I) = (1-\lambda)^2$,

By making $\det(A - \lambda I) = (1-\lambda)^2 = 0$ we get that the eigenvalues are $\lambda = 1$

b) $A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}$ $\quad A - \lambda I = \begin{pmatrix} 0-\lambda & 1 \\ -2 & -3-\lambda \end{pmatrix}$, $\quad \det(A - \lambda I) = -\lambda(-3-\lambda) + 2 = \lambda^2 + 3\lambda + 2$

By making $\det(A - \lambda I) = \lambda^2 - 3\lambda + 2 = 0$ we get that the eigenvalues are $\lambda = \frac{-3 \pm \sqrt{1}}{2} = -1, -2$

Once we have the eigenvalues, we can proceed to compute the eigenvectors. Write the eigenvector equation in matrix form and solve for v. Continuing with the previous examples

a) $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, with eigenvalue $\lambda = 1$, $Then\ write\ A - \lambda I = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. Which vectors are the eigenvectors? Any vector v such that $(A - \lambda I)v = 0$. In this case all vectors are eigenvectors. Not surprising, since for any vector v we have that $Iv = v$.

b) $A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}$, for $\lambda = -1$, we have that $(A - \lambda I)v = \begin{pmatrix} -1 & 1 \\ -2 & -4 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

From here we get two equations $v_2 = -v_1$ and $-2v_1 - 4v_2 = 0$ Both equations are $v_1 = -v_2$. This means that any vector that satisfies that property is an eigenvector. For instance, $v = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

We can also get the other eigenvector which comes from repeating the same analysis but writing $(A - \lambda I)v$ with λ=-2.
$(A - \lambda I)v = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -2v_1 \\ -2v_2 \end{pmatrix}$, we get $v = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$

It is important to note that the eigenvalues that we chose are only one possibility. Actually any scaled version of them is also an eigenvalue, for instance $v = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$ and $v = \begin{pmatrix} 3 \\ -6 \end{pmatrix}$. You can verify these computations in Matlab using the command [V,D] = eig(A). The matrix V will contain the eigenvectors and the matrix D the eigenvalues.

Exercise:

Compute the eigenvalues and eigenvectors of the following matrices. Verify the results with Matlab and by computing Av. Is there any difference between your computations and Matlab's result? Explain.

a) $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$

b) $A = \begin{pmatrix} -6 & 3 \\ 4 & 5 \end{pmatrix}$

c) $A = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$

Part C – Eigenvalues, Eigenvectors, and the Covariance Matrix

Background

We saw in Part A that the distribution of the data of a multivariate Gaussian random variable depends on the covariance matrix. In Part B, we discussed eigenvalues and eigenvectors. The goal of this part is to relate these notions.

Exercise: Compute the eigenvalues and eigenvectors of the covariance matrices computed in Part A. What are the principal axes of the data distribution for the datasets? How do you interpret these results?