# Statistical Pattern Recognition
## Assignment 1

Ali Gholami

Department of Computer Engineering & Information Technology
Amirkabir University of Technology

*http://ceit.aut.ac.ir/~aligholamee*
*aligholamee@aut.ac.ir*

**Abstract**

This is an introductory assignment to the world of *Statistics* and *Probability* in the context of *Pattern Recognition*. We'll introduce some key concepts like *Probability Distribution Function, Cumulative Distribution Function, Probability Density Function, Probability Mass Function, Joint Probability Density Function, Joint Cumulative Density Function, Marginal Density* & more details as the probabilistic point of view. Furthermore, we'll review the concepts of *Expected Value, Variance, Standard Deviation, Covariance & Correlation of Random Variables(e.g. Random Vectors), Univariate & Multivariate Gaussian Distribution, Total Probability & Bayes Theorem, Geometric & Mahalanobis Distances, Central Limit Theorem, Independence & Correlation* as the statistics point of view. Also, a principal concept called *Linear Transformation* is discussed. The relationship between these fields is far more important than each separately.

**Keywords.** *PDF, PMF, JPDF, JPMF, CDF, JCDF, Covariance Matrix, Correlation Coefficient, Correlation, Variance, Expected Vector, Gaussian Distribution, Marginal Probability, Linear Tranformation, Eigenvector, Eigenvalue, Rank.*

# 1 Expectation & Variance

A random variable $X$ has $E(X) = -4$ and $E(X^2) = 30$. Let $Y = -3X + 7$. Compute the following.

(a) $V(X)$

(b) $V(Y)$

(c) $E((X+5)^2)$

(d) $E(Y^2)$

## Solution

The main equation to calculate the *Variance* of a random variable $X$ is given in 1.1.

$$V(X) = E[(X - E[X])^2] \tag{1.1}$$

Expanding the equation 1.1, we'll have the equation 1.2 using simple calculus.

$$V(X) = E[X^2 + E[X]^2 - 2XE[X]]$$
$$V(X) = E[X^2] + E[X]^2 - 2E[X]^2$$
$$V(X) = E[X^2] - E[X]^2 \tag{1.2}$$

---

(a) The equation 1.2 can be conducted directly to compute the *Variance*. Replacing the values from the problem description we get the following as result.

$$V(X) = E[X^2] - E[X]^2$$
$$V(X) = 30 - 16 = 14$$

---

(b) It is important to mention the *Linearity* of *Expectation* operator as formally described in 1.3.

$$E[aX + b] = aE[X] + b \tag{1.3}$$

Using property 1.3, we can write the *V(Y)* as

$$V(Y) = V(-3X + 7) = E[(-3X + 7)^2] - E[-3X + 7]^2$$
$$V(Y) = E[9X^2 + 49 - 42X] - E[-3X + 7]^2$$
$$V(Y) = 9E[X^2] + E[49] - 42E[X] - 9E[X]^2 - E[49]$$
$$V(Y) = 9 * 30 + 49 - 42 * (-4) - 9 * 30 - 49$$
$$V(Y) = 168$$

---

(c) Expanding the internals of the expectation, we'll get the following.

$$E[(X + 5)^2] = E[X^2 + 10X + 25]$$
$$E[(X + 5)^2] = E[X^2] + 10E[X] + E[25]$$
$$E[(X + 5)^2] = 30 + 10 * (-4) + 25$$
$$E[(X + 5)^2] = 15$$

---

(d) Same as above, we'll use 1.2 to get the following.

$$E[Y^2] = E[(3X + 7)^2]$$
$$E[Y^2] = E[9X^2 + 49 - 42X]$$
$$E[Y^2] = 487$$

## 2 Eigenvector & Eigenvalue

(a) Compute eigenvalues and eigenvectors of $A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 9 & -5 \end{bmatrix}$ and compare your results with Matlab outputs.

(b) A $2 * 2$ matrix $A$ has $\lambda_1 = 2$ and $\lambda_2 = 5$, with corresponding eigenvectors $V_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ and $V_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. Find $A$.

### Solution

(a) First of all, we'll find the *eigenvalues* using the *characteristic equation* given in 2.1.

$$|A - \lambda I| = 0 \tag{2.1}$$

Using this equation, all of the diagonal components of matrix $A$ will be decremented by a $\lambda$ term. The determinant of the resulting matrix will be as following.

$$A - \lambda I = \begin{bmatrix} 4-\lambda & 0 & 0 \\ 0 & 2-\lambda & 2 \\ 0 & 9 & -5-\lambda \end{bmatrix}$$

$$|A - \lambda I| = (4-\lambda)(\lambda^2 + 3\lambda - 28)$$

This will result in the following two roots for the $\lambda$ term.

$$\boxed{\lambda_1 = 4} \quad \boxed{\lambda_2 = -7}$$

These are the *eigenvalues* of the given matrix $A$. We'll continue using the *Gaussian Elimination* technique to compute the *eigenvectors* of matrix $A$. According to the equation 2.2 we are looking for all possible vectors that can be substitute with vector X.

$$|A - \lambda I|X = 0 \tag{2.2}$$

Thus, for each *eigenvalue* determined in the previous computations, we'll find the proper *eigenvector* by converting the equation 2.2 to a *Row Echelon Form* and solving the resulting linear system by *Back Substitution*. Using the computed *eigenvalues*, we'll have 2.3 and 2.4

$$A - 4I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & 9 & -9 \end{bmatrix} \tag{2.3}$$

$$A + 7I = \begin{bmatrix} 11 & 0 & 0 \\ 0 & 9 & 2 \\ 0 & 9 & 2 \end{bmatrix} \tag{2.4}$$

Now we can find the proper $X$ for each *eigenvalue*, using *augmented* version of 2.3 and 2.4. We'll have 2.5 and 2.6 as result.

$$E_1 = (A - 4I \mid 0) = \left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 9 & -9 & 0 \end{array}\right] \tag{2.5}$$

$$E_2 = (A + 7I \mid 0) = \left[\begin{array}{ccc|c} 11 & 0 & 0 & 0 \\ 0 & 9 & 2 & 0 \\ 0 & 9 & 2 & 0 \end{array}\right] \tag{2.6}$$

The above matrices will be converted to the *Row Echelon Form* below using *Row Operations.*

$$E_1 = \left[\begin{array}{ccc|c} 0 & 9 & -9 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 9 & -9 & 0 \end{array}\right]$$

$$E_1 = \left[\begin{array}{ccc|c} 0 & 9 & -9 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$E_1 = \left[\begin{array}{ccc|c} 0 & 1 & -1 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$E_1 = \left[\begin{array}{ccc|c} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The resulting equation for $E_1$ will be as following.

$$(0)(X_1) + (1)(X_2) - (1)(X_3) = 0$$

Thus, we'll have the following *eigenvector* and *eigenvalue.*

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 0 \\ X_2 \\ X_2 \end{bmatrix}$$

$$X = X_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \lambda = 4$$

Calculating the *eigenvector* using the same procedure, we'll get the following results.

$$E_2 = (A + 7I \mid 0) = \left[\begin{array}{ccc|c} 11 & 0 & 0 & 0 \\ 0 & 9 & 2 & 0 \\ 0 & 9 & 2 & 0 \end{array}\right] \tag{2.7}$$

Using *Row Operations* we'll have:

$$E_2 = \left[\begin{array}{ccc|c} 11 & 0 & 0 & 0 \\ 0 & 9 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Since $E_2$ is in the *Row Echelon Form*, we can infer the equation for $E_2$ as:

$$(11)(X_1) + (1)(X_2) = 0$$

For the *eigenvector* we'll have:

$$X = X_2 \begin{bmatrix} \frac{-1}{11} \\ \\ 1 \\ \\ 0 \end{bmatrix}, \quad \lambda = -7$$

---

(b) Since $A$ is a $2 * 2$ matrix, we have:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Given *eigenvectors* and *eigenvalues* in the problem description, we can conduct the *Characteristic Equation* presented in 2.2 and get the following results.

$$(A - 2I)X = \begin{bmatrix} a-2 & b \\ c & d-2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$(A - 5I)X = \begin{bmatrix} a-5 & b \\ c & d-5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Simply solving these equations will give us the following results.

$$\boxed{a=2} \quad \boxed{b=3} \quad \boxed{c=0} \quad \boxed{d=5}$$

Finally, we'll rewrite the matrix $A$ replacing the new parameters in it.

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}$$

# 3 Probability Density Function

Let

$$f(x,y) = \begin{cases} c(x+y^2) & if \;\; 0<x<1 \; ; \; 0<y<1 \\ 0 & elsewhere. \end{cases}$$

Find each of the following.

(a) $c$

(b) $f_X(x)$

(c) $f_{X|Y}(y)$

(d) Are $X$ and $Y$ independent?

(e) What is the probability $Pr(X < \frac{1}{2} | Y = \frac{1}{2})$?

## Solution

(a) The *Joint Probability Distribution* is given above. It is obvious that the total probability is always equal to 1; Hence we can integrate over the variables in the *JPDF* and that will give us the *Total Probability* which is equal to *1*.

$$\int_0^1 \int_0^1 f(x,y) \text{ d}x \text{ d}y = 1$$

Replacing the given equation we'll have the following.

$$\int_0^1 \int_0^1 c(x+y^2) \text{ d}x \text{ d}y = 1$$

$$\int_0^1 (\frac{c}{2} + cy^2) \text{ d}y = 1$$

$$\frac{c}{2} + \frac{c}{3} = 1$$

$$\boxed{c = \frac{6}{5}}$$

---

(b) *Marginal Density* can easily be computed by integrating over the variable of not interest.

$$f_X(x) = \int_0^1 cx + cy^2 \text{ d}y$$

$$f_X(x) = \int_0^1 \frac{6x}{5} + \frac{6y^2}{5} \, \mathrm{d}y$$

$$\boxed{f_X(x) = \frac{6x+2}{5}}$$

---

(c) While having the relation for the *Joint Probability* and the ease of finding *Marginal Density* of each of the variables, we can simply find the conditional probability using the *Bayes* theorem.

$$f_{X|Y}(x) = \frac{f(x,y)}{g(y)} = \frac{joint\ density\ of\ x\ \&\ y}{marginal\ density\ of\ y} \tag{3.1}$$

The $g(y)$ can be found by integrating the *JPDF* over the variable $X$.

$$g(y) = \int_0^1 \frac{6}{5}(x + y^2) \, \mathrm{d}x$$

$$g(y) = \frac{6y^2}{5} + \frac{3}{2}$$

Finally, the $f_{X|Y}(x)$ will be:

$$f_{X|Y}(x) = \frac{\frac{6}{5}(x+y^2)}{\frac{6y^2}{5} + \frac{3}{2}}$$

---

(d) In the case the $X$ and $Y$ are independent, the conditional probability $f_{X|Y}(x)$ is equal to $f_X(x)$. In the previous section, we calculated the $f_{X|Y}(x)$ using *Bayes* theorem. We can simply find $f_X(x)$ from the *JPDF* by integrating over $Y$. If the results are equal:

$$f_{X|Y}(x) = f_X(x)$$

we say that $X$ and $Y$ are independent. Computing the integral of *JPDF* over $Y$ we get:

$$f_X(x) = \frac{6x+2}{5}$$

So the variables $X$ and $Y$ are not *independent*.

---

(e) Given the conditional density introduced in 3.1, we'll have:

$$Pr(X < \frac{1}{2} \mid Y = \frac{1}{2}) = \frac{f(X < \frac{1}{2}, Y = \frac{1}{2})}{g(\frac{1}{2})}$$

$$Pr(X < \frac{1}{2} \mid Y = \frac{1}{2}) = \frac{\frac{3}{10}}{\frac{36}{20}}$$

$$Pr(X < \frac{1}{2} \mid Y = \frac{1}{2}) = \frac{1}{6}$$

# 4 Multivariate Gaussian Distribution

Recall the multivariate Gaussian for a vector $x \epsilon R^n$.

$$p(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{\frac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \tag{4.1}$$

(a) Let $y = v_i^T x$, where $v_i$ is an eigenvector of $\Sigma$ with an eigenvalue of $\lambda_i$. Find the probability density of $p(Y)$.

(b) Let $X$ be a zero mean Gaussian random vector with an isotropic covariance ($\Sigma = I$). Let $Y = AX + B$. Compute the mean and variance of $Y$.

(c) (*Matlab*) Generate 500 random samples from a 2 dimensional Gaussian with an isotropic $\Sigma$ using the *Matlab* command ***randn***. Transform the data as above with $B = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$, and $A = \begin{bmatrix} -5 & 5 \\ 1 & 1 \end{bmatrix}$. Plot the original and transformed points.

## Solution

(a) The represented map in section a is a *Linear Map*. Applying a *Linear Map* to a random vector $X$ with *mean* $\mu_X$ and *covariance matrix* $\Sigma_X$ will give us a new random vector $Y$ with the following properties.

$$\boxed{\Sigma_Y = v_i \Sigma v_i^T} \tag{4.2}$$

$$\boxed{\mu_Y = v_i \mu} \tag{4.3}$$

It is important to mention that the result of a *Linear Map* on a random vector $X$ with *Gaussian* distribution will result in variable Y with *Gaussian* distribution; Hence we can simply represent the new density function for the variable $Y$ using the *covariance matrix* and *mean* vector given 4.1, 4.2 and 4.3.

$$p(y \mid \mu_Y, \Sigma_Y) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_Y|^{\frac{1}{2}}} e^{\frac{-1}{2}(y-\mu_Y)^T \Sigma_Y^{-1}(y-\mu_Y)} \tag{4.4}$$

Thus we will have the following probability distribution for random vector $Y$.

$$p(y \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |v_i \Sigma v_i^T|^{\frac{1}{2}}} e^{\frac{-1}{2}(y-v_i\mu)^T (v_i \Sigma v_i^T)^{-1}(y-v_i\mu)}$$

---

(b) A formal representation of the problem description is given below.

$$X \sim \mathcal{N}(\mu_X,\ \Sigma_X) = \mathcal{N}(0,\ I)$$

Note that the isotropic covariance will look like $C = \lambda I$ which in this special case $\lambda = 1$. Applying the linear map $y = Ax + B$ on $X$ will have the following effect on the *mean* and *covariance matrix*.

$$Y \sim \mathcal{N}(A\mu_X + B,\ v_i \Sigma_X v_i^T)$$

$$Y \sim \mathcal{N}(B, I)$$

The new *mean* will be equal to $B$ and the new *covariance matrix* will be equal to $I$ because:

$$Cov(Y) = Cov(AX + B) = A\ \Sigma\ A^T = A\ I\ A^T = A\ A^T\ I = I$$

# Correlation & Expectation