Assignment 1
Warm-up for Basic Pattern Recognition Prerequisites

Please note:

1. What you must hand in includes the assignment report (.pdf) and – if necessary – source codes (.m). Please zip them all together into an archive file named according to the following template: HW1_XXXXXXXX.zip
   Where XXXXXXXX must be replaced with your student ID.
2. Some problems are required to be solved *by hand* (shown by the icon), and some need to be implemented (shown by the icon).
3. As for the first type of the problems, you are free to solve them on a paper and include the picture of it in your report. Here, cleanness and readability are of high importance.
4. Your work will be evaluated mostly by the quality of your report. Don't forget to explain what you have done, and provide enough discussions when it's needed.
5. 5 points of each homework belongs to compactness, expressiveness and neatness of your report and codes.
6. By default, we assume you implement your codes in MATLAB. If you're using Python, you have to use equivalent functions when it is asked to use specific MATLAB functions.
7. Your codes must be separated for each question, and for each part. For example, you have to create a separate .m file for part b. of question 3. Please name it like p3b.m.
8. Problems with bonus points are marked by the icon.
9. **Please upload your work in Moodle, before the end of October 20th.**
10. If there is ***any*** question, please don't hesitate to contact me through the following email address:
    - **ali.the.special@gmail.com**
11. Unfortunately, it is quite easy to detect copy-pasted or even structurally similar works, no matter being copied from another student or internet sources. Try to send us your own work, without being worried about the grade! ;)

1. Designing Simple Pattern Analysis Systems (16 Pts.)

Keywords: *Pattern Recognition System, Feature Extraction, Prediction Problems, Classification, Regression, Clustering*

A **Pattern Analysis System** is a system responsible for automated recognition of patterns in data, in order to identify which of a set of categories (or classes) a new observation belongs to, or to estimate the value of a specific attribute. It contains several parts, which must be carefully designed.

In this problem, you will get hands-on experience in designing a pattern analysis system for various different scenarios:

a. Managing students attendance in a classroom
b. Predicting US Dollar to Iranian Rial exchange rate in the coming year
c. Detecting suspicious behaviors in a prison
d. Labeling a collection of movies by their genres
e. Grouping football teams based on their playing styles

In each one of the scenarios, please answer the following questions:

1. Which types of prediction problems (classification, regression, etc.) does it belong to?
2. What sensors (if any) are needed?
3. What is your training set?
4. How do you gather your data?
5. Which features do you select?
6. Is there any pre-processing stage needed? Explain.
7. Express the challenges and difficulties that may affect the outcome of your system.
8. How beneficial do you think it is to design such a system? Express the pros and cons of applying these systems instead of using a human observer.

Note 1: There is no limitations on the methods you choose. As an example, a student attendance management system could be based on student's faces (face recognition), fingerprints (fingerprint recognition), gaits (gait recognition), etc.

Note 2: Your design must be as practical as possible, e.g. features must be discriminative and measurable.

Note 3: We know it's impossible to predict Dollar/Rial exchange rate, but try your best!

2. Getting More Familiar with the Art of Feature Extraction (12 Pts.)

Keywords: *Feature Extraction, Classification Problems, Transfer Learning*

The success of a pattern recognition system is heavily dependent on the **Feature Extraction** stage, where the goal is to extract distinctive properties of input patterns that best help in differentiating between the categories of the input data.

In this problem, you are going to get more familiar with the importance of feature extraction stage. Here, the focus is mainly on classification problems.

a. Figure 1 shows different types of cucumber. In each of the following parts, please state what features might be used to best distinguish between each sets of cucumbers.
   a1. {Persian} and {English}
   a2. {Lemon} and {Japanese}
   a3. {Persian} and {Middle Eastern}
   a4. {Kirby} and {Japanese}
   a5. {Kirby} and {Lemon}
   a6. {Kirby}, {English} and {Japanese}
   a7. {Kirby}, {Persian}, {English} and {Lemon}
   a8. All cases

Figure 1 Six different types of cucumber

b. Let's continue with a more complicated scenario. Figure 2 demonstrates different stages in banana ripeness process. Again, determine the most discriminative features in order to classify the following sets.
   b1. {1} and {3}
   b2. {3} and {4}
   b3. {3} and {7}
   b4. {1} and {2}
   b5. {1}, {2} and {3}
   b6. {3}, {4} and {5}
   b7. {4}, {5}, {6} and {7}
   b8. All cases

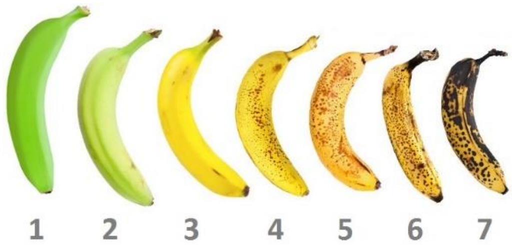

Figure 2 Banana ripeness chart including 7 different states, from raw (1) to ripe (7)

c. Now consider a more tricky case. In Figure 3, a set of US Dollar banknotes are displayed. Like the previous parts, you must state the features you think are best discriminative for classifying the following bills.
   c1. {$1 bill} and {$100 bill}
   c2. {$1 bill} and {$2 bill}
   c3. {$10 bill}, {$20 bill} and {$50 bill}
   c4. All cases

Figure 3 United States Dollar banknotes, from $1 bill to $100 bill

d. What features do you choose in order to build a classifier capable of recognising cucumbers and banknotes (of any types)?
e. Suppose you want to train a classifier capable of distinguishing bananas from banknotes, and then use it in a similar problem, where the goal is to distinguish cucumbers from banknotes (**Transfer Learning**). What would be your selected features?
f. Do you think it is possible to build a classifier which is able to distinguish the following sets?
   {$1 bill}, {$2 bill}, {$5 bill}, {Kirby cucumber} and {Persian cucumber}
   If yes, clarify how and using what features, and if no, explain why.

Note: Again, your features must be properly measurable.

3. Feature Selection: Evaluating Features to Select '*Good*' Ones (9 Pts.)

Keywords: *Feature Selection, Linear/Nonlinear Separable Data, Data Correlation, Multi-modality Data*

After **Feature Extraction** stage, it is advised to apply a process to select a subset of relevant features. The motivation behind this process, called **Feature Selection**, is that the data may contain redundant on irrelevant features, and can thus be reduced without incurring much loss of information. One can apply this process for many reasons, from simplifying the model and avoiding *the curse of dimensionality* to reducing training times and enhancing generalisation.

In this problem, you are going to practice simple feature selection tasks. You will find out more about **Feature Selection** as the course goes on.

Download the famous [Iris dataset]() and load it in MATLAB. As you can see, the dataset contains 3 classes, and each sample has 4 attributes.

a. Suppose you are going to classify the samples with only one of their attributes. Plot each feature distribution in a separate figure (4 figures in total). Then answer the following questions.
   a1. Which features are *good*, and which ones are *bad*? Explain your reasons.
   a2. Investigate each feature in terms of linear/non-linear separability.
   a3. Investigate each feature in terms of correlation among the samples.
   a4. Investigate each feature in terms of modality between the samples.
   Hint: In order to clearly highlight the differences, you can plot the features belonging to each class in different vertical levels, as can be seen in Figure 4.

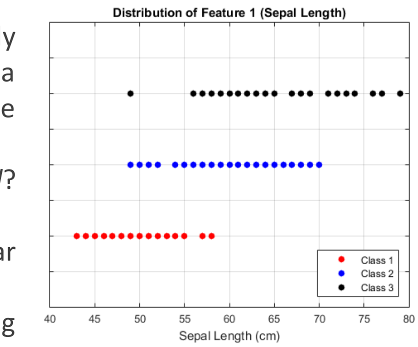

Figure 4 Distribution of the first feature

b. Now, suppose you want to keep two features for classification. Plot distributions for all feature pairs (6 figures in total), and answer the previous questions – a1 to a4 – for these feature pairs.
   Hint: The first one has been done for you (Figure 5).

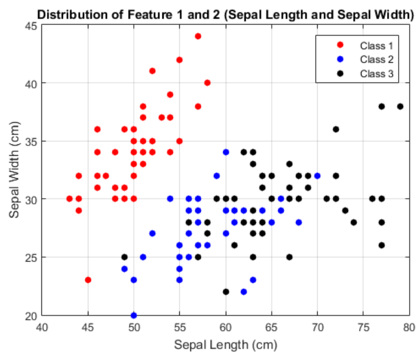

Figure 5 Distribution of the first two features

c. Repeat part a. for sets containing three features (4 figures in total), and answer the same questions for these feature trios.
   Hint: The first one has been done for you (Figure 6).

Hint: In MATLAB, you can easily load Iris dataset by using `load iris.dat`

Note 1: Remember to highlight samples of each class with different colors.

Note 2: Include the resultant figures of each part in your report.

Recommended MATLAB functions: `plot()`, `plot3()`

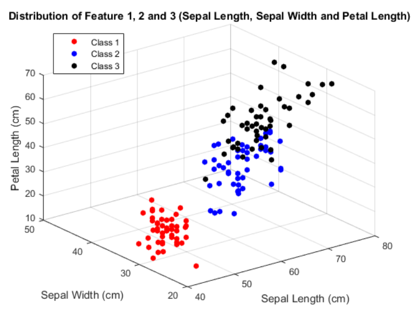

Figure 6 Distribution of the first three features

4. Basic Statistics Warm-up (Part I) (10 Pts.)

Keywords: *Probability Theory, Random Variable, Discrete Variable, Conditional Probability, Marginal Probability, Expected Value, Independent Variables, Correlated Variables*

In **Statistical Pattern Recognition**, the goal is to use **Statistical Techniques** for analysing data measurements in order to extract meaningful information and make justified decisions. Therefore, mastering basic statistical properties and to be able to understand and use them is highly important.

In this problem, you are to review your knowledge in this area.

a. A biased coin is so that the probability of obtaining tail is $2/3$. Suppose that X is the number of obtained heads. Find the following quantities after the coin is tossed 1400 times.
   a1. The mean of X
   a2. The standard deviation of X
b. When Lionel Messi takes a free-kick, 4 times out of 10 he will score. In El Clásico, he takes 6 free-kicks.
   b1. Find the probability that he scores a hat-trick (3 goals) from free-kick spot.
   b2. Find the probability that he scores for the first time in his third try.
c. A random variable X is distributed according to a Poisson distribution. If $\mathrm{P}(X=3)=\mathrm{P}(X=0)+\mathrm{P}(X=1)$;
   c1. Find the value of mean
   c2. For the value you calculated for mean, evaluate $P(2 \le X \le 4)$
d. Assuming X and Y are discrete random variables and a and b are constant, show that:
   d1. $\mathrm{E}[aX+bY]=a\mathrm{E}[X]+b\mathrm{E}[Y]$
   d2. $\sigma_X^2=\mathrm{E}[X^2]-\mu_X^2$
   d3. Independence implies uncorrelatedness
   d4. uncorrelatedness doesn't necessarily implies independence

5. Basic Statistics Warm-up (Part II) (10 Pts.)

Keywords: *Probability Theory, Random Variable, Probability Distribution, Density Function, Continuous Variable, Cumulative Distribution Function, Expected Value*

Following the previous problem, here you are going to continue reviewing statistical principals in a different manner.

Suppose a continuous random variable X with a probability density function $f(x)$, where

$$f_k(x)=\begin{cases} e-ke^{kx}, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$

a. Find a proper value for k which makes $f_k(x)$ a valid density function.
b. Determine $P(0.25 \le \mathrm{X} \le 0.5)$ in terms of e.
c. Find the cumulative distribution function (CDF) of X.
d. Calculate the mean and variance of the distribution in terms of e.

Suppose that X represents the lifetime of a certain type of battery (in years).

e. What is the probability that a battery lifetime lasts more than 9 months.

Three of these batteries are fitted in a TV remote control. Assuming that each battery fails independently of the other two, find the probability that after six months;

f. None of the three batteries has failed

g. Exactly one of the three batteries has failed

6. Hanging Around with Covariance Matrix and Linear Transformations (8 Pts.)

Keywords: *Covariance, Covariance Matrix, Data Dimension, Data Correlation, Eigenvalues, Eigenvectors, Linear Transformations, Whitening Transformation*

In statistical pattern recognition, **Covariance Matrix** concept is highly important. In general, it is defined as a matrix whose element in position i, j is the covariance between $i-th$ and $j-th$ elements of a random vector. It somehow generalises the concept of variance to multiple dimension.

In this problem, you are going to examine your knowledge of covariance matrices and their attributes.

Consider a dataset with covariance matrix $\Sigma = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 4 & 0 \\ -2 & 0 & -1 \end{bmatrix}$, and answer the following questions.

a. Specify the dimensionality of the dataset, i.e. the number of features each sample has.

b. Determine the number of samples in the dataset.

c. Find the correlations between different data dimensions.

d. On which dimension are the data scattered more?

e. Calculate eigenvalues and eigenvectors associated with the covariance matrix, and then find the angle between each of the eigenvector pairs. What can you infer from the three obtained values? Does it hold in every arbitrary covariance matrix? Justify your answer.

f. Find a transformation to whiten data associated with the given covariance matrix.

g. Show that Σ is a valid covariance matrix.

7. Mastering Eigenvalues and Eigenvectors and Their Properties (10 Pts.)

Keywords: *Eigenvalues, Eigenvectors, Eigenspace, Invertible Matrix, Diagonalizable Matrix*

Eigenvalues and **Eigenvectors** are vastly used in various scientific areas, from geology and ecology to computer vision and data mining. They also play an important role in the field of pattern recognition, where it is applied in many applications such as calculating covariance matrix or principle component analysis (PCA).

In this problem, we are going to take a deeper look into these concepts.

a. For each of the following pairs, determine whether v is an eigenvector of $\mathbf{A}$ or not. If so, specify the corresponding eigenvalue.

a1. $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}, v = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$

a2. $\mathbf{A} = \begin{bmatrix} 5 & 6 & 2 \\ 0 & -1 & -8 \\ 1 & 0 & -2 \end{bmatrix}, v = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

b. Determine a basis for the eigenspace corresponding to each of the following eigenvalues.

b1. $\mathbf{A}=\begin{bmatrix}1 & -6\\ -3 & 4\end{bmatrix}, \lambda=-2$ b2. $\mathbf{A}=\begin{bmatrix}-2 & 4 & 2\\ 2 & 1 & -2\\ 4 & -2 & 5\end{bmatrix}, \lambda=6$

c. Calculate the eigenvalues and eigenvectors associated with the following matrices.

c1. $\mathbf{A}=\begin{bmatrix}3 & 2\\ -1 & 6\end{bmatrix}$ c2. $\mathbf{A}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$ c3. $\mathbf{A}=\begin{bmatrix}3 & 1 & -2\\ 2 & 3 & -2\\ 2 & 1 & -1\end{bmatrix}$

d. Find a 2×2 matrix $\mathbf{A}$ with eigenvalues $\lambda_1=-1$ and $\lambda_2=7$, and corresponding eigenvectors $v_1=\begin{bmatrix}-\frac{3}{2}\\ 1\end{bmatrix}$ and $v_2=\begin{bmatrix}\frac{1}{2}\\ 1\end{bmatrix}$.

e. Let $\mathbf{A}$ be an invertible matrix with eigenvalue λ. Assuming a nonzero x satisfies $\mathbf{A}x=\lambda x$ show that $\mathbf{A}^{-1}$ has an eigenvalue equal to λ^{-1}.

f. Assume $\mathbf{A}$ is a matrix of the size 3×3 with two eigenvalues, and the corresponding eigenspaces are one-dimensional. Justify whether $\mathbf{A}$ is diagonizable or not.

g. Show that if a $n\times n$ matrix $\mathbf{A}$ has an eigenvalue equals to 3, then $\mathbf{A}^2$ has an eigenvalue equals to 9.

h. Consider the sequences x_n and y_n, such that for each $n\geq 1$,

$$x_n=2x_{n-1}-3y_{n-1}, \quad y_n=-4x_{n-1}+y_{n-1}$$

Assuming $x_0=2$ and $y_0=3$, specify each of x_n and y_n explicitly in terms of n.

Useful Links: [1]

8. Simple Sample Generation and Beyond (15 Pts.)

Keywords: *Sample Generation, Normal Distribution, Linear Transformations, Simultaneous Diagonalisation, Whitening Transformation*

In many pattern recognition applications, **Sample Generation** plays an important role, where it is necessary to generate samples which are to be normally distributed according to a given expected vector and a covariance matrix.

In this problem, you are going to do this technique yourself. You will also practice some more complicated matrix operations as well.

a. Generate samples from three normal distributions specified by the following parameters:

$$n=1, \quad N=1000, \quad \mu=2, \quad \sigma=1,2,3$$

Plot the samples, as well as the histograms associated with each of the distributions. Compare the results.

b. Generate samples from a normal distributions specified by the following parameters:

$$n=2, \quad N=1000, \quad M=\begin{bmatrix}1\\ 2\end{bmatrix}, \quad \Sigma=\begin{bmatrix}3 & 1\\ 1 & 2\end{bmatrix}$$

Display the samples, as well as the associated contour plot.

c. Consider a normal distribution specified by the following parameters:

$$n = 2, \quad N = 1000, \quad M = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Determine appropriate values for each of the unknown variables, so that the shape of the distribution becomes:

c1. A circle in the upper left of the Euclidean coordinate system.

c2. A diagonal line (/ shape) in the centre

c3. A horizontal ellipsoid in the lower right of the Euclidean coordinate system

Display the generated samples.

d. Calculate the sample mean $\hat{M}$, and sample covariance matrix $\hat{\Sigma}$ of the distribution in part b., and comment on the results.

e. Simultaneously diagonalise Σ and $\hat{\Sigma}$, and form a vector $V = [\lambda_1, \lambda_2]^T$.

f. Find a transformation for covariance matrix of the distribution in part b., such that when applied on the data, the covariance matrix of the transformed data becomes $\mathbf{I}$. Transform the data and display the distribution in the new space.

g. Calculate the eigenvalues and eigenvectors associated with the covariance matrix of the distribution in part b. Plot the eigenvectors. What can you infer from them?

h. Again, consider the distribution and samples you generated in part b. Construct a 2×2 matrix $\mathbf{P}$, which has eigenvectors associated with Σ as its columns ($\mathbf{P} = [v_1, v_2]$, such that v_1 is corresponding to the largest eigenvalue). Project your generated samples to a new space using $\mathbf{Y}_i = (\mathbf{X}_i - M) \times \mathbf{P}$, and plot the samples. What differences do you notice?

i. Find the covariance matrix associated with the projected samples in part h. Also calculate its eigenvalues and eigenvectors, and comment on the results.

Recommended MATLAB functions: `meshgrid(), mvnpdf(),mvnrnd(),eig()`

9. Some Explanatory Questions (5 Pts.)

Please answer the following questions as clear as possible:

a. Why do you think Central Limit Theorem is important? Where and how can it be used?

b. What is the difference between a feature and a measurement?

c. Does a covariance matrix need to be symmetric? Why?

d. What does zero eigenvalue mean?

e. When does the whitening transformation come into use?

f. Explain how does Google's PageRank algorithm use eigenvalues and eigenvectors concepts.

Good Luck!
Ali Abbasi