Assignment 1
Warm-up of Basic Pattern Recognition Concepts

Homeworks Guidelines and Policies

- ***What you must hand in.*** It is expected that the students submit an assignment report (HW1_[student_id].pdf) as well as required source codes (.m or .py) into an archive file (HW1_[student_id].zip).
- ***Pay attention to problem types.*** Some problems are required to be solved *by hand* (shown by the icon), and some need to be implemented (shown by the icon).
  Please don't use implementation tools when it is asked to solve the problem by hand, otherwise you'll be penalized and lose some points.
- ***Don't bother typing!*** You are free to solve by-hand problems on a paper and include picture of them in your report. Here, cleanness and readability are of high importance. Images should also have appropriate quality.
- ***Reports are critical.*** 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.
- ***Appearance matters!*** In each homework, 5 points (out of a possible 100) belongs to compactness, expressiveness and neatness of your report and codes.
- ***Python is also allowable.*** 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.
- ***Be neat and tidy!*** 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.
- ***Use bonus points to improve your score.*** Problems with bonus points are marked by the icon. These problems usually include uncovered related topics or those that are only mentioned briefly in the class.
- ***Moodle access is essential.*** Make sure you have access to Moodle because that's where all assignments as well as course announcements are posted on. Homework submissions are also done through Moodle.

- ***Assignment Deadline.*** Please submit your work **before the end of December 4th**.
- ***Delay policy.*** During the semester, students are given 7 free late days which they can use them in their own ways. Afterwards there will be a 25% penalty for every late day, and no more than three late days will be accepted.
- ***Collaboration policy.*** We encourage students to work together, share their findings, and utilize all the resources available. However you are not allowed to share codes/answers or use works from the past semesters. Violators will receive a zero for that particular problem.
- ***Any questions?*** If there is any question, please don't hesitate to contact us through the following email address: **ali.the.special@gmail.com** and **fardin.aiar@gmail.com**.

1. Improving the Height Requirement Systems in Disneyland Theme Parks (12 Pts.)

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

In Walt Disney World and Disneyland theme parks, there are strict restrictions on the availability of some rides to the children. Most of these restrictions are based on height requirements, which demand the visitors to be taller than a certain threshold in order to be able to safely enjoy the ride. However, there are some items which are designed for children of specific age groups, in which guests are also evaluated by their heights. Although the correlation between age and height in young children has been established, yet it is known to be erroneous and inaccurate, which may cause dissatisfaction and frustration among parents.

Figure 1 Some items at Walt Disney World and Disneyland theme parks have minimum and a few have maximum height requirements.

Now, you are asked to suggest a practical and feasible **Pattern Analysis System** capable of detecting children between 7 and 14. Please specify the following:

a. Which types of prediction problems (classification, regression, etc.) does it belong to?
b. What sensors (if any) are needed?
c. What is your training set?
d. How do you gather your data?
e. Which features do you select?
f. Is there any pre-processing stage needed? Explain.
g. Express the challenges and difficulties that may affect the outcome of your system.
h. How beneficial do you think it is to design such a system? Express the pros and cons of applying these systems instead of using ordinary height measurement method.

Note 1: There is no limitations on the method you choose. As an example, a system capable of grouping students by their personalities could be based on surveillance cameras (**Computer Vision** techniques), paper-based surveys (**Sentiment Analysis** techniques), etc.

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

2. Feature Selection in Some Challenging Computer Vision Problems (15 Pts.)

Keywords: *Feature Selection, Classification Problems, Fingerprint Recognition, Emotion Recognition, Gait Recognition, Activity Recognition, Object Recognition*

One of the core concepts in pattern recognition which hugely impacts the performance of the selected model is **Feature Selection**. Selecting the appropriate set of features used for data modelling has been shown to improve the performance of any **Supervised** and **Unsupervised** learning method. Feature selection is greatly responsible for reducing computational cost (such as training time or required sources), model interpretability and avoiding high-dimensional input data.

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

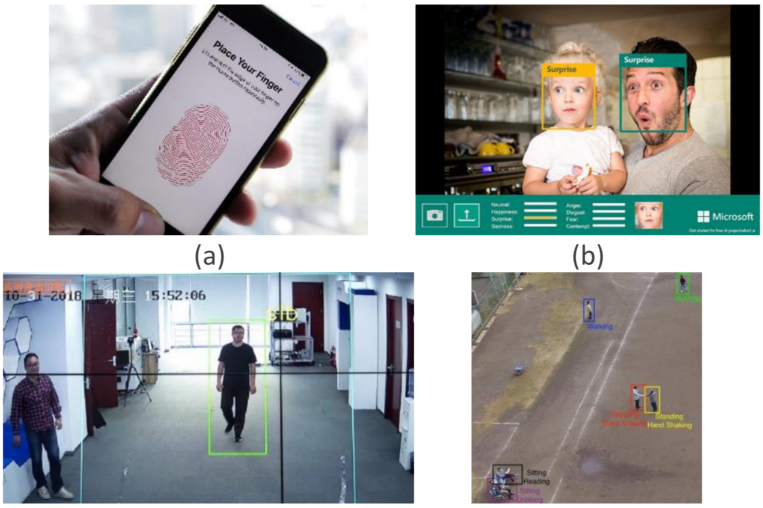
First, assume the following problems. Suggest at least two discriminative and measurable features that might be used for each problem.

a. **Fingerprint Recognition**: Aims to identify or confirm the identity of a person based on the comparison of two fingerprints. It is considered as the most popular biometric solution used for authentication on computer systems.

b. **Emotion Recognition**: Application of image processing and computer vision algorithms to locate faces in the image, extract facial features and classify them into some facial expression-interpretative categories such as “happiness” or “surprised”.

c. **Gait Recognition**: The process of determining person’s identity by the manner of his walking. It has been shown that the way a person walks – called “gait” – is a biometric identifier, i.e., a unique biological or behavioral identification characteristic such as fingerprint or face.

d. **Activity Recognition**: Deals with recognising actions and behaviors of one or more subjects from a series of observations from sensor data. Actions are often involves typical activities performed indoors, such as sitting, standing, walking and talking.
Note: Here, consider a human activity recognition system which is able to recognise the following activities: “standing”, “walking”, “running”, “reading”, “handshaking”, “drinking” (See part d of Figure 2).



(a) (b)

(c) (d)

Figure 2 Four different computer vision applications (a) iPhone 8 fingerprint recognition featre, known as “Touch ID” (b) Microsoft tool for real-time emotion recognition (c) A surveillance camera capable of distinguishing people by their gaits (d) A surveillance camera in a prison yard, equipped with an activity recognition algorithm

Now let’s investigate simple feature selection in another computer vision task. Consider an **Object Recognition** problem which tries to find simple geometric shapes in an image, Figure 3. In each part, suggest feature(s) that could be used to distinguish between the desired sets. Note that once again your features must be properly measurable.

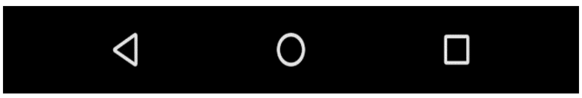


Figure 3 Android home screen soft keys. Here, the goal is to introduce a simple object recognition system which is capable of locating these three geometric shapes in an input image

e. {Circles} and {Squares}
f. {Triangles} and {Squares}
g. {Triangles} of any sizes (robust against *size*)
h. {Circles} of any sizes (robust against *size*)
i. {Squares} of any rotational degrees (robust against *rotations*)
j. {Triangles}, {Circles} and {Squares}

3. Introducing Subspace Learning as a Feature Extraction Method (12 Pts.)

Keywords: *Feature Extraction, Subspace Learning, Feature Projection, Linear Transformation, Dimensionality, Dimensionality Reduction, Curse of Dimensionality*

Feature Extraction and **Feature Selection** are two very close concepts. Feature selection often involves filtering irrelevant or redundant features. Therefore, the result would be a subset of the original features. Feature extraction, on the other hand, creates brand new set of features that still captures most of the useful information. The process can be supervised (e.g., **PCA**) or supervised (e.g., **LDA**).

In pattern recognition, it is usually more desirable to keep the **Dimensionality** low. Lower dimensionality often means simpler models and more generalisation, while less training times and risk of **The Curse of Dimensionality**. As dimensionality can be referred to the number of features in dataset, feature extraction methods and **Subspace Learning** methods – which are approaches to **Dimensionality Reduction** – are heavily related.

In this problem, you are going to get familiar with feature extraction by transferring features to new subspaces. You will find out more about feature extraction and dimensionality reduction methods as the course goes on.

Download the well-known [Iris dataset]() and load it in MATLAB or Python. As can be seen, the dataset contains 3 categories and each samples has 4 features. Suppose we want to keep two features for classification purpose.

a. Plot the distributions for the following feature pairs: feature 1 and 2, feature 1 and 3 and feature 2 and 4 (three figures in total).
b. Transform data in each of the figures using the following linear transformations, and plot the transformed data in new subspaces (12 figure in total):

$$T_1 = \begin{bmatrix} 1.2 \\ -0.3 \end{bmatrix} \qquad T_2 = \begin{bmatrix} -1.8 \\ 0.6 \end{bmatrix} \qquad T_3 = \begin{bmatrix} 1.4 \\ 0.5 \end{bmatrix} \qquad T_4 = \begin{bmatrix} -0.5 \\ -1 \end{bmatrix}$$

c. For each of the distributions in part a, investigate which one of the transformations in part b are more appropriate for classification purpose.

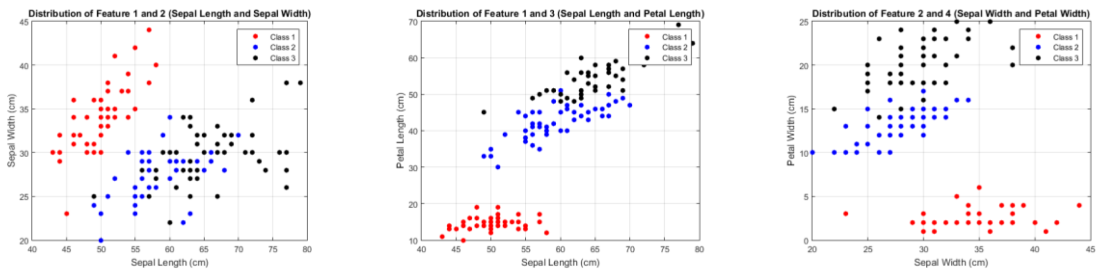


Figure 4 Distributions used in the first part

Now let's consider a different approach.

d. Plot the distributions for the remaining feature pairs, i.e., feature 1 and 4, feature 2 and 3 and feature 3 and 4 (three figures in total).

e. Use try and error to approximately find the best linear transformation for each of the distributions for classification purpose. Display the projected data in the new subspace (three figures in total).

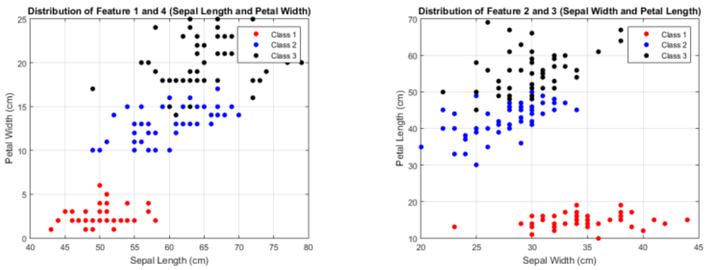


Figure 5 Distributions used in the second part

Recommended MATLAB functions: `plot()`

4. Basic Statistics Warm-up (12 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. The Fate of Khosrow and Shirin Love Story (20 Pts.)

Keywords: *Dynamic Systems, Differential Equations, State Transition Matrix, Steady States, Eigenvalues, Eigenvectors*

Predicting the fate of a love affair using the concept of eigenvalues and eigenvectors may sound awkward at the first glance, yet it is perfectly possible. Let's see how.

Khosrow II was a Sasanian king who had a love affair with Armenian princess, Shirin, the later queen of Persia. Their romantic story was influenced by the arrival of Farhad, a sculpture who also fell in love with Shirin and became Khosrow's love-rival. We now want to see how their romance continued based on the idea of the paper written by Steven H. Strogatz in 1988.

Figure 6 Khosrow seeing Shirin bathing. Written by Iranian poet Nizami Ganjavi, the tragic romance of Khosrow and Shirin is a fictional version of the story of the love of a Persian king for an Armenian princess.

Let $K[n]$ be Khosrow's emotions about Shirin on day n, and $S[n]$ be Shirin's feelings about Khosrow on day n. The two parameters are scalars, and their signs denote like or dislike. Thus, $K[n] > n$ means Khosrow likes Shirin, $S[n]$ indicates Shirin dislikes Khosrow, and $K[n] = 0$ denotes Khosrow has a neutral stance towards Shirin. The intensity of that feeling is also represented by the magnitude of $K[n]$ and $S[n]$, i.e., their absolute values.

First, we consider the following linear system for their relationship:

$$K[n+1] = aK[n] + bS[n], \quad n = 0,1,2,\dots$$
$$S[n+1] = cK[n] + dS[n], \quad n = 0,1,2,\dots$$

which can be written as $\vec{d}[n+1] = \mathbf{A}\vec{d}[n]$, where $\vec{d}[n] = \begin{bmatrix} K[n] \\ S[n] \end{bmatrix}$ represents the state vector and $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ indicates the state transition matrix. The fate of Khosrow and Shirin's love story depends on the model parameters a, b, c, and d, in the state transition matrix and the initial state $\vec{d}[0]$. Our goal in this problem is to explore some of these possibilities.

First, consider the case where $a + b = c + d$.

a. Show that $\vec{v}_1 = [1 \quad 1]^T$ is an eigenvector of $\mathbf{A}$, and specify its corresponding eigenvalue λ_1.

b. Show that $\vec{v}_2 = [b \quad -c]^T$ is an eigenvector of $\mathbf{A}$, and find its corresponding eigenvalue λ_2.

c. Write down the first and second eigenvalues and their eigenspaces in terms of the parameters a, b, c, and d.

Now, consider $\mathbf{A} = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$ as the state-transition matrix.

d. Calculate the eigenvalues and their corresponding eigenvectors for this system.

e. Find all of the non-zero steady states of this system. In other words, determine all possible state vectors $\vec{d}_*$ such that if Khosrow and Shirin start at, or enter, any of those state vectors, their states will stay in place forever: $\left\{\vec{d}_* \mid \mathbf{A}\vec{d}_* = \vec{d}_*\right\}$.

f. If Khosrow and Shirin start from an initial state $\vec{d}[0] \in \text{span}\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$, what happens to their love affair over time? Specifically, what is $\vec{d}[n]$ as $n \to \infty$?

g. Repeat the previous part for the case when the initial state is $\vec{d}[0] = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

Next, suppose the following state-transition matrix: $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

h. Calculate the eigenvalues and their corresponding eigenvectors for this system.

i. If Khosrow and Shirin start from an initial state $\vec{d}[0] \in \text{span}\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$, what happens to their love affair over time? Specifically, what is $\vec{d}[n]$ as $n \to \infty$?

j. Repeat the previous part for the case when the initial state is $\vec{d}[0] = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Finally, let's consider $\mathbf{A} = \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}$ as the state-transition matrix.

k. Calculate the eigenvalues and their corresponding eigenvectors for this system.

l. If Khosrow and Shirin start from an initial state $\vec{d}[0] = \begin{bmatrix} K[0] \\ S[0] \end{bmatrix}$, where $\vec{d}[0] \in \text{span}\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\} = \text{span}\left\{\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$, what happens to their love affair over time if initially Khosrow likes Shirin but Shirin dislikes Khosrow? What about if initially Khosrow dislikes Shirin but Shirin likes Khosrow? Specifically, what is $\vec{d}[n]$ as $n \to \infty$?

m. Repeat the previous part for the case when the initial state is $\vec{d}[0] \in \text{span}\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$.

Note: You may not be familiar with some of the terms used in this problem. Hence, extra readings may be required.

6. Application of Simple Statistical Calculations in a Real-World Problem (16 Pts.)

Keywords: *Probability Theory, Random Variable, Expected Value, Multivariate Gaussian Distribution, Heat-map*

Following the previous problems on statistics and probability, here we are going to deal with a more realistic and interesting problem.

Load the file “first_half_logs.csv” attached to this homework. This dataset contains activity logs of 11 players of a football team in the first half of a football match. The coordinate system is illustrated in Figure 7. As can be seen, $0 \leq x \leq 105$ and $0 \leq y \leq 68$.

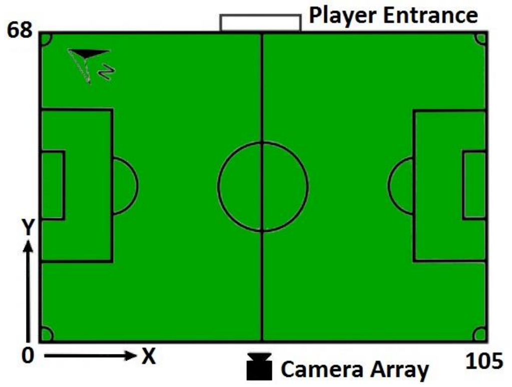


Figure 7 Coordinate system and other assumptions in the “first half logs” dataset

The data format as well as a part of the dataset is represented in Figure 8. Note that the `timestamp` (`string`) is Local Central European Time (CET), `tag_id` (`int`) is player’s identifier, and `x_pos` (`float`) and `y_pos` (`float`) are relative positions (in meters) of the player in the field’s x and y direction, respectively.

```
‘timestamp’,’tag_id’,’x_pos’,’y_pos’,’heading’,’direction’,’energy’,’speed’,’speed’,’total_distance’
’2013-11-03 18:30:00.000612’,31278,34.2361,49.366,2.2578,1.94857,3672.22,1.60798,3719.61
’2013-11-03 18:30:00.004524’,31890,45.386,49.8209,0.980335,1.26641,5614.29,2.80983,4190.53
’2013-11-03 18:30:00.013407’,0918,74.5904,71.048,-0.961152,0,2.37406,0,0.285215
’2013-11-03 18:30:00.015759’,109,60.2843,57.3384,2.19912,1.22228,4584.61,8.14452,4565.93
’2013-11-03 18:30:00.023466’,909,45.0113,54.7307,2.23514,2.27993,4170.35,1.76589,4070.6
```

Figure 8 A part of the given dataset as well as the associated header

a. Calculate and plot the mean of each of the 11 player’s locations in the first half.
b. For each of the players, find a Gaussian distribution for their locations. Plot the results for as well as a heat-map or 3D figure. Also report the mean and covariance of the distributions.
c. Randomly select three players, and for each player randomly select three different locations on the football field. For each location, report the probability of the player being at that locations.
d. Try to determine player’s positions in the football field (e.g., goalkeeper, central midfielder, right forward, etc.).
   Hint: See here.

7. Some Explanatory Questions (8 Pts.)

Please answer the following questions as clear as possible:

a. When does the whitening transformation come into use?
b. Regression is said to be a generalisation of classification task. How can you explain that?
c. What is the definition of an eigenvalue and an eigenvector for a matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$
d. Specify the geometric interpretation of the λ for matrix $\mathbf{M}$.
e. The concept of eigenvectors plays a major role in Google’s search engine algorithm. Briefly explain how.

Good Luck!
Ali Abbasi, Fardin Ayar