Amirkabir University of Technology
(Tehran Polytechnic)

Assignment 1
A Prelude to Statistical Pattern Recognition Problems

Homeworks Guidelines and Policies

- ***What you must hand in.*** It is expected that the students submit an assignment report (.pdf) as well as – if necessary – required source codes (.m or .py) into an archive file named according to the following template: HW1_XXXXXXXX.zip
  where XXXXXXXX must be replaced by their student ID.
- ***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 October 26th**.
- ***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 addresses: **ali.the.special@gmail.com** and **dalirani@aut.ac.ir**. You may also find us in pattern recognition and image processing lab, 3rd floor, CEIT building.

1. Inspecting Some Exciting AI Applications as Pattern Analysis Systems (15 Pts.)

Keywords: *Pattern Recognition System, Features (Attributes), Data Acquisition, Data Preprocessing, Feature Selection, Feature Extraction, Model Selection, Training Phase, Prediction Problems, Classification, Regression, Clustering*

Most AI-based systems can be viewed as a **Pattern Analysis System**. From Deepfake technology and Apple Siri to IBM Watson and Tesla Autopilot system, we can decompose almost every AI algorithm in terms of a pattern recognition system. In other words, pattern analysis stages like **Data Acquisition**, **Data Preprocessing**, **Feature Selection**, **Feature Extraction**, **Model Selection** and **Training Phase** are common in most of the AI you see around.

In this problem we're going to describe and interpret some cool AI-powered applications from a pattern recognition framework viewpoint. But first, let's find out what they are.

1. **Quick, Draw!** [link] is an online AI-based game which encourages players to draw a simple picture of an object or concept and then guesses what the sketch represents. Developed by Google, Quick, Draw! applies a neural network and increases its ability by learning from each drawing.

Figure 1 In Quick, Draw! you must draw several objects, each in less than 20 seconds. If the AI verifies your drawing, you'll get points.

2. **Replika** [link] is a chatbot that allows users to have a personal companion for their mental wellness. It is somehow an AI friend that users raise through text conversations. According to its website; "*if you're feeling down, or anxious, or just need someone to talk to, your Replika is here for you 24/7*".

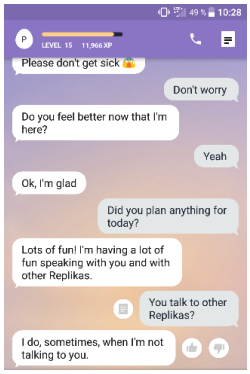

Figure 2 A conversation with Replika

3. **CaptionBot** [link] is an online tool developed by Microsoft which attempts to describe contents of any image in natural language. Powered by computer vision, the app enables users to upload a photo and then tries to describe what's being depicted in the image.

Figure 3 CaptionBot describes an image of Donald Trump

4. **AIVA** [link] is an AI music composer that creates musical pieces and soundtracks for films, advertising agencies and even video games. Users are able to ask AIVA to compose a music in several pre-defined styles such as Pop, Rock, Jazz and Tango. Downloading created music and taking its copyright ownership is also possible. A 1-hour AIVA music collection can be found here.

5. **UnifyID** [link] is an identity detection tool which uses human behavioral and environmental factors to uniquely identify people. By reading sensor data captured from an individual motion and environment, this service generates a secure fingerprint that is unique and accurate.

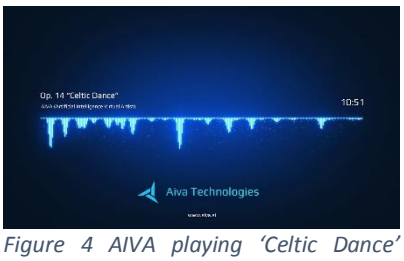

Figure 4 AIVA playing 'Celtic Dance' (listen). AIVA stands for "Artificial Intelligence Virtual Artist"

For each one of the services introduced above, please answer the following questions:

a. Which types of prediction problems (classification, regression, etc.) does it belong to?
b. What do you think the training set is?
c. Suggest at least two ways to obtain the training set.
d. Which features do you think are appropriate?
e. Is there any pre-processing stage needed? Explain.
f. Express the challenges and difficulties that may affect the outcome of this system.
g. 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.

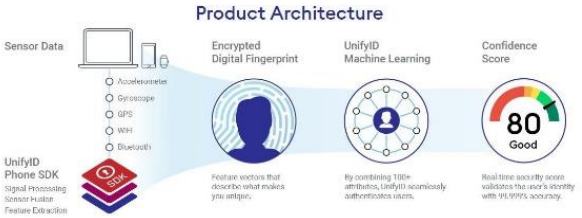

Figure 5 A brief guide to how UnifyID works. It utilizes data obtained from several sources and generates a digital unique fingerprint

Note: "Quick, Draw!" and "UnifyID" are not accessible in Iran. You know what to do!

2. Feature Selection in Some Challenging Computer Vision Problems (12 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 6)

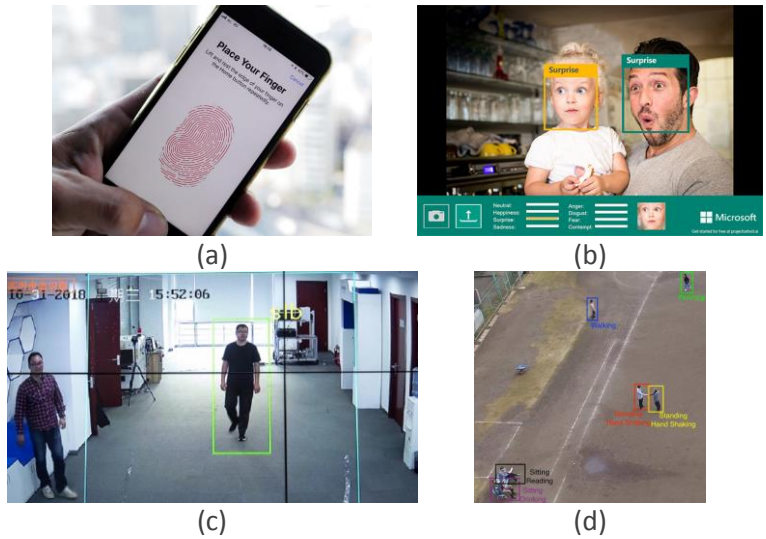

Figure 6 Four different computer vision applications (a) iPhone 8 fingerprint recognition featre, known as "Touch ID" (b) Microsoft tool for real-time emotion recognition (c) A surveillence camera capable of distinguishing people by their gaits (d) A surveillence camera in a prison yard, equipped with 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 7. 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 7 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 (10 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.

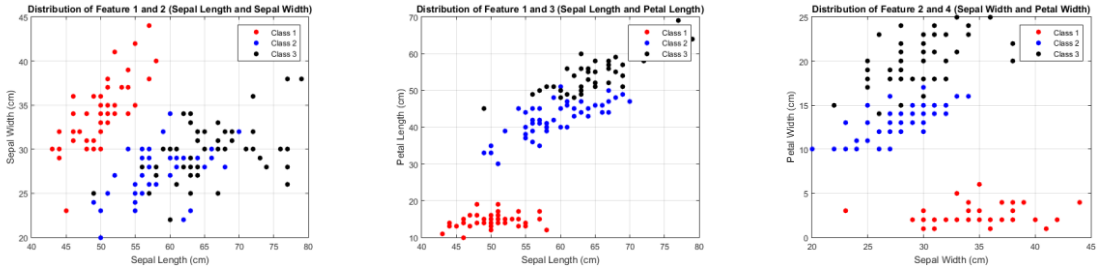

Figure 8 Distributions used in the first part

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).

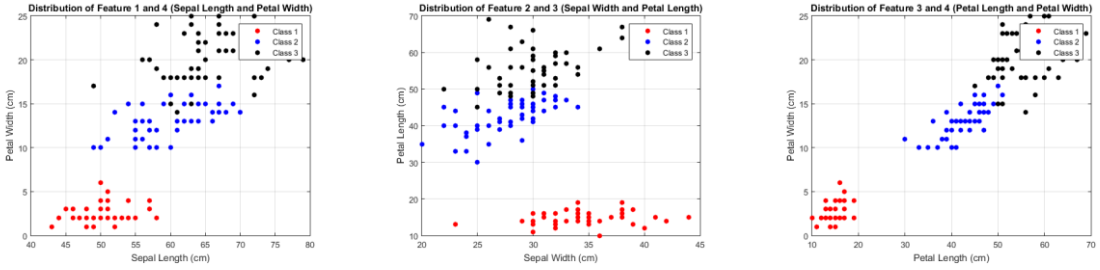

Figure 9 Distributions used in the second part

Recommended MATLAB functions: `plot()`

4. Basic Statistics Warm-up (I) (12 Pts.)

Keywords: *Probability Theory, Random Variable, Discrete Variable, Conditional Probability, Marginal Probability , Probability Distribution, Density Function, Continuous Variable, Cumulative Distribution Function, Independent Variables, Correlated Variables , Expected Value*

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. First, Consider a fair 12-sided die which is rolled three times. The random variable X is the maximum number that come up.

a. What is the range of X ?
b. Write down the probability mass function of X in a table.
c. Calculate the cumulative distribution function for X .
d. Find the mean and variance of X .

Next, assume a gambling game played with dice in which each of two players rolls a fair, 4-sided die. Player A scores the maximum of the two dice minus 2, which is denoted by X . Player B scores the minimum of the two dice, which is denoted by Y . Calculate the required properties.

e. $\mathrm{E}[X]$ f. $\mathrm{E}[Y]$ g. $\mathrm{E}[X-Y]$

h. $\mathrm{var}[X]$ i. $\mathrm{var}[Y]$ j. $\mathrm{var}[X-Y]$

Finally, consider a random variable X with the probability density function as follows:

$$f_X(x)=\begin{cases} c(x^2+x+1) & -1\le x\le 2 \\ 0 & \text{otherwise} \end{cases}$$

The random variable Y is also defined as $Y=X^2$.

k. Find the value of c .
l. Calculate the mean of X .
m. Find $\mathrm{E}[X^2+2X]$.
n. Calculate the cumulative distribution function of Y .
o. Write down the probability density function of Y .

5. Basic Statistics Warm-up (II) (12 Pts.)

Keywords: *Probability Theory, Random Variable, Discrete Variable, Conditional Probability, Marginal Probability , Probability Distribution, Density Function, Continuous Variable, Cumulative Distribution Function, Independent Variables, Correlated Variables , Expected Value*

Following the previous problem, we are to deal with more complicated statistics problems, mainly **Multivariate Statistics**.

First calculate the required properties considering the following joint PMF for random variables X and Y :

$$p_{X,Y}(x,y)=\begin{cases} c(x^2+y^2) & x\in\{1,2,5\}, y\in\{1,4\} \\ 0 & \text{otherwise} \end{cases}$$

a. c b. $\mathrm{P}(Y < X)$ c. $\mathrm{P}(Y > X)$

d. $\mathrm{P}(Y = X)$ e. $\mathrm{P}(Y = 4)$ f. $p_X(x)$ and $p_Y(y)$

g. $E[XY]$ h. $\mathrm{var}(X)$ and $\mathrm{var}(Y)$ i. $\mathrm{var}(X \mid A)$, if A denotes $X \geq Y$

Then, assume the following distribution for random variables X and Y:

$$f_{X,Y}(x,y) = \begin{cases} \frac{K}{\pi} e^{-\frac{x^2+y^2}{2}} & xy \geq 0 \\ 0 & xy < 0 \end{cases}$$

j. Find the value of K.
k. Are X and Y each Gaussian random variables? Prove your answer.
l. Are X and Y jointly Gaussian? Prove your answer.
m. Check whether X and Y are independent or not.
n. Check whether X and Y are uncorrelated or not.
o. Calculate $f_{X|Y}(x \mid y)$.
p. Is the result you obtained in the previous part a Gaussian distribution? Prove your answer.

6. Dealing with Covariance Matrix and Its Properties (16+2 Pts.)

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

Covariance Matrix is a generalisation of the concept of variance to multiple dimension. It has numerous uses in many different areas. For example, a transformation matrix, called **Whitening Transformation**, can be derived from covariance matrix that allows one to completely decorrelate the data. It also plays a key role in many pattern recognition algorithms.

In this problem, you are going to examine your knowledge of covariance matrices and their properties. You are also going to review the concepts of **Eigenvalues** and **Eigenvectors**.

First, Consider the following dataset:

1	2	-1	1	0	2	2	1	-1	0
1	-2	0	-1	-1	0	2	0	1	2
0	-1	-2	0	1	-2	0	2	1	1

a. Calculate the covariance matrix associated with the data.
b. Find the correlations between different data dimensions.
c. On which dimension are the data scattered more?
d. 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.
e. Find a transformation to whiten data associated with the given covariance matrix.
f. Show that the obtained covariance matrix is a valid one.

Note: You may use MATLAB or Python in you calculations, but you are not allowed to use their functions.

Next, assume matrix A and vectors N and M are given as below:

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$$

$$N = \begin{bmatrix} 6 \\ -5 \end{bmatrix} \qquad M = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

g. Investigate whether N and M are eigenvectors of A, without directly calculating them.
h. Show that one eigenvalue of A is 7. Find the corresponding eigenvector.
i. Compute $\exp(A)$.

Hint: First diagonalise A as $A = P\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}P^{-1}$ where λ_1 and λ_2 are the eigenvalues of A and P is a 2×2 invertible matrix. Then use $\exp(A) = P\begin{pmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{pmatrix}P^{-1}$

Finally, consider a matrix A with the following eigenvalues and eigenvectors

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \text{ with eigenvalue } \lambda_1 = -2,$$

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \text{ and } \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \text{ with eigenvalue } \lambda_2 = 4$$

j. Find A.

7. Application of Simple Statistical Calculations in a Real-World Problem (8 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 10. As can be seen, $0 \le x \le 105$ and $0 \le y \le 68$.

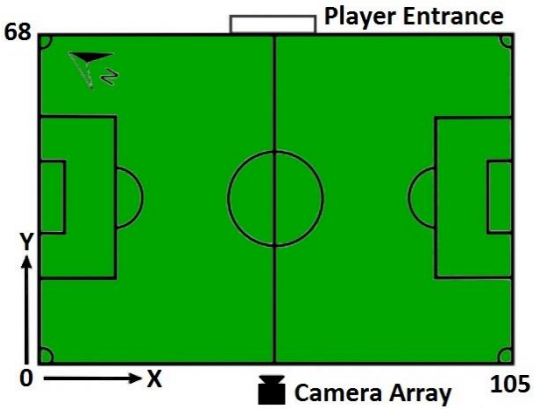

Figure 10 Coordinate system and other assumptions in the "first half logs" dataset

The data format as well as a part of the dataset is represented in Figure 11. 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 11 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.

8. Some Explanatory Questions (10 Pts.)

Please answer the following questions as clear as possible:

a. Regression is said to be a generalisation of classification task. How can you explain that?
b. Imagine you have two options for transferring a document: fax it, that is, send the image, or use an optical character reader (OCR) and send the corresponding text file. Discuss the advantages and disadvantages of the two methods and determine when one would be preferable over the other.
c. Pattern recognition is often defined as an ill-posed problem. Explain why.
d. Let's assume an OCR system which stores the bitmap of every character as a template and matches these templates with the observed character pixel by pixel. Discuss when this system would fail. Why do you think barcode readers are still used?
e. Imagine a face recognition system, and a 200x200 face image which would be a 40,000-dimenstional vector. With one pixel shift to the left, this would be a completely different vector in the 40,000-dimensional space. How would you construct a face recognition algorithm which is robust to these variations?

Good Luck!
Ali Abbasi, Farhad Dalirani