Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

Exercise 2.1 – *Principal component analysis*

a) *Consider the following dataset consisting of 4 2-dimentional vector:*

$$x^{(1)} = (1,1)^T, x^{(2)} = (2,1)^T, x^{(3)} = (3,1)^T, x^{(4)} = (4,1)^T$$

Compress this dataset to a 1-dimensional set using the PCA i.e. derive the encoder function $f(x) = D^T \cdot x$·as defined in the lecture. Then apply f to the dataset in order to compress it.

$$\sum x_1 = 10\,; \Rightarrow \frac{10}{4} = \frac{5}{2}$$

$$\sum x_2 = 5\,; \Rightarrow \frac{5}{4}$$

$$\bar{x}^{(1)} = \begin{pmatrix} -\frac{3}{2} \\ -\frac{1}{4} \end{pmatrix}; \bar{x}^{(2)} = \begin{pmatrix} -\frac{1}{2} \\ \frac{3}{4} \end{pmatrix}; \bar{x}^{(3)} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{4} \end{pmatrix}; \bar{x}^{(4)} = \begin{pmatrix} \frac{3}{2} \\ -\frac{1}{4} \end{pmatrix}$$

Recourse Basis:

$$\bar{X} = \begin{pmatrix} -3/2 & -1/4 \\ -1/2 & 3/4 \\ 1/2 & -1/4 \\ 3/2 & -1/4 \end{pmatrix}$$

$$\bar{X}^T\bar{X} = \begin{pmatrix} -3/2 & -1/2 & 1/2 & 3/2 \\ -1/4 & 3/4 & -1/4 & -1/4 \end{pmatrix} \begin{pmatrix} -3/2 & -1/4 \\ -1/2 & 3/4 \\ 1/2 & -1/4 \\ 3/2 & -1/4 \end{pmatrix}$$

$$\bar{X}^T\bar{X} = \begin{pmatrix} 5 & -1/2 \\ -1/2 & 3/4 \end{pmatrix}$$

Eigenvalues: 5.0580; 0.6920

$V = \begin{bmatrix} -8.6106 \\ 0.11606 \end{bmatrix}$; eigenvector corresponds to largest eigenvalue.

Encoding training data:

$$f(x) = D^T\bar{X}^T = [-8.6106 \quad 0.11606] \begin{bmatrix} -3/2 & -1/2 & 1/2 & 3/2 \\ -1/4 & 3/4 & -1/4 & -1/4 \end{bmatrix}$$

$$= (1.4612 \quad 0.5831 \quad -0.5255 \quad -1.262)$$

$$x^{(1)} = 1.4612;\; x^{(2)} = 0.5831;\; x^{(3)} = -0.5255;\; x^{(4)} = -1.262$$

Reconstructing Data:

$$DD^T\bar{x}^1 = \begin{pmatrix} -1.0922 \\ 0.14664 \end{pmatrix}$$

$$DD^T\bar{x}^2 = \begin{pmatrix} -0.44741 \\ 0.06007 \end{pmatrix}$$

$$DD^T\bar{x}^3 = \begin{pmatrix} 0.39741 \\ -0.05336 \end{pmatrix}$$

$$DD^T\bar{x}^4 = \begin{pmatrix} 1.1422 \\ -0.15335 \end{pmatrix}$$

b) *Now consider the set:*

$$x^{(1)} = (-1,1)^T, x^{(2)} = (-2,2)^T, x^{(3)} = (-1,3)^T, x^{(4)} = (-1,4)^T$$

As in part (a) compress this set by deriving the encoder function f and apply it to the set.

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

$$\sum x_1 = -5\,; \implies -\frac{5}{4} = -1.25$$

$$\sum x_2 = 10\,; \implies \frac{10}{4} = 2.5$$

$$\bar{x}^{(1)} = \begin{pmatrix} 0.25 \\ -1.5 \end{pmatrix};\ \bar{x}^{(2)} = \begin{pmatrix} -0.75 \\ -0.5 \end{pmatrix};\ \bar{x}^{(3)} = \begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix};\ \bar{x}^{(4)} = \begin{pmatrix} 0.25 \\ 1.5 \end{pmatrix}$$

Recourse Basis:

$$\bar{X} = \begin{pmatrix} 0.25 & -1.5 \\ -0.75 & -0.5 \\ 0.25 & 0.5 \\ 0.25 & 1.5 \end{pmatrix}$$

$$\bar{X}^T\bar{X} = \begin{pmatrix} 0.25 & -0.75 & 0.25 & 0.25 \\ -1.5 & -0.5 & 0.5 & 1.5 \end{pmatrix} \begin{pmatrix} 0.25 & -1.5 \\ -0.75 & -0.5 \\ 0.25 & 0.5 \\ 0.25 & 1.5 \end{pmatrix}$$

$$\bar{X}^T\bar{X} = \begin{pmatrix} 0.75 & 0.5 \\ 0.5 & 5 \end{pmatrix}$$

Eigenvalues: 5.0580; 0.6919

$V = \begin{bmatrix} 0.11587 \\ -8.63027 \end{bmatrix}$; eigenvector corresponds to the largest eigenvalue.

Encoding training data:

$$f(x) = D^T\bar{X}^T = [0.11587 \quad -8.6327] \begin{bmatrix} 0.25 & -0.75 & 0.25 & 0.25 \\ -1.5 & -0.5 & 0.5 & 1.5 \end{bmatrix}$$

$$= (1.3235 \quad 0.3446 \quad -0.4025 \quad -1.2655)$$

$$x^{(1)} = 1.3235;\ \ x^{(2)} = 0.3446;\ x^{(3)} = -0.4023;\ x^{(4)} = -1.2655$$

Reconstructing Data:

$$DD^T\bar{x}^1 = \begin{pmatrix} 0.5335 \\ -1.1422 \end{pmatrix}$$

$$DD^T\bar{x}^2 = \begin{pmatrix} 0.03993 \\ -0.29741 \end{pmatrix}$$

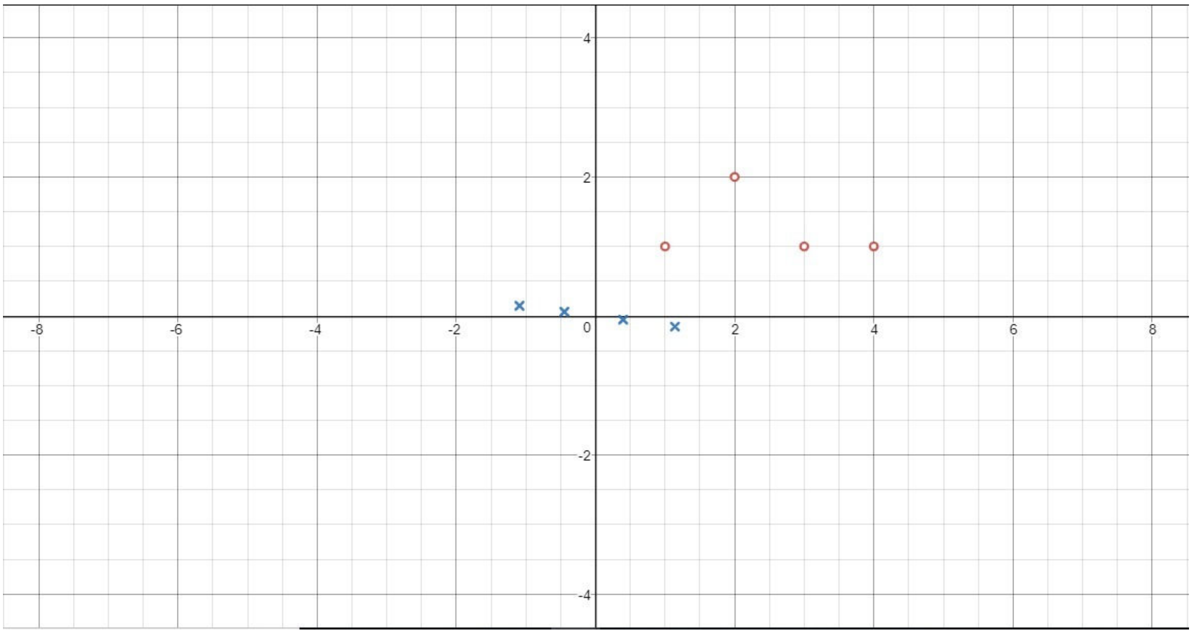
$$DD^T\bar{x}^3 = \begin{pmatrix} -0.04664 \\ 0.34741 \end{pmatrix}$$

$$DD^T\bar{x}^4 = \begin{pmatrix} -0.14664 \\ 1.09222 \end{pmatrix}$$

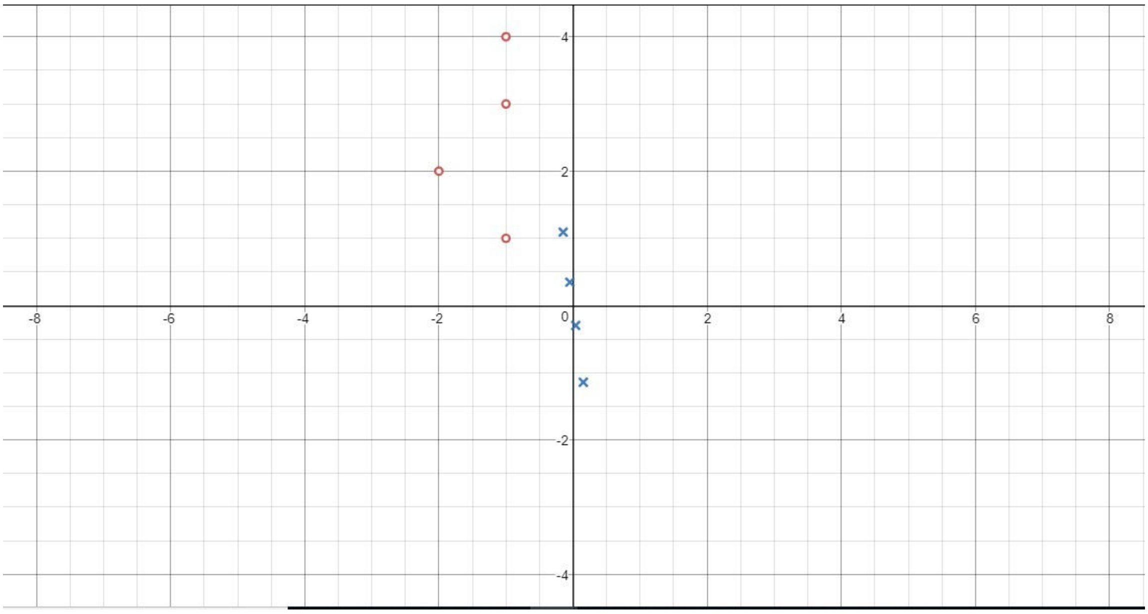
c) *For both the parts (a) and (b) sketch the corresponding datasets in a separate figure. Also include the reconstructed vectors into the corresponding figures. Explain the values of the reconstructed vectors.*

Corresponding dataset for part (a). Red circles represent uncompressed data; blue circles represent encoded data points.

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564



Corresponding dataset for the part b). Red circles represent uncompressed data; blue circles represent encoded data points.



For the cases, which lie along a 1-dimensional manifold (despite of being plotted in 2D), the dimensionality has been reduced which is clearly shown by the encoded data points. Those cases have an increasing order of difficulty, based on existing techniques. The dimensionality reduction technique is based on an implicit assumption that the data lies along some low-dimensional manifold.

Note: Each value is approximate, since it can slightly differ, depending on the used tool (Matlab, Python, Anaconda, some online tools etc.)

d) *PCA can be used for a lot of applications. One of them is image recognition. Discuss briefly how PCA can be utilized in such task.*

The main idea is to express large 1-D vector of pixels constructed from 2-D facial image into the compact principal components of the feature space (eigenspace projection).

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

In PCA, each image is represented as a linear combination of weighted eigenvectos – eigenfaces. These eigenvectors are obtained from covariance matrix of a training image set. The weights are found out after selecting a set of most relevant Eigenfaces. Recognition is performed by projecting a test image onto the subspace spanned by the eigenfaces and then
classification is done by measuring minimum Euclidean distance[1].
However, using PCA for the image processing has certain drawbacks:

- By rearranging pixels to 1-dimentional vector (column by column), relations of given pixels regarding neighboring-rows are not taken into consideration;
- Small change or an error in input image will influence the complete eigen-representation (like in all linear algorithms)

e) *Is PCA a supervised or unsupervised algorithm? Explain your answer and discuss what the tunable parameter in PCA is. How can we choose this parameter?*
In general, in supervised learning there always will be a "teacher", e.g. in regression – it will be dependent variable "telling" the algorithm whether you learnt something right, or inching towards the correct prediction.
On the other hand, in unsupervised learning there is no "teacher".
Thus, PCA is unsupervised learning algorithm, since it just applies a linear transformation to the input data and gives the output, and there the dataset has neither the target variable to supervise the learning process, nor response value and the data is unlabeled.
Tunable parameter – basically the parameter, user can change while running the algorithm. In PCA this parameter controls the sparsity of the estimated matrix and the number of outliers as a byproduct. One of the methods to choose it would be to estimate this parameter from the data resulting in a fully automated system.

f) *Why is PCA a linear dimensionality reduction? What are non-linear dimensionality reduction techniques?*
In general, PCA is used to reduce the same dataset into two dimensions. There the space onto the original data-points are projected as well as the transformation (mapping) are linear. Thus, this is possible to represent the entire transformation in terms of linear algebraic operation (matrices and vectors). Additionally, vectors (or eigenvectors!) in n-dimensional space are lines.
Hence, PCA is linear dimension reduction method.
Non-linear dimensionality reduction techniques are Maniford learning algorithms (e.g. Isomap, Fast Manifold Learning, Locally-linear embedding, Laplacian eigenmaps, Sammon's mapping, Autoencoders etc).

Exercise 2.2 – Derivatives and Critical points

a) *Let $f(x)$ be a twice differentiable function that satisfies the following equation: $f(x) = \sin(\pi e^x)$. What are the values $f'(0)$ and $f''(0)$?*
Let's use Matlab functions to calculate first and second derivatives. The code will look following:

```
>> syms x
>> f=sin(pi*exp(x)); %define the function
```

[1] Face Recognition Using Principal Component Analysis Method, Liton Chandra Paul, Abdulla Al Sumam, Nov.2012

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

```
>> f_1=diff(f) %first derivative

f_1 =

pi*exp(x)*cos(pi*exp(x))

>> f_2=diff(f,2) %second derivative

f_2 =

pi*exp(x)*cos(pi*exp(x)) - pi^2*sin(pi*exp(x))*exp(2*x)

>> f_1_of_0=pi*exp(0)*cos(pi*exp(0))

f_1_of_0 =

   -3.1416

>> f_2_of_0=pi*exp(x)*cos(pi*exp(x)) - pi^2*sin(pi*exp(x))*exp(2*x)

f_2_of_0 =

pi*exp(x)*cos(pi*exp(x)) - (2778046668940015*sin(pi*exp(x))*exp(2*x))/281474976710656

>> f_2_of_0=pi*exp(0)*cos(pi*exp(0)) - pi^2*sin(pi*exp(0))*exp(2*0)

f_2_of_0 =

   -3.1416
```

Thus, $f'(0)$ actually equal $f''(0)$ and equal to -3.1416.

b) *Let $f(x) = 9x^2 - 3x^3$ where x defined in $-4 < x < 4$, find all critical points of the function $f(x)$, and indicate if they are saddle, local or global min/max points.*
Critical value c of the function is such that for $f'(c) = 0$ or $f'(c)$ does not exist[2].
So let's start with finding the first derivative:

$$f'(x) = 18x - 9x^2 = -9x^2 + 18 = -9x(x - 2)$$

Given that $f'(x) = 0$, we can solve the equation above as:

$-9x = 0 \Rightarrow x = 0$ $\qquad$ $x - 2 = 0 \Rightarrow x = 2$

Thus, needed points are: $(0, 0)$ and $(2,0)$. Clearly they both belong to the given interval of x.

Now we have to apply second derivative check to find out if those points are local min/max or saddle.

[2] https://sites.math.washington.edu

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

The condition will be following:
Assume, $x = c$ – is the critical value, then:
1. $f''(c) > 0$ – then $x = c$ gives local min
2. $f''(c) < 0$ – then $x = c$ gives local max

Thus, by substituting c with critical values, we get:
$f''(2) = 18 - 18 \cdot 2 = -18 < 0$ – local max
$f''(0) = 18 - 18 \cdot 0 = 18 > 0$ – local min

Exercise 2.3 – *Matrices*

Consider a real symmetric matrix $A \in R^{m\times n}$ with the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m$. Prove that the eigenvalues of A^k are $\lambda_1^k, \lambda_m^k, \dots, \lambda_m^k$ for $k \in N$.

Suppose A has m linearly independent eigenvectors. Let Q be the matrix that has eigenvectors of A as it columns. Then we can write following:
$AQ = Q\Lambda$, where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_m)$.
Multiplying by Q^T from right we get:

$$Q^{-1}\Lambda Q = \Lambda = \begin{bmatrix} \lambda_1^k & 0 & \dots & & 0 \\ 0 & \lambda_2^k & & & \vdots \\ \vdots & & \ddots & \dots & \vdots \\ & 0 & 0 & \lambda_{m-1}^k & 0 \\ 0 & 0 & 0 & 0 & \lambda_m^k \end{bmatrix}$$

The characteristic polynomial of the matrix A^k is given by:

$$\begin{aligned} p(t) &= \det(A^k - tI) \\ &= \det(Q^{-1})\det\big(QA^k - tI\big)\det(Q) \\ &= \det\big(Q^{-1}\big(A^k - tI\big)s\big) \end{aligned}$$

$$= \det\big(Q^{-1}A^kQ - tI\big) = \begin{vmatrix} \lambda_1^k - t & & \dots & \\ & \ddots & \dots & \vdots \\ \vdots & & \lambda_{m-1}^k - t & \dots \\ & & \dots & \lambda_m^k - t \end{vmatrix}$$

$$= \prod_{i=1}^{m}(\lambda_i^k - t)$$

Since the roots of the characteristic polynomial are all the eigenvalues, we see that $\lambda_1^k, \dots, \lambda_n^k$ are all the eigenvalues of A^k.