Exercise Sheet 2

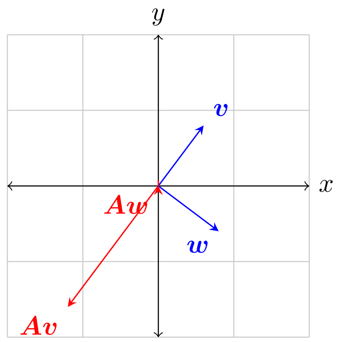
December 12, 2020

Exercise 2.1 - Eigenvalue decomposition & SVD

(a)

We know that

$$\boldsymbol{Av} = av = -2\begin{bmatrix}0.6\\0.8\end{bmatrix} = \begin{bmatrix}-1.2\\-1.6\end{bmatrix}$$
$$\boldsymbol{Aw} = aw = 0\begin{bmatrix}0.8\\-0.6\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$



The matrix A causes vector v to stretch and reverse direction, while causes vector w to equal to zero.

(b)

There are infinite eigenvectors for the eigenvalue $a = -2$. Any multiple of an eigenvector is an eigenvector, which spans the set of eigenvectors.
The general solution $\boldsymbol{X} = \begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}cx_2\\x_2\end{bmatrix}$

$$\begin{aligned} x_2 &= 0.8 \\ x_1 &= cx_2 \\ 0.6 &= c(0.8) \\ \Rightarrow c &= 0.75 \\ \boldsymbol{X} &= \begin{bmatrix}0.75x_2\\x_2\end{bmatrix} \end{aligned}$$

One such eigenvector is $\begin{bmatrix}1.5\\2\end{bmatrix}$.

(c)

$$\boldsymbol{A} = \boldsymbol{V} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{V}^{-1}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix $(\lambda_1, ..., \lambda_n)$ and $\boldsymbol{V}$ is the matrix of corresponding eigenvectors. However $\boldsymbol{V}$ is an orthogonal Matrix which implies that:

$$\boldsymbol{V}^{-1} = \boldsymbol{V}^T$$

and $\boldsymbol{V}$ is a symmetric matrix which implies that: $\boldsymbol{V} = \boldsymbol{V}^T$

$$\boldsymbol{A} = \boldsymbol{V} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{V}$$

$$\boldsymbol{A} = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix} = \begin{bmatrix} -0.72 & -0.96 \\ -0.96 & -1.28 \end{bmatrix}$$

(d)

$$\boldsymbol{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\boldsymbol{V}^{-1} = \frac{1}{-0.36 - 0.64} \begin{bmatrix} -0.6 & -0.8 \\ -0.8 & 0.6 \end{bmatrix}$$

$$\boldsymbol{V}^{-1} = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix} = \boldsymbol{V}$$

(e)

singular matrix - a square matrix that is not invertible. A matrix is singular if its determinant is 0.
positive definite - a symmetric matrix whose eigenvalues are all positive.
positive semi-definite - a symmetric matrix whose eigenvalues are all positive or zero valued.
negative definite - a symmetric matrix whose eigenvalues are all negative.
negative semi-definite a symmetric matrix whose eigenvalues are all negative or zero valued.

The matrix $\boldsymbol{A}$ is a negative semi-definite matrix because its eigenvalues (-2 and 0) are all negative or zero-valued.

(f)

Singular value decomposition (SVD) is defined for all matrices of size $m \times n$ and can decompose a matrix into several components. It can be used to extract useful information from the matrix unlike eigendecomposition. SVD is used to generalize matrix inversion to non-square matrices. Other applications of SVD include matrix approximation, and determining the rank, range, and null space of a matrix.

We use both eigendecomposition and SVD during principal components analysis (PCA). Apart from that, if the given matrix is symmetric and positive definite, SVD simplifies to eigendecomposition, hence we can use both of these techniques.

Exercise 2.2 - Principal Component Analysis

(a)

Centering the data is not a requirement; it does however simplify notation and helps us calculate the covariance matrix- which gives us direction of variability in the sample. Figure 1

(b)

Encoding matrix is computed by performing SVD on the covariance matrix (size $n \times n$). This gives us a matrix with all eigenvectors (size $n \times n$) of which we pick out only k eigenvectors. The resulting matrix of size $n \times k$ is the encoding matrix. Figure 2

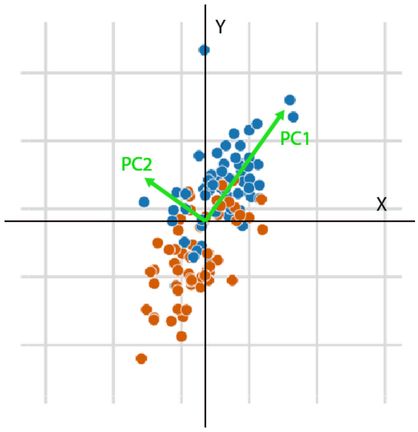


Figure 1: *Centered dataset with all principle components marked*

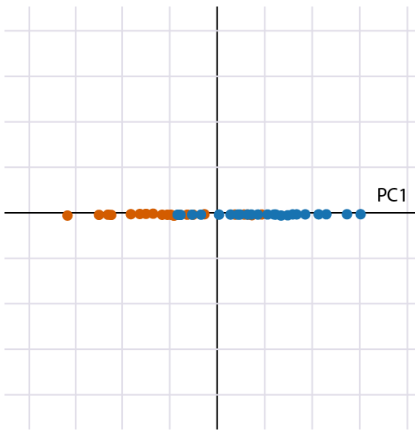


Figure 2: *1-D data after encoding*

(c)

The Decoding Matrix D contains eigenvectors corresponding to the largest eigenvalues of

$$X^T X$$

Where X is the design matrix or the Data Matrix, each column of the Data matrix is a training vector. Figure 3

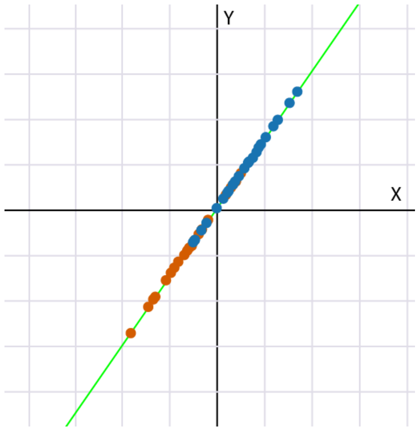


Figure 3: *Data in original space after decoding*

(d)

4 show a toy data set with tow classes, where pc1 fails to separate the two classes as it's not in the direction of the highest variance. Figure 4

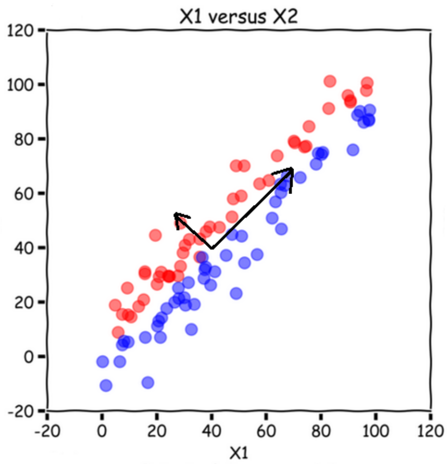


Figure 4: *Date set in 2-dimensional space*

since variance is a measure of the "variability" or "spread" of the data, we want our principle components to be in the direction where the data is most separable
when projecting the data points on pc1 , we notice that we can't separate the one dimensional data into 2 classes.

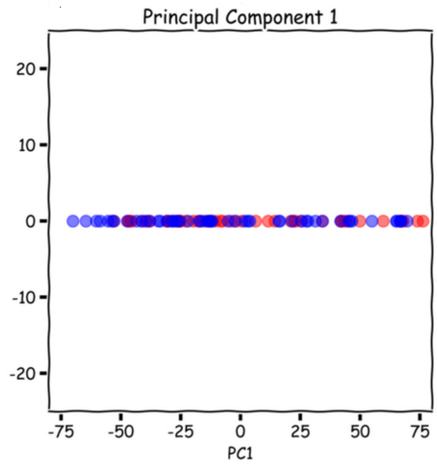


Figure 5: *Data projected to pc1*

5 shows the data points on pc1.

Exercise 2.3 - Applications of eigendecomposition in machine learning

Some of the applications of eigendecomposition are:

1. **k-means clustering** [1] - PCA can be exploited with k-means to viualize high dimensional data such clusters in National health and nutrition survey
2. **Linear Discriminant Analysis** [2] - LDA, similarly as PCA, is used as dimensionality reduction technique in the pre-processing step for pattern-classification and machine learning applications, but, in contrast to PCA, LDA is "supervised".
3. **Word embeddings** [3]- eigendecomposition has been used to view word vectors as semantic groups from different domains.
4. **Spectral clustering** [4]- spectral clustering techniques uses eigenvalues of the similarity matrix of the data to perform dimensionality reduction before clustering in fewer dimensions .

References

[1] Dmitriy Kavyazin. *Principal Component Analysis and k-means Clustering to Visualize a High Dimensional Dataset*, 2019.

[2] Sebastian Raschka. Linear discriminant analysis, 2014.

[3] Madotto Andrea Shin, Jamin and Pascale Fung. Interpreting word embeddings with eigenvector analysis. In *Workshop on Interpretability and Robustness in Audio, Speech, and Language (IRASL)*. NeurIPS IRASL, 2018.

[4] Wikipedia. Spectral clustering, 2020-11-24.