CE888 Assignment 1

Ogulcan Ozer

***Abstract*—The traditional unsupervised machine learning algorithms like k-Means, PCA and other linear factor models are proven to be very useful for many important applications. But in this day and age we are able to collect very complex, high dimensional data and these simple but efficient algorithms are failing to learn or represent it [1]. In this work we are going to use auto encoders to reduce the dimensions of three different datasets and try to extract good features using different activation functions. Then finally, we will cluster the data by using the extracted features and compare the results with the traditional techniques.**

I. Introduction

UNSUPERVISED learning is a branch of machine learning that deals with unlabeled data to find patterns, groups and relations. This sub section of machine learning is useful to describe and understand unknown data, which also helps us understand the properties of future data samples. Right now we are collecting data much faster than we can process it. And most of the time the data we acquire is unlabeled, and these huge amounts of data are complex, noisy and it might include unnecessary information.

There are different ways to reduce the complexity of a given data and learn good features from it. In this work, we will look at auto encoders, one of those procedures used to learn features from a given data. Auto encoders are a part of representational learning, which can be trained with or without labels. They play a key role in deep learning applications. They are used in dimensionality reduction and feature extraction, and together with deep learning they gained a lot of traction after Hinton proposed a method to train deep belief networks one layer at a time [2]. The motivation for using auto encoders is, in what they do they are very similar to traditional feature selection algorithms, but they are also capable of learning good features from complex and high dimensional data. This aspect of auto encoders and unsupervised learning in general is important. As LeCun states in [3], people are not paying attention to unsupervised learning because of the success of the supervised learning in recent years. But it will be far more important in the future because of how living beings perceive and learn, they do not need labels to recognize the world around them.

In this paper we are going to look at previous work (Section II), look at different methods and present their results to show how they compare. In Section III, we are going to present our goals and what we want to accomplish. Then we will talk about the datasets that we are going to use for the task and give detailed information about them. The planned experiments and analyses for the task will be mentioned in Section IV. Evaluation methods that are going to be used after the experiments will be presented in Section V.

II. Background

The conclusion goes here.

III. Conclusion

The conclusion goes here.

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgment

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.