Report of Mini Project 2

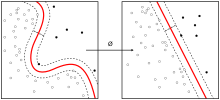
Individual Task

* **Introduction**

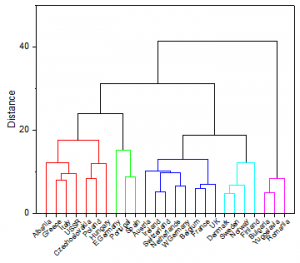
Unsupervised learning is a type of [machine learning](https://www.mathworks.com/solutions/machine-learning.html) algorithm used to draw inferences from datasets consisting of input data without labeled responses[1]. As it doesn’t have a teacher to evaluate its work, it have to form and evaluate on its own. In other words, in unsupervised learning, a developer only need to provide a goal and set a reward mechanism for it, and then the program will train itself to achieve its goal.

According to IBM, machine learning could be roughly classified into three categories, supervised learning, unsupervised learning and reinforcement learning. The supervised learning required a labeled data set, so it can provide the program with the accurate error rate as a feedback. Basically, the main goal for supervised learning is to reduce the error rate. While unsupervised learning doesn’t need a feedback. Instead, the main goal for it is to perform better in the test and get a reward.

Cluster analysis is the most applicable and most popular method of unsupervised learning. Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters)[2].



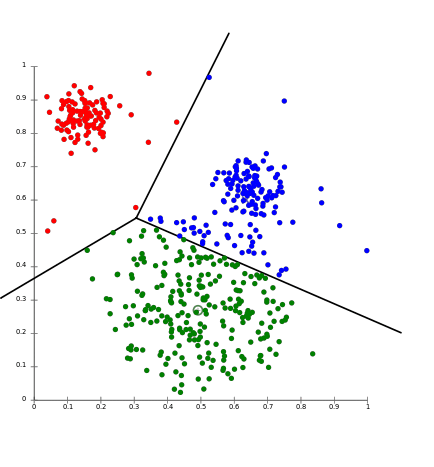
This is a basic cluster analysis.



<https://www.originlab.com/doc/Origin-Help/Cluster-Analysis>

And this is much more detailed.

In cluster analysis, there are many ways to finish the classification. For example, the K-means.

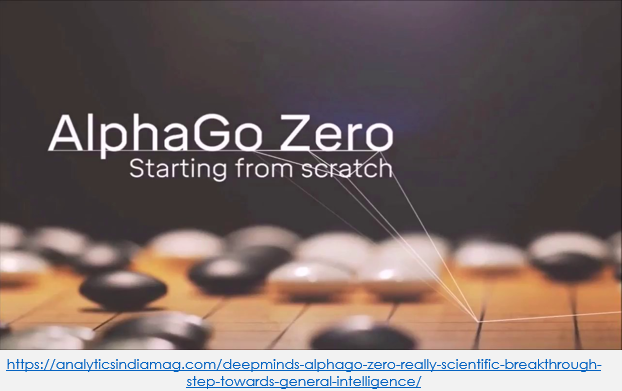


<https://aws.amazon.com/cn/blogs/machine-learning/k-means-clustering-with-amazon-sagemaker/>

By k-means, Points that are more likely to be classified as a class can be classified as a class according to the distance between the feature dimensions of the data.

* **Pros and Cons**

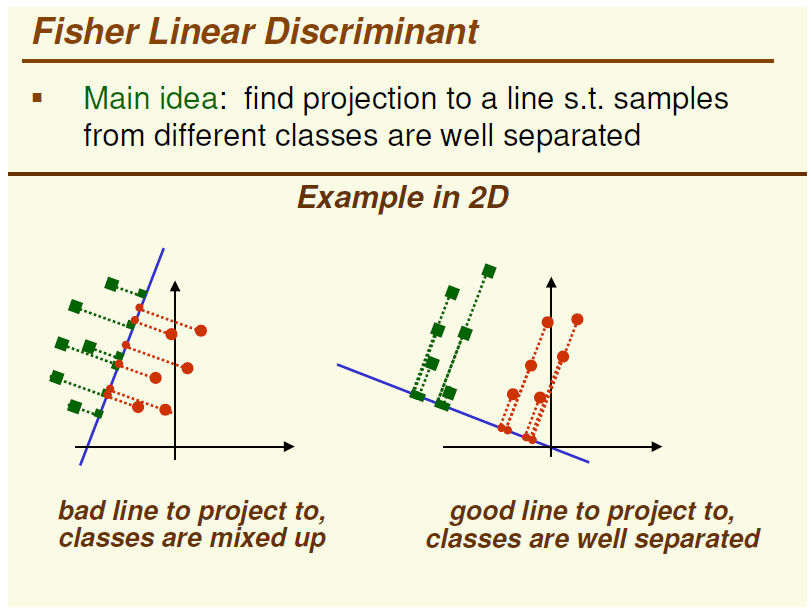
Of course unsupervised learning has many pros. For example, because the machine could do the ‘dirty work’ itself and doesn’t required a labeled data set, it could be more convenient and cheap. It also need less complicated hardware and equipment to train the model. What’s more, an unsupervised learning could avoid biases caused by human intervention.



In a paper just published in nature, DeepMind researchers have unveiled the latest version of Alpha Go Zero. It performs better in games, learns faster, and does well with less computing hardware. More importantly, though, unlike the original version, Alpha Go Zero successfully taught itself the game without consulting human experts.

Nobody knows the best solution of Go, therefore the unsupervised provide the program with no limitation to solve this problem. Guidance from human experts may limit computer power.

However, an unsupervised learning machine may be in trouble as its training algorithm is written by human and a human always have some blind spot. In Chopra’s work, when a manipulator was asked to move a block of wood to a designated spot on the table, it moved the table. Another weird situation showed up in the work of Sims. Their program was asked to evolve creatures that could run very fast, yet it evolve a strange creature with extremely height and it gained super high speed in the process of falling. In other occasions, unsupervised learning did not turned out to be as accurate as supervised learning.



https://blog.csdn.net/qq\_28620327/article/details/82345511

Here’s the example of LDA(right) and PCA(left). They are quite similar, except the LDA is supervised while the PCA is not. Apparently, LDA did a better job in this classification.

* **Recommendation**

Unsupervised learning is a wonderful tool in some situation, such as developing a app which could learn from the habits of each user and recommend their favorite songs. Also, it is convenient for city planner to divide the city up based on the function of the building.

Unsupervised learning has distinct advantages on classifying big data. Therefore, a cybersecurity worker could complete the identification of fraudulent mails through large data set.

* Conclusion

Unsupervised learning is a mature method of training a model and it has clear pros and cons. Compared to supervised learning, it is ideal for people who need to deal with big data and have limited budgets

* Reference

[1].https://www.mathworks.com/discovery/unsupervised-learning.html#targetText=Unsupervised%20learning%20is%20a%20type,patterns%20or%20grouping%20in%20data.

[2].https://en.wikipedia.org/wiki/Cluster\_analysis#targetText=Cluster%20analysis%20or%20clustering%20is,in%20other%20groups%20(clusters).