1) What are the three primary types of machine learning, i.e., what type of learning do the classifiers do?

a. Describe the three types concisely.

1. Supervised learning
2. Unsupervised learning
3. Reinforce learning

b. Give an example of an application for two of those types.

* + - Supervised learning (Decision Tree)
    - Unsupervised learning (KMeans)

2) Give a short description of Adaline and perceptron models.

a. What are the primary components of each?

* Adaline

Input vectors

Output values

Weights

Bias

Threshold

Weighted sum

Activation Function

* Perceptron

Input vectors

Output values

Weights

Bias

Threshold

Weighted sum

Activation Function

b. Compare and contrast the two models.

The two models are differentiated in their activation function. In Adaline, the output of activation function is continuous linear function and in Perceptron the output of activation function will be binary.

3) Give a short description of one method for extending a binary classifier to many classes.

a. What primary algebraic component is used to represent features in data?

Vector is a key data structure in linear algebra, used to represent features in data.

b. How is this algebraic component used in other classifiers such as SVMs?

Vectors in SVM is to maximize the margin of data and the hyperplane.

4) Briefly describe what is an SVM?

a. What are the key concepts?

Finding the hyper plane to separate different classes, then maximize the margin to get the better result.

b. How does it operate in higher dimensions - what common method does it use?

It used kernel to project the vectors in higher dimension, and find a hyperplane to split the data.

5) Compare briefly Adaline and Logistic regression

a. When would you use one over the other method?

If we wish to get the probability of the classification, then logistic regression will be the better choice, because the output will be between 0-1.

If we expect the output be binary, then Adaline will be the better choice.

b. What single step or “exchange” would convert Adaline into logistic regression?

Convert the activation function to sigmoid function.

6) What is overfitting and underfitting?

a. What could you do to each to avoid these situations?

To avoid overfitting, we can use

1. Cross-validation to examine the data
2. Increase training data.
3. Early stopping (stop the training before overfitting)
4. Regularization
5. Use dropout for neural networks to tackle overfitting.

To avoid underfitting, we can add the epodes or we can reduce the learning rate.

1. Increase model complexity
2. Standardization
3. Increase number of features.
4. Remove noise from the data.
5. Increase the number of epochs or increase the duration of training to get better results.

b. Are linear or nonlinear separable data classes more prone to these?

Nonlinear separable data will be more prone to Underfitting/Overfitting.

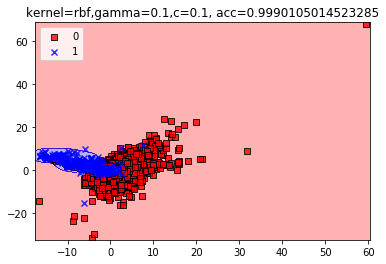
Underfitting, because the separate procedure will be more complex, it’s harder to produce a perfect decision boundary.

Overfitting, if the model produces a perfect decision boundary, it may negatively impact the performance of the model on new data, since it picked up noise or random fluctuations in the training data and learned as concepts by the model.

7.

1. You will use parts of Lab1a to classify the provided credit card data as either fraudulent or valid. You may use either ADALINE or the SVM. Keep in mind that you will need to replace the IRIS dataset with the credit card dataset described above. Show the resulting graph of fraudulent and valid transactions separated out.

SVM.



1. You will use parts of Lab1b to classify the provided credit card data into clusters that define each of the feature categories. You will justify your choice of clustering algorithm with a sentence or two that describes why it is better suited than the other clustering approaches, what methods you use to measure similarity and distance. Show graphically, and also show if there is any relationship between some clusters and fraudulent transactions.

|  |  |
| --- | --- |
| Class=0 | Class=1 |
|  |  |

In Kmeans algorithm, we have to pre-specified number of clusters or often use elbow method to decide how many clusters should assign to the classifier.

In Approx Agglomerative Clustering, the algorithm automatically chose datapoints as prototypes that are most far in the plane, then recursively assign nearest points to the prototypes before meets the threshold. So, its better deals with data and greater data. In addition, this algorithm is less efficient and memory demanding.

The both algorithms used Euclidean Distance to compute distance.

In conclusion, we clearly notice we have two clusters, so Kmeans will be a more efficient way to split the data. Also, with large amount of data, K-means compute faster.

Results

|  |  |
| --- | --- |
| Approx Agglomerative Clustering | Kmeans |
|  |  |