EML Assignment 1 Problem 1

Muhammed Saeed and Ali Salaheldin Ali Ahmed

November 2022

1 Unsupervised and supervised learning

In supervised machine learning, we have the input and output pairs of observation and target (x_i, y_i), the goal is to find an approximation function $\hat{f}(x)$ such that

$$y_i \approx \hat{f}(x_i), \forall i$$

including unseen (x_i, y_i) pairs.

On the other hand for unsupervised learning, we have a set of input observations $\{x_1, x_2, ..., x_n\}$ and the goal is to elucidate relationships between the variables or the observations, often equated with clusters analysis. Examples of unsupervised tasks are Clustering, PCA "dimensional reduction"

2 Prediction and inference

In prediction, we often treat the approximation function $\hat{f}(x)$ as a black box, i.e. we are not interested in the shape or the form of the function.

In inference, the goal is insight into the relationship between the input and the output, thus we are interested in the form as much as the response of the approximation function.

3 Classification and regression

Classification is the problem of predicting a discrete class, where each class is associated with a label output. For example: sentiment analysis with classes "positive", "neutral", "negative".

Regression is the task of predicting and estimating a continuous value or output, for example: house pricing prediction using information about the house.

4 Training and test data

The training dataset is the data used to train the model to identify a hidden pattern in the data so that the model is able to generalize to unseen data.

The testing dataset must not be seen during the training phase, and it is used to evaluate how good the model is able to generalize to unseen data.

5 Parametric and non-parametric

Parametric models are those models which has countable number of parameters.

Non-parametric models we have uncountable number of parameters the model is able to identify very complex patterns