Fundamentals of Data Mining - Homework 1

Mohammad Bahrami - 9724133

February 25, 2022

---

1 Question 1

1. **Problem Understanding:** We need to understand what the problem is. In our case, the problem is to predict the price of a car using some of its features that we have in the data.

2. **Data preparation:** In this phase we prepare the data by extracting it from data repositories. Then, we need to clear and prepare the data by:

   - Identifying and handling outliers
   - Standardizing the data
   - Reclassifying categorical variables
   - Binning numerical variables
   - Adding index

3. **Exploratory data analysis:** We can now explore our data and do some basic analysis. we can plot different features to see how they behave. For our example, we can plot the fuel consumption of different cars to see how it is changing over prices, etc.

4. **Setup:** Now our data is ready to be fed into a model, but before that, we need to establish a baseline, split the data to train, validation and test portions and balance it.

5. **Modeling:** Now we implement our machine learning model of choice, for example a MLP to discover relationships in our data. that is, predicting the price of a car by its features in our case. After implementing the model, fine tuning the hyper parameters is done to reach optimal performance.

6. **Evaluation:** In this phase, we need to compare our model's performance with the baseline. also we need to see if we are actually solving the problem that was set in phase 1. In our case, are we really predicting car prices? Are our predictions good?

7. **Deployment:** Now that our model is ready, we can start using it on real world deployment. that is, we can start predict real car prices based on their features.

All other questions are answered in their respective ipynb file.