Mandatory Assignment 3

Regression & NN Classification

Markus Kinn

Regression:

Dataset 1:

As the first dataset I used a Concrete Strength dataset. The goal of this dataset is to predict the strength of concrete given some attributes.
(https://www.kaggle.com/datasets/prathamtripathi/regression-with-neural-networking)

The has dataset has 9 columns and 1031 rows.

The dataset contains these columns:

> Cement: Percent of mixture which is cement, Blast Furnace Slag: Percent of mixture which is blast furnace slag, Water: Percent of mixture which is water, Superplasticizer: Percent of mixture which is superplasticizer, Coarse aggregate: Percent of mixture which is coarse aggregate, Fine aggregate: Percent of mixture which is fine aggregate, Age, Concrete Compressive strength.

I got the best result with regards to r-squared score using: XGBoost

Dataset 2:

The second dataset I used for regression was a medical cost dataset. The point of this dataset is to predict a person's medical cost based on attributes, mainly, about their health.
(https://www.kaggle.com/datasets/mirichoi0218/insurance)

The has dataset has 7 columns and 1338 rows.

The dataset contains these columns:

> Age, Sex, BMI, children, smoker, region and the ground truth, charges.

I got the best result with regards to r-squared score using: XGBoost

Neural Network Classification:

Dataset 1:

As the first dataset for this task, I revisited the titanic classification dataset so I could compare the results.

I got the best result with the second Neural Network Architecture, with a training score of 0.8308 and test score of 0.8161. This was slightly worse than the XGBoost one.

Dataset 2:

As the second dataset I once again revisited an old dataset, this time the Heart Attack Analysis one.

I got the exact same result with both Neural Network Architectures. Training score of 0.8950 and test score of 0.8553. However, the second Neural Network Architecture got this result with less layers.

I got the best score on this dataset with SVM.

Dataset 3:

As the third dataset I tried a new one: Star Prediction. The goal of this datatype is to predict which of 6 classes a star belongs to. (https://www.kaggle.com/datasets/deepu1109/star-dataset).

The has dataset has 7 columns and 240 rows.

The dataset contains these columns:

> Temperature, Luminosity, Radius, Absolute magnitude, Star color, Spectral class and the ground truth, Star type.

I got the best average score using the 2nd Neural Network Architecture. With this architecture I got a training score of 0.9389 and a test score of 0.9167