Advanced Machin Learning

DR.

Ahmed Hesham

Contents

Feature Extraction:

In our project, we study the feature extraction phases in three popular machine-learning models:

- SVM Regression
- ANN Artificial Neural Network
- Decision Tree Classifier

We thoroughly examine the training dynamics by analyzing the absence of cross-validation and the key hyperparameters.

1. SVM Regression Model:

SVM regression is a highly effective technique for predicting continuous target variables, but it requires careful feature selection. In our implementation, we navigate this challenge as follows:

- Input features: We have selected 'Close' (which represents the closing price) and 'MA_30' (which represents the 30-day moving average) as our input features.

- Number of features extracted: We use a succinct duo of features, comprising 'Close' and 'MA_30'.

- Names of features: Simply put, the feature landscape is adorned with 'Close' and 'MA_30'.

- Dimension of resulted features: We represent the feature space in a compact form with a representation of (number of data points, 2).

2. ANN Artificial Neural Network:

Our ANN model excels in effectively leveraging a diverse range of input features, allowing it to discern complex data patterns with ease.

- Input features: The model works with a rich tapestry of carefully processed and standardized features. These features include 'Gender', 'Specialization', and more.

- Number of features extracted: The number of features used for training is dynamic and depends on the specific requirements of the model.

Names of features: The feature space is a cornucopia of descriptors, including 'Gender', 'Specialization', and beyond.

- Dimension of resulted features: The feature space is a versatile framework, denoted as (the number of data points, and number of features), which reflects the inherent flexibility of the feature space.

3.Decision Tree classifier:

The decision tree classifier relies on careful feature selection for success, leveraging its well-known transparency and interpretability.

- Input features: We meticulously analyze the data, dropping any unnecessary columns and selecting key features such as '10percentage', 'CollegeTier', and more.

- Number of features extracted: We tailor our selection to the unique characteristics of each dataset, resulting in a nuanced assortment that highlights the flexibility of the decision tree paradigm.

- Names of features: Our selection of features is an eclectic mix of descriptors, including '10percentage', 'CollegeTier', and beyond.

- Dimension of resulted features: The resulting mosaic of data points is intertwined with the selected features, resulting in a matrix format of (number of data points, number of features).

Cross-validation:

Our implemented models lack cross-validation, omitting a pivotal role in model evaluation. This simplified evaluation approach may raise eyebrows, but it is employed in this context.

Hyperparameter:

Hyperparameters are critical in guiding model training. Below is a detailed breakdown of the hyperparameters that govern each model:

The SVM Regression Model follows a unique approach by embracing a linear kernel without any specified regularization, batch size, or epochs. This aligns with the distinctive tenets of SVM training.

The ANN Model, on the other hand, adopts Stochastic Gradient Descent (SGD) as the optimizer, with dropout layers introducing a dropout rate of 0.5 for regularization. Robust model training is ensured by setting a batch size of 32 and 200 epochs.

Finally, **The Decision Tree Classifier** operates under the aegis of Gini impurity as the criterion, with no specific constraints on maximum depth or features. The best splitter is used, and parameters governing minimum samples split and leaf fine-tune the decision tree's branching dynamics.