|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Code** | **ITM 526** | | **Title** | | **Business Analytics** | | |  |
| **Type** | **Assignment** | **Questions** | | **1** | | **Weighting** | **5%** |  |

Assigment 1



1. Build classification models → (1), (2), (3), (5): please refer to the code

**20102122 정효안**

(1) Load the dataset (10pts)

(2) Split the dataset (Training / Validation) (10pts)

(3) Train the models while changing some hyperparameters (30pts)

- Use two different learning algorithms you know.

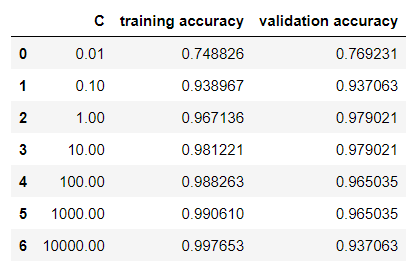
(4) Describe the meaning of the hyperparameters you adjusted.(10pts)

- min\_samples\_leaf: The minimum number of samples required to be at a leaf node. A split point at any depth will only be considered if it leaves at least min\_samples\_leaf training samples in each of the left and right branches.

- max\_depth: The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min\_samples\_split samples. Too large max\_depth can cause overfitting.

- C: Inverse of regularization strength; must be a positive float. Like in support vector machines, smaller values specify stronger regularization.

(5) Performance tables for training set / validation set according to hyperparameter settings (30pts)

테이블이(가) 표시된 사진

자동 생성된 설명테이블이(가) 표시된 사진

자동 생성된 설명- Performance metric: accuracy

(6) Find the best hyperparameters for each learning algorithms (10pts)

→Learning algorithms used (- best hyperparameter)

1. Decision Tree

* min\_samples\_leaf: 2
* max\_depth: 7

1. Logistic Regression

* C: 10.0

→This is because the validation accuracy is the highest in these cases. (The ultimate goal of training is to get a high score on test set.)