

Homework 9

Regression and Classification Using Classical Techniques

For the following two problems, find what is the best model for each problem. In one pdf file, submit your code and the score for each experiment. Bonus: Use hyper-parameter tuning and data preparation to improve the best score.

1. Diabetes Dataset

Scikit-Learn Diabetes dataset has 10 baseline variables (age, sex, body mass index, average blood pressure, and six blood serum measurements) that were obtained for each of 442 diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after the baseline. For more detail, check: https://scikit-learn.org/stable/datasets/toy_dataset.html#diabetes-dataset

You can download this dataset using the following code:

```
from sklearn import datasets
X, y = datasets.load_diabetes(return_X_y=True)
```

Using `cross_val_score(reg, X, y, scoring="neg_mean_squared_error", cv=3)`, find which is the best regressor among the following regressors to predict diabetes disease progression one year after the baseline:

1. `sklearn.linear_model.LinearRegression`
2. `sklearn.neighbors.KNeighborsRegressor`
3. `sklearn.tree.DecisionTreeRegressor`
4. `sklearn.ensemble.RandomForestRegressor`
5. `sklearn.svm.LinearSVR`
6. `sklearn.svm.SVR`

2. Iris Dataset

A famous dataset that contains the sepal and petal length and width of 150 iris flowers of three different species: Setosa, Versicolor, and Virginica.

For more detail, check: https://scikit-learn.org/stable/datasets/toy_dataset.html#iris-dataset

You can download this dataset using the following code:

```
from sklearn import datasets
X, y = datasets.load_iris(return_X_y=True)
```

Using `cross_val_score(clf, X, y, scoring="accuracy", cv=3)`, find which is the best classifier among the following classifiers to predict the iris species:

1. `sklearn.linear_model.SGDClassifier`
2. `sklearn.neighbors.KNeighborsClassifier`
3. `sklearn.tree.DecisionTreeClassifier`
4. `sklearn.ensemble.RandomForestClassifier`
5. `sklearn.svm.LinearSVC`
6. `sklearn.svm.SVC`
7. `sklearn.ensemble.VotingClassifier`