

Exercise

Use iris flower dataset from sklearn library and use `cross_val_score` against following models to measure the performance of each. In the end figure out the model with best performance,

1. Logistic Regression
2. SVM
3. Decision Tree
4. Random Forest

```
In [1]: from sklearn.datasets import load_iris
from sklearn.model_selection import cross_val_score
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
import numpy as np
```

```
In [2]: iris = load_iris()
```

Logistic Regression

```
In [19]: l_scores = cross_val_score(LogisticRegression(solver='liblinear'), iris.data, iris.target, cv=3)
l_scores
```

```
Out[19]: array([0.96, 0.96, 0.94])
```

```
In [20]: np.average(l_scores)
```

```
Out[20]: 0.9533333333333333
```

Decision Tree

```
In [21]: d_scores = cross_val_score(DecisionTreeClassifier(), iris.data, iris.target, cv=3)
d_scores
```

```
Out[21]: array([0.98, 0.92, 0.98])
```

```
In [22]: np.average(d_scores)
```

```
Out[22]: 0.96
```

Support Vector Machine (SVM)

```
In [23]: s_scores = cross_val_score(SVC(), iris.data, iris.target, cv=3)
s_scores
```

```
Out[23]: array([0.96, 0.98, 0.94])
```

```
In [24]: np.average(s_scores)
```

```
Out[24]: 0.96
```

Random Forest

```
In [25]: r_scores = cross_val_score(RandomForestClassifier(n_estimators=40), iris.data, iris.target, cv=5)
r_scores
```

```
Out[25]: array([0.98, 0.92, 0.94])
```

```
In [26]: np.average(r_scores)
```

```
Out[26]: 0.9466666666666667
```

Best score so far is from DecisionTree & SVM: 0.96

Date	Author
2021-10-02	Ehsan Zia