

Practical - 2

- Aim: Write a python program to calculate evaluation using Sklearn library to find mean square error.

- Theory:

- Mean Square Error:

The measure of mean squared error needs a target of prediction or estimation along with a predictor or estimator, which is said to be the function of the given data. MSE is the average of squares of the 'errors'.

The mean square error may be called a risk function which agrees to the expected value of the loss of squared error. This difference or the loss could be developed due to the randomness or due to the randomness or due to the estimator is not representing the information which could ~~not~~ provide a more accurate estimate.

If an estimator is an unbiased estimator, then its MSE is the same as the variance of the estimator.

- MSE Formula :

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_{obsi} - x_{modeli})^2$$

Here x_i is the vector denoting values of n number of predictions.

Also x_i is a vector representing n number of true values.

- Code :

```
import pandas as pd.  
from sklearn.linear_model import LinearRegression  
from sklearn.model_selection import KFold, cross_val_score  
  
url = 'https://raw.githubusercontent.com/.../housingdata'  
names = ['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE',  
         'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT',  
         'MEDV']  
  
df = pd.read_csv(url, delim_whitespace = True,  
                 names = names)  
  
array = df.values  
X = array[:, :-1]
```



```
y = array[:, :-1]
```

```
Kfold = Kfold(n_splits=10, random_state=7,  
              shuffle=True)
```

```
model = LinearRegression()
```

```
scoring = 'neg-mean-squared-error'
```

```
results = cross_val_score(model, x, y, cv=Kfold,  
                          scoring=scoring)
```

```
print("MAE: %.3f (%.3f)"%(results.mean(),  
                          results.std()))
```

- Output:

⇒ MAE :- 23.747 (11.143)

- Conclusion:

Hence, we have successfully performed the calculation of Mean Squared Error (MSE) using the sklearn library.

Code:



MSE.ipynb ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

+ Code + Text

Write a Python program to calculate evaluation using sklearn library to find Mean Square Error.

Shivam Tawari A-58

```
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import KFold, cross_val_score

url = 'https://raw.githubusercontent.com/jbrownlee/Datasets/master/housing.data'
names=['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT', 'MEDV']
df = pd.read_csv(url, delim_whitespace=True, names=names)

df.head()

array = df.values
X = array[:, :-1]
y = array[:, -1]

kfold= KFold (n_splits=10, random_state=7, shuffle=True)

model = LinearRegression()
scoring = 'neg_mean_squared_error'

results = cross_val_score (model, X, y, cv=kfold,scoring=scoring)

print("MAE: %.3f (%.3f)" % (results.mean(), results.std()))
```

MAE: -23.747 (11.143)