Logistic Regression

Diabetes dataset – lines 11-14

Harsha Sathish

Tag J

print(“Accuracy:”,metrics.accuracy_score(y_test,y_pred))

- This line prints the accuracy classification score.
- i.e. It compares the y_test with y_pred where
- y_test : Ground truth labels
- y_pred : Predicted labels, returned by classifier

It returnes the fraction of values of y_test correspondingly matching with y_test

Info : matrics.accuracy_score(y_test,y_pred,normalize = true)

- Best Case Scenario : True (When all val in y_test correspondingly matches with y_pred.
- Normalize is an optional parameter when false, returns the number of correctly classified sample. It is by default true.

12.
import pickle
import os

- Pickle is a module for implementing binary protocols for serializing and de-serializing python object structure.
- Os is a module providing portable way of using OS dependent functionality like reading/writing file(open()), manipulating paths,etc.
- The above 2 packages are imported.

```
#Saving the model
if not os.path.exists(‘models’):
        os.makedirs(‘models’)
```

- os.path.exists() is used to check whether a specific path exists or not and os.makedirs() used for recursive directory creation.
- Therefore, if the ‘models’ path does not exists, a new directory called ‘models’ is created.

What is pickling??

- The process whereby a python object hierarchy is converted into a byte stream.
- Unpickling – The byte stream being converted back into object hierarchy.

```
model_path = “models/logistic_reg.sav”
pickle.dump(logreg,open(model_path,’wb’))
```

- First, the file model_path is opened in write(binary) mode using the open() function.
- Then, we use pickle.dump() to put the dictionary logreg into opened file.
- Thus, the pickle operation is done to serialize the ML algorithm and will save the serialized format to file.

13.
#INITIALIZE LIST OF LISTS
data = [[6,0,33.6,50,148,72,0.627]]

- An array with the above values is created.

df = pd.dataframe(data,columns = [‘pregnant’,’insulin’,’bmi’,’age’,’glucose’,’bp ’,’pedigree’])

- A pandas dataframe will be created.
- The columns given are labels which are used for resulting frame.
- Will default to RangeIndex(0,1,2….n) if no column labels are provided.

#PREDICT ON NEW DATA new_pred = logreg.predict(df) new_pred

Now, using the data given in the above lines, we predict if the subject(person) we have considered is a diabetic patient or not using the logreg.predict(df) function.

If the return value is 1, it means the person is diabetic. If the return value is 0, the person is not diabetic.