3. Logistic Regression

March 28, 2022

[1]:

```
import numpy as np
from sklearn import datasets
from sklearn.model_selection import train_test_split

data = datasets.load_breast_cancer()
X = data.data
y = data.target
print(X.shape)

from sklearn.preprocessing import StandardScaler
sc = StandardScaler().fit(X)
X = sc.transform(X)

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2)
```

```
(569, 30)
```

[2]:

```
weights = np.random.rand(X_train.shape[1])
t = y_train
N = len(X_train)
loss = []
eta = 0.05
epsilon = 1e-5    # constant to prevent log(0) condition
```

[3]:

```
for _ in range(100):
    y = np.dot(X_train, weights)
    p = 1 / (1 + np.exp(-y))                        # logistic function
    predict_1 = t * np.log(p+epsilon)               # t*log(p)
    predict_0 = (1 - t) * np.log(1 - p+epsilon)     # (1-t)*log(1-p)
    # - [ t*log(p) + (1-t)log(1-p) ] / N
    cost = -sum(predict_1 + predict_0) / len(X)
    loss.append(cost)
    # w_j = w_j -  [(Σ[(y-t)(x_j)])/N]
    weights = weights - eta * np.dot(X_train.T,p-t)/N

y_pred = np.dot(X_test, weights)
p_pred = 1 / (1 + np.exp(-y_pred))
```

```
p_test = [1 if i>0.5 else 0 for i in p_pred]       # Returning binary result

accuracy = np.sum(p_test == y_test)/len(p_test)
print("Training Accuracy: ", accuracy)
```

```
Training Accuracy:  0.9210526315789473
```

[4]:

```
from sklearn.metrics import accuracy_score, precision_score, recall_score,␣
 ↪f1_score , classification_report, confusion_matrix
print('Accuracy:', accuracy_score(y_test, p_test))
print('Precision:', precision_score(y_test, p_test))
print('Recall:', recall_score(y_test, p_test))
print('F1 score:', f1_score(y_test, p_test))
print('\n Classification report:\n', classification_report(y_test,p_test))
print('\n Confusion matrix:\n',confusion_matrix(y_test, p_test))
```

```
Accuracy: 0.9210526315789473
Precision: 0.9714285714285714
Recall: 0.9066666666666666
F1 score: 0.9379310344827586

 Classification report:
                precision    recall  f1-score   support

           0       0.84      0.95      0.89        39
           1       0.97      0.91      0.94        75

    accuracy                           0.92       114
   macro avg       0.91      0.93      0.91       114
weighted avg       0.93      0.92      0.92       114


 Confusion matrix:
 [[37  2]
 [ 7 68]]
```