

We implement hard- and soft-margin SVMs to classify whether or not a person with given characteristics has diabetes. We attach our code and output.

Code:

GNU nano 2.5.3

File: SVM.py

```
import numpy as np
import pandas as pd
from sklearn.svm import LinearSVC
from sklearn.model_selection import GridSearchCV
from sklearn import metrics

param_grid = dict(C=np.linspace(0.1, 2, 20))

df=pd.read_csv('diabetes_scale.csv', sep=',',header=None)
datavals=df.values
data=datavals.astype(np.float)

trainlabels = data[0:499,0]
trainfeatures = data[0:499,1:]
testlabels = data[500:767,0]
testfeatures = data[500:767,1:]

print("Soft-Margin SVM:")
SoftClf =
GridSearchCV(LinearSVC(loss='hinge',penalty='l2',random_state=42),param_grid=param_grid, cv=5)
SoftClf.fit(trainfeatures,trainlabels)
print("The best parameter is %s with a training accuracy of %f"
      % (SoftClf.best_params_, SoftClf.best_score_))
SoftTestGuess=SoftClf.predict(testfeatures)
errs = metrics.accuracy_score(testlabels,SoftTestGuess)
print("The test accuracy is %f"
      % (errs))

print("Hard-Margin SVM:")
HardClf = LinearSVC(C=1000000,loss='hinge',penalty='l2',random_state=42)
HardClf.fit(trainfeatures,trainlabels)
HardTrainGuess=HardClf.predict(trainfeatures)
TrainErrs = metrics.accuracy_score(trainlabels,HardTrainGuess)
print("The train accuracy is %f"
      % (TrainErrs))
HardTestGuess=HardClf.predict(testfeatures)
TestErrs = metrics.accuracy_score(testlabels,HardTestGuess)
print("The test accuracy is %f"
      % (TestErrs))
```

Output:

Soft-Margin SVM:

The best parameter is {'C': 0.5} with a training accuracy of 0.769539

The test accuracy is 0.775281

Hard-Margin SVM:

The train accuracy is 0.739479

The test accuracy is 0.734082

We note that the soft-margin SVM performed better on both the training and testing set. This is likely because the soft-margin SVM is more robust, and is able to ignore noise that the hard-margin SVM is forced to consider. Thus, the hard-margin SVM will fit to the noise and select a less-than ideal decision boundary, while the soft-margin SVM is able to ignore noise when defining its margin.