(b)

import numpy as np

from sklearn.linear\_model import LogosticRegression

data = np.loadtxt(“q1\_data.csv”, delimiter=”,”)

clf = LogisticRegression()

from sklearn import cross\_validation

scores = cross\_validation.cross\_val\_score(clf, data[:,:2], data[:,2], cv=10)

scores

Out: array([ 0.43, 0.45, 0.47, 0.43, 0.37, 0.47, 0.46, 0.43, 0.44, 0.45])

Print(“Accuracy: %0.2f (+/- %0.2f)” %(scores.mean(), scores.std() \*2))

(d)

import numpy as np

from sklearn import RandomForestClassifier

from sklearn.ensemble import RandomForestClassifier

clf = RandomForestClassifier()

data = np.loadtxt("/Users/jayant/Work/mll/q1\_data.csv", delimiter =",")

data

Out[7]:

array([[ 0.894 , 0.448 , 0. ],

[-0.764 , 0.238 , 1. ],

[-0.285 , -0.747 , 1. ],

...,

[ 0.996 , -0.0879, 0. ],

[ 0.754 , 0.266 , 1. ],

[-0.514 , 0.858 , 0. ]])

from sklearn import cross\_validation

scores = cross\_validation.cross\_val\_score(clf, data[:, :2], data[:, 2], cv=10)

scores

Out[12]: array([ 1. , 1. , 1. , 0.99, 1. , 0.99, 0.99, 1. , 1. , 0.99])

print("Accuracy: %0.2f (+/- %0.2f)" % (scores.mean(), scores.std() \* 2))

Accuracy: 1.00 (+/- 0.01)

Question 2:

1. Graph
2. SVM Linear

import numpy as np

from sklearn.svm import SVC

data = np.loadtxt("/Users/jayant/Work/mll/q2\_data.csv", delimiter =",")

data

Out[44]:

array([[-0.88130967, -6.24187192, 0. ],

[ 1.54029478, 4.18513032, 0. ],

[-1.13016968, 4.92378479, 0. ],

...,

[ 0.07013473, -1.79114042, 1. ],

[-1.46368246, -1.05353431, 1. ],

[-0.53926228, 1.03086848, 1. ]])

clf = SVC(kernel = "linear")

from sklearn import cross\_validation

scores = cross\_validation.cross\_val\_score(clf, data[:,:2], data[:,2], cv=10)

scores

Out[9]:

array([ 0.71856287, 0.71856287, 0.71856287, 0.71856287, 0.72 ,

0.72 , 0.71943888, 0.71943888, 0.71943888, 0.71943888])

print("Accuracy: %0.2f (+/- %0.2f)" % (scores.mean(), scores.std() \* 2))

Accuracy: 0.72 (+/- 0.00)

1. SVM RBF (Non Linear)

import numpy as np

from sklearn.svm import SVC

data = np.loadtxt("/Users/jayant/Work/mll/q2\_data.csv", delimiter =",")

data

Out[44]:

array([[-0.88130967, -6.24187192, 0. ],

[ 1.54029478, 4.18513032, 0. ],

[-1.13016968, 4.92378479, 0. ],

...,

[ 0.07013473, -1.79114042, 1. ],

[-1.46368246, -1.05353431, 1. ],

[-0.53926228, 1.03086848, 1. ]])

clf = SVC(kernel = "rbf")

scores = cross\_validation.cross\_val\_score(clf, data[:,:2], data[:,2], cv=10)

scores

Out[86]:

array([ 0.94810379, 0.96407186, 0.94810379, 0.9500998 , 0.964 ,

0.952 , 0.96192385, 0.95591182, 0.95390782, 0.96793587])

print("Accuracy: %0.2f (+/- %0.2f)" % (scores.mean(), scores.std() \* 2))

Accuracy: 0.96 (+/- 0.01)

1. Random Forest Classifier

from sklearn.ensemble import RandomForestClassifier

data

Out[56]:

array([[-0.88130967, -6.24187192, 0. ],

[ 1.54029478, 4.18513032, 0. ],

[-1.13016968, 4.92378479, 0. ],

...,

[ 0.07013473, -1.79114042, 1. ],

[-1.46368246, -1.05353431, 1. ],

[-0.53926228, 1.03086848, 1. ]])

clf = RandomForestClassifier()

scores = cross\_validation.cross\_val\_score(clf, data[:,:2], data[:,2], cv=10)

scores

Out[59]:

array([ 0.97205589, 0.98403194, 0.9740519 , 0.97005988, 0.978 ,

0.978 , 0.98797595, 0.97795591, 0.98396794, 0.97995992])

print("Accuracy: %0.2f (+/- %0.2f)" % (scores.mean(), scores.std() \* 2))

Accuracy: 0.98 (+/- 0.01)