

PRINTED OUTPUT

QUESTION 2.

Question 2(b):

bias term w_0 :

[-6.03039659]

weight vector w :

[[1.64236665 1.74189139]]

Question 2(c):

accuracy1:

0.8586666666666667

accuracy2:

0.8586666666666667

accuracy difference (should be zero):

0.0

Question 2(g):

$P(C = 1 | x) = 0.6$

precision

0.76546037563

recall

0.6684

$P(C = 1 | x) = 0.5$

precision

0.755378774182

recall

0.8356

$P(C = 1 | x) = 0.05$

precision

0.474248316418

recall

1.0

QUESTION 4.

Question 4(a):

Accuracy of Gaussian Bayes clf 4(a):

0.8762

Question 4(c):

Accuracy of Gaussian Bayes clf 4(c):

0.8213333333333334

Question 4(d):

Accuracy of Gaussian Bayes clf 4(d):

0.9007666666666667

Question 4(e):

I don't know

QUESTION 5.

Question 5(b):

C:\tools\Anaconda2\lib\site-packages\sklearn\discriminant_analysis.py:692: UserWarning: Variables are collinear

warnings.warn("Variables are collinear")

Time to fit model 5(b): 9.22499990463

Accuracy of classifier 5(b):

Training: 0.14235

Testing: 0.1419

Question 5(c):

Time to fit model 5(c): 0.427999973297

Accuracy of classifier 5(c):

Training: 0.5649

Testing: 0.5558

Question 5(e):

Time to fit model (full Bayes from 5(b) repeated with noisy data): 10.1540000439

Accuracy of classifier (full Bayes from 5(b) repeated with noisy data):

Training: 0.99335

Testing: 0.9474

Time to fit model (Naive Bayes from 5(c) repeated with noisy data): 0.960999965668

Accuracy of classifier (Naive Bayes from 5(c) repeated with noisy data):

Training: 0.8249333333333333

Testing: 0.8152

Question 5(f):

Time to fit model (full Bayes from 5(e) repeated with only 6000 elements): 0.993999958038

Accuracy of classifier (full Bayes from 5(e) repeated with only 6000 elements):

Training: 1.0

Testing: 0.1787

Time to fit model (Naive Bayes from 5(e) repeated with only 6000 elements): 0.0739998817444

Accuracy of classifier (Naive Bayes from 5(e) repeated with only 6000 elements):

Training: 0.8438333333333333

Testing: 0.819

Question 5(h):

Accuracy of classifier:

Training: 0.8435

Testing: 0.8189

Difference in training and test accuracy vs 5(f):

Diff. train (should be below 0.001): 0.0003333333333333

Diff. test: 0.0001