Machine Learning - Homework 2

Parita Pooj (psp2133)

September 28, 2016

Problem 1

(a) $\mu_{y,j} = \sum_{i=1}^{n} \frac{\mathbb{1}\{y_i = y\} x_{i,j}}{\mathbb{1}\{y_i = y\}}$

(b) Source code: hw2_p1b.m
Training Error Rate: 21.6257%
Test Error Rate: 37.6016%
Running Time: 30 secs

(c) Source Code: hw2_p1c.m
Training Error Rate: 5.7794%
Test Error Rate: 13.1383%
Running Time: 1.15 secs

(d) Source Code: hw2_p1d.m
α_0 = -20.7849
20 most positive words (corresponding to 20 largest values of α_j)
1: 'firearms'
2: 'occupied'
3: 'israelis'
4: 'serdar'
5: 'argic'
6: 'ohanus'
7: 'appressian'
8: 'sahak'
9: 'melkonian'
10: 'villages'
11: 'cramer'
12: 'armenia'
13: 'cpr'
14: 'sdpa'
15: 'handgun'
16: 'optilink'
17:'palestine'

18: 'firearm'
19: 'budget'
20: 'arabs'

20 most negative words (corresponding to 20 smallest values of α_j)
1: 'athos'
2: 'atheism'
3: 'atheists'
4: 'clh'
5: 'teachings'
6: 'revelation'
7: 'testament'
8: 'livesey'
9: 'atheist'
10: 'wpd'
11: 'solntze'
12: 'scriptures'
13: 'theology'
14: 'believers'
15: 'ksand'
16: 'alink'
17: 'benedikt'
18: 'jesus'
19: 'prophet'
20: 'mozumder'

Problem 2

(a) The classifier f^* predicts 1 when $c\frac{2}{3}N(0,1) \leq \frac{1}{3}N(2,\frac{1}{4})$
Thus, the range in which f^* predicts 1 can be given by the roots of:
$c\frac{2}{3}N(0,1) = \frac{1}{3}N(2,\frac{1}{4})$
Thus, we get the quadratic equation
$3x^2 - 16x + 16 + 2log(c) = 0$
which give the roots:

$$x = \frac{16 \pm \sqrt{256 - 12(16 + 2log(c))}}{6}$$

For $1 \leq c \leq 14$,
f^* predicts 1 when x is in the interval $[\frac{16-\sqrt{256+12(16+2log(c))}}{6}, \frac{16+\sqrt{256+12(16+2log(c))}}{6}]$

(b) For $c \geq 15$, the roots are imaginary i.e. they don't exist. Thus, the classifier will always predict 0.

Problem 3

(a) Covariance matrix can be written as:
$\Sigma = U\Lambda U^T$ where U is orthonormal
Thus, $\Sigma + \sigma^2 I = U\Lambda U^T + \sigma^2 U U^T$
$\Sigma + \sigma^2 I = U(\Lambda + \sigma^2 I)U^T$
Thus, the eigenvalues of $\Sigma + \sigma^2 I$ are

$$\lambda_1 + \sigma^2, \lambda_2 + \sigma^2, \ldots, \lambda_d + \sigma^2$$

(b) Similarly, we get the eigenvalues of $(\Sigma + \sigma^2 I)^{-2}$ as

$$(\lambda_1 + \sigma^2)^{-2}, (\lambda_2 + \sigma^2)^{-2}, \ldots, (\lambda_d + \sigma^2)^{-2}$$

References

[1] `https://www.cs.ubc.ca/~murphyk/Teaching/Stat406-Spring08/Lectures/linalg1.pdf`

[2] `http://cs229.stanford.edu/section/gaussians.pdf`

[3] `https://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/least-squares-determinants-and-eigenvalues/diagonalization-and-powers-of-a/MIT18_06SCF11_Ses2.9sum.pdf`