

INTRO TO DATA SCIENCE HW 3

Ben Green

Contents

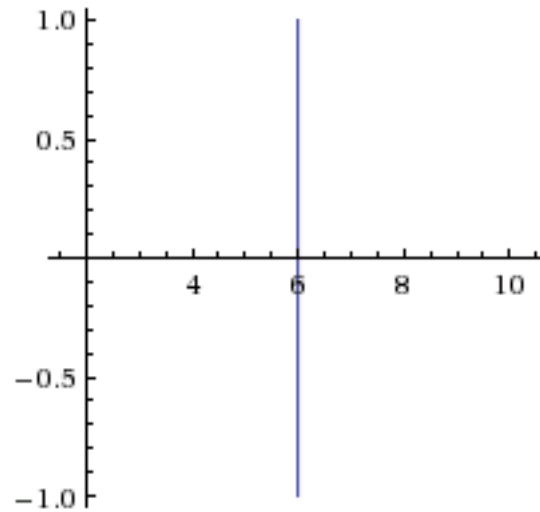
Question 1	1
Question 2	1
Question 3	1
3a	1
3b	1
Question 4	2
4a	2
4b	2
Question 5	2
5a	2
5b	2
5c	2
Question 6	2
Question 7	2
Question 8	2
Question 9	3
9a	3
9b	3
Question 10	3
Question 11	3
11a	3
11b	3
Question 12	3
Question 13	3
13a	3
13b	3
13c	4
Question 14	4

Question 15

4

April 13, 2015

Question 1



where x_1 is the horizontal axis, x_2 is the vertical axis, objects to the right of the line are + and objects to the left are -.

Question 2

x - 14×4

θ - 4×1

y - 4×1

Question 3

3a

Feature scaling is necessary because the classifier (midterm exam)²'s range of data value varies widely. One strategy that we can use is rescaling.

3b

Decrease the learning rate until we have $j\theta$ decrease over time.

Question 4

4a

Boosting can fail to perform well given overly complex base classifiers or base classifiers that are too weak.

4b

Boosting weak base classifiers (or vice versa if necessary) can provide a significant improvement in performance. Brownboosting also deemphasizes outliers which may help.

Question 5

5a

The SVM is making an underfit which means the classifier has high bias and low variance.

5b

C should be increased.

5c

σ^2 should be decreased.

Question 6

Bagging can reduce predictive accuracy and is computationally hard.

Question 7

Naïve Bayesian classification is called naïve because it assumes class conditional independence. That is, the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is made to reduce computational costs, and hence is considered naïve.

Question 8

Accuracy - $95/1000 = .095$

Recall - $85/100 = .85$

Precision - $85/975 = .087$

F1-score - $2 * \text{product} / \text{sum of precision \& recall} = .158$

Question 9

9a

The accuracy would be .99. This would be a bad fraud detection system because fraud is not actually ever being detected.

9b

Recall - 1

Precision - .01

Question 10

Led7 is the specified point.

The Z-statistic is .16.

Question 11

11a

Minimize the Schwarz criterion.

11b

This results in a low amount of distortion

Question 12

Spectral clustering is optimal because it isn't NP complete.

Question 13

13a

The SSE for cluster C1 is 36.

13b

The optimal partitioning splits C1 into two groups - one with (0,8) and one with (6,5) and (3,2)

13c

The new SSE is 9, so the reduction is 27.

Question 14

Point 3 fits these criteria.

Question 15

B, C, F, B