B529: Homework 2

Nathan Byers

Friday, February 24, 2015

Question 1

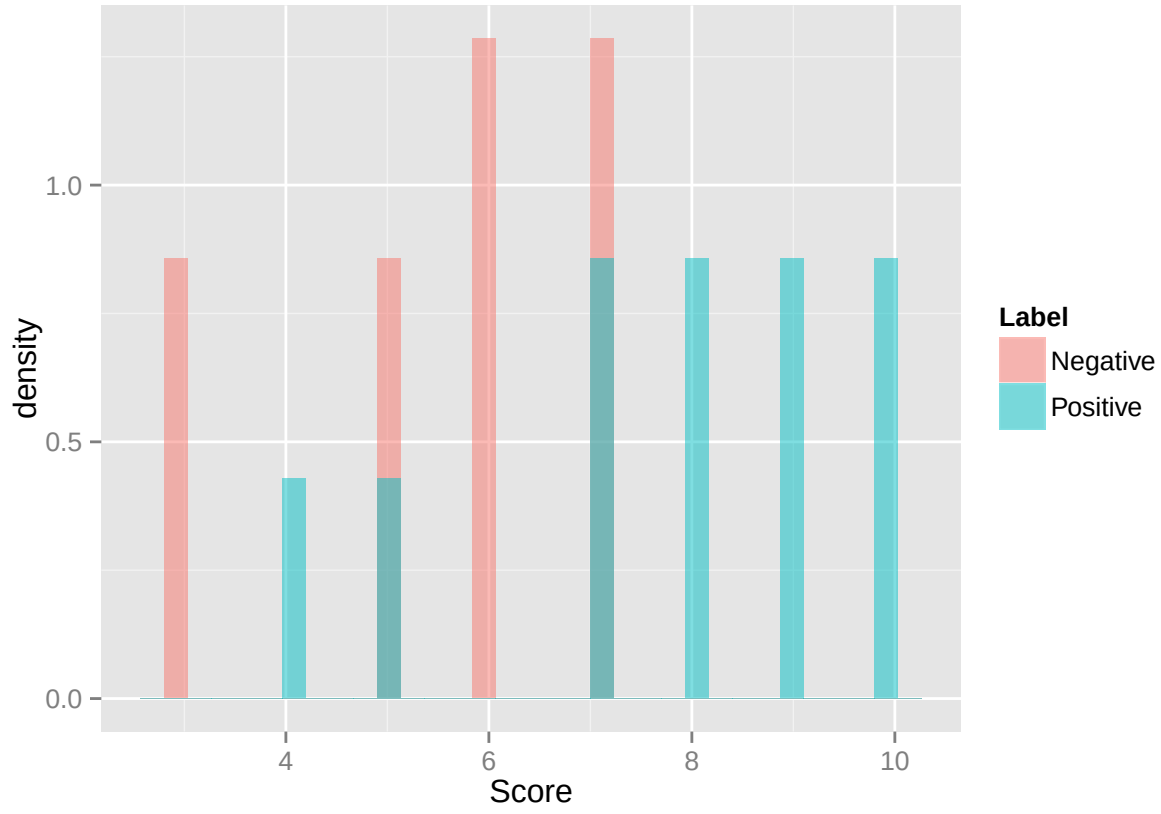
A scoring function (classifier) assigned scores for 10 positive data points and 10 negative ones as follows:

No.	Label	Score
1	Positive	10
2	Positive	10
3	Positive	9
4	Positive	9
5	Positive	8
6	Positive	8
7	Negative	7
8	Positive	7
9	Negative	7
10	Negative	7
11	Positive	7
12	Negative	6
13	Negative	6
14	Negative	6
15	Positive	5
16	Negative	5
17	Negative	5
18	Positive	4
19	Negative	3
20	Negative	3

Compute FPR and TPR for each score threshold in 3, 4, 5, 6, 7, 8, 9, 10, and plot the ROC curve of the scoring function (25 points).

Answer 1

Below is a histogram for the assigned scores.



Looking at the densities, it appears that Negative has lower scores than Positive. If we choose 3 as the cutoff (≤ 3 is Negative, > 3 is Positive), then we get the table below (the column headers indicate the target function and the first column indicates the outcome function).

	+1	-1
+1	2	0
-1	8	10

The false positive rate is $\frac{0}{0+10} = 0$ and the true positive rate is $\frac{2}{2+10} = 0.2$.

If 4 is the cutoff:

	+1	-1
+1	2	1
-1	8	9

$FPR = \frac{0}{0+10} = 0$ and $TPR = \frac{2}{2+10} = 0.2$.

For all of the cutoffs, we get the following table and ROC curve:

Score	TP	P	TPR	FP	N	FPR
≤ 10	10	10	1	10	10	1
≤ 9	10	10	1	8	10	0.8
≤ 8	10	10	1	6	10	0.6
≤ 7	10	10	1	2	10	0.2
≤ 6	7	10	0.7	2	10	0.2
≤ 5	4	10	0.4	2	10	0.2
≤ 4	2	10	0.2	1	10	0.1
≤ 3	2	10	0.2	0	10	0