Solution 1:

a) First, sort the table:

ID	Actual Class	Score	Predicted Class
6	0	0.63	1
7	1	0.62	1
10	0	0.57	1
4	1	0.38	0
1	0	0.33	0
8	1	0.33	0
2	0	0.27	0
5	1	0.17	0
9	0	0.15	0
3	1	0.11	0

	Actual Class - 0	Actual Class - 1
Prediction - 0	3	4
Prediction - 1	2	1

so we get

FN	FP	TN	TP
4	2	3	1

b)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{1}{3}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{1}{5}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{4}{10}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{3}{5}$$

$$\text{Error Rate} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{6}{10}$$

$$\text{F-measure} = \frac{2 \cdot \text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} = 0.25$$

$$\text{Negative Predictive Value} = \frac{\text{TN}}{\text{TN} + \text{FN}} = \frac{3}{7}$$

c) First we sort the results by the score:

	true_labels	scores
6	0	0.63
7	1	0.62
10	0	0.57
4	1	0.38
1	0	0.33
8	1	0.33
2	0	0.27
5	1	0.17
9	0	0.15
3	1	0.10

Here we see that $\frac{1}{n_+} = \frac{1}{5} = 0.2$ and $\frac{1}{n_-} = \frac{1}{5} = 0.2$. Now we follow the algorithm as described in the lecture slides:

(i) Set $\alpha = 1$, so we start in $(0, 0)$; we predict everything as 1.

(ii) Set threshold $\tau = 0.625$ yields TPR 0 and FPR $0 + \frac{1}{n_-} = 0.2$. (Obs. 6 is "0")

(iii) Set threshold $\tau = 0.6$ yields TPR $0 + \frac{1}{n_+} = 0.2$ and FPR 0.2. (Obs. 7 is "1")

(iv) Set threshold $\tau = 0.5$ yields TPR 0.2 and FPR $0.2 + \frac{1}{n_-} = 0.4$. (Obs. 10 is "0")

(v) Set threshold $\tau = 0.35$ yields TPR $0.2 + \frac{1}{n_+} = 0.4$ and FPR 0.4. (Obs. 4 is "1")

(vi) Set threshold $\tau = 0.3$ yields TPR $0.4 + \frac{1}{n_+} = 0.6$ and FPR $0.4 + \frac{1}{n_-} = 0.6$. (Obs. 1/8 is "0"/"1")

(vii) Set threshold $\tau = 0.2$ yields TPR 0.6 and FPR $0.6 + \frac{1}{n_-} = 0.8$. (Obs. 2 is "0")

(viii) Set threshold $\tau = 0.16$ yields TPR $0.6 + \frac{1}{n_+} = 0.8$ and FPR 0.8. (Obs. 5 is "1")

(ix) Set threshold $\tau = 0.14$ yields TPR 0.8 and FPR $0.8 + \frac{1}{n_-} = 1$. (Obs. 9 is "0")

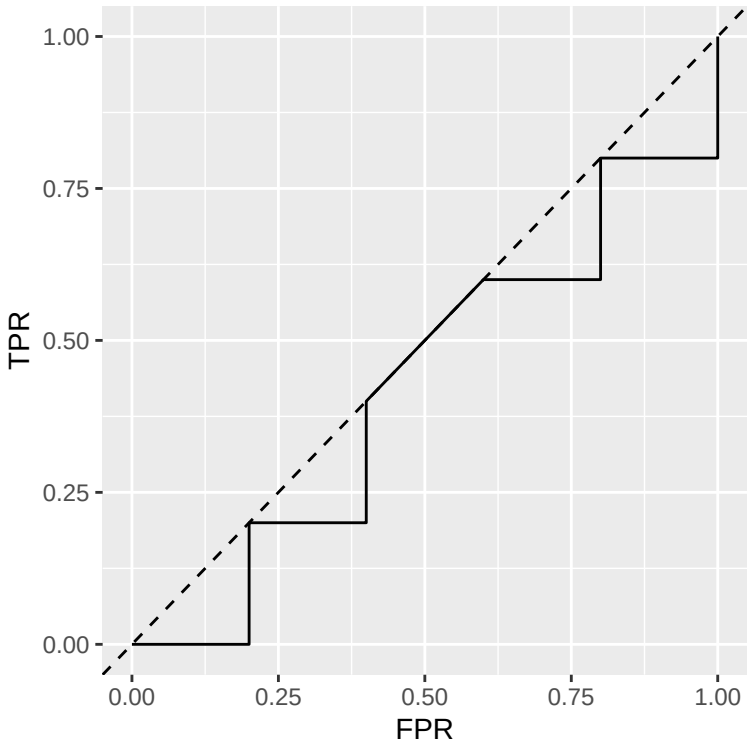
(x) Set threshold $\tau = 0.09$ yields TPR $0.8 + \frac{1}{n_+} = 1$ and FPR 1. (Obs. 3 is "1")

Therefore we get the polygonal path consisting of the ordered list of vertices

$$(0, 0), (0.2, 0), (0.2, 0.2), (0.4, 0.2), (0.4, 0.4), (0.6, 0.6), (0.8, 0.6), (0.8, 0.8), (1, 0.8), (1, 1).$$

```r
library(ggplot2)
roc_data <- data.frame(TPR = c(0, 0,   0.2, 0.2, 0.4, 0.6, 0.6, 0.8, 0.8, 1),
                       FPR = c(0, 0.2, 0.2, 0.4, 0.4, 0.6, 0.8, 0.8,  1,  1))

ggplot(roc_data, aes(x = FPR, y = TPR)) + geom_line() +
  geom_abline(slope = 1, intercept = 0, linetype = 'dashed')
```



We see that the resulting ROC lies below the line from the origin with a slope of 1, which represents a random classifier, i.e., the scoring algorithm performs worse than a random classifier. If this happens while evaluating the training data, the labels of the scoring algorithm should be inverted.

d) We can compute the AUC (*area under the curve*) by looking at the ROC, s.t.

$$AUC = 0.5 - 4 \cdot (0.2 \cdot 0.2 \cdot 0.5) = 0.42.$$