

Thomas Dinh

Sept. 29, 2023

DSE 6211

Week 5

Exercises

1) In the ROC curve above, what is the TPR and FPR associated with the threshold value of 0.3?

The fpr is 0.4917 and tpr is 0.9035 for my ROC curve.

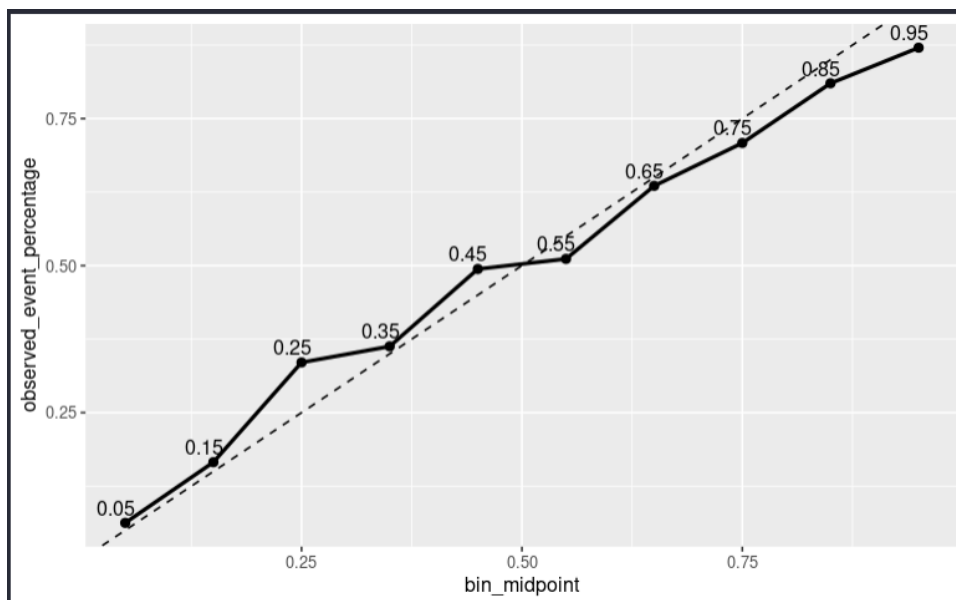
```
```{r}
over_threshold <- test_set[test_set$p_prob >= 0.3,]
fpr <- sum(over_threshold$lodgepole_pine==0)/sum(test_set$lodgepole_pine==0)
fpr

tpr <- sum(over_threshold$lodgepole_pine==1)/sum(test_set$lodgepole_pine==1)
tpr
```

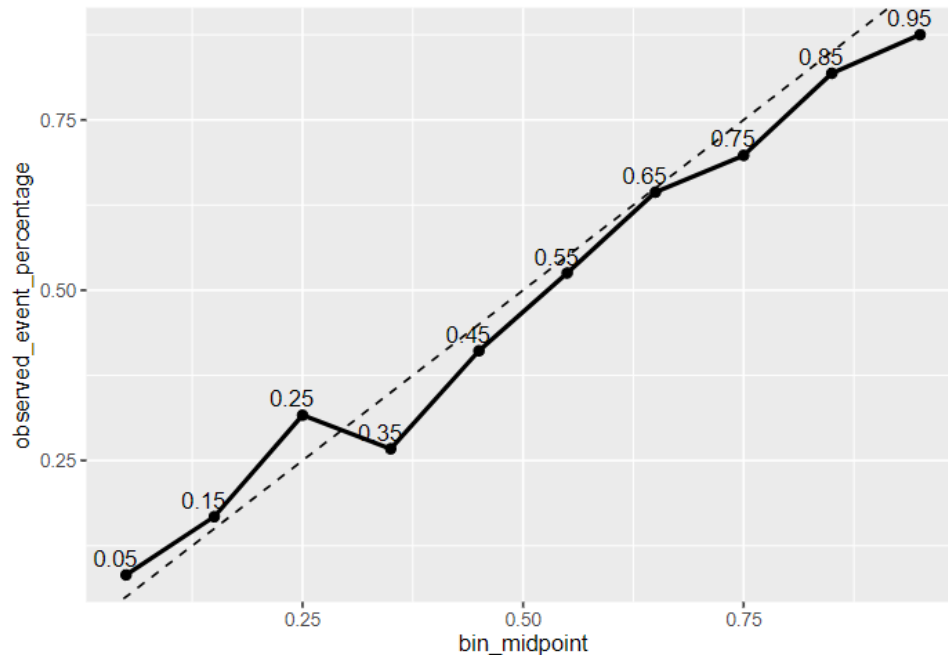
[1] 0.4917431
[1] 0.9035813
```

2) In the calibration curve above, are the predicted probabilities in the interval (0.2, 0.3) under-confident or over-confident?

In my model (0.2, 0.3) is under-confident



In the example model (0.2, 0.3) is also under-confident.



3) The 'AppliedPredictiveModeling' R package contains several datasets. One such dataset is the 'logistic-CreditPredictions' dataframe, which contains the predictions and predicted probabilities for a credit dataset containing a binary target variable with the classes 'Good' and 'Bad'. The positive class is the 'Bad' class, since we are trying to identify customers with bad credit. The 'logisticCreditPredictions' dataframe has 4 columns: the columns 'Bad' and 'Good' contain the predicted probabilities of class membership, the column 'pred' contains the predicted class using the threshold 0.5, and the column 'obs' contains the actual class. Use the code below to plot an ROC curve and calibration curve for the predicted probabilities. To do this, fill in the question marks with the appropriate column names and values. Copy and paste the ROC curve and calibration curve.

