LogitModel.RMD

Bart Ciastkowski

12/2/2019

# Slide 1: Determining whether wine good value or not

## Logit Model

### logit\_mod <- glm( well\_priced ~ …… )

## How is Well\_Priced determined?

### Well\_priced == whether we think wine is well priced

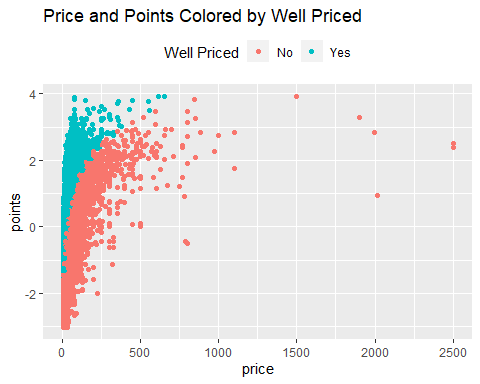
### Well\_priced == f(median log(price) to points ratio)

## In Summary: Well\_Priced takes into consideration diminishing returns i.e. marginal increase in points is accompanied by a higher and higher increase in price

# Slide 2: Points vs Price by Well Priced

## Purpose: showing the effects of dimishing returns

PointsVsPricePlot



# Slide 3: Let’s create the model

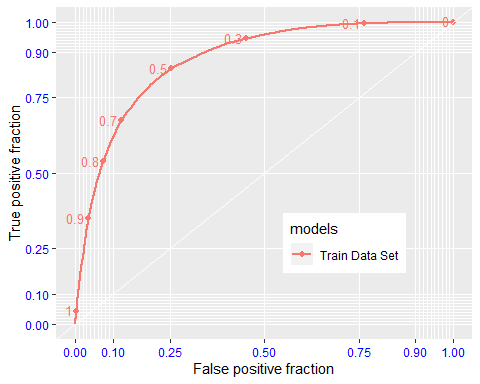
## Notes: Price removed from the dataset as well as variety\_and\_color as it provided no added value (i.e. NA values)

logit\_mod <- glm( well\_priced ~ .,  
 data = wine\_train\_logit %>%   
 select (  
 -variety\_and\_color,   
 ),  
 family = binomial) #our varaible can be 0 or 1, a binomial

# Slide 4: ROC Curves and AUC – Train vs Test

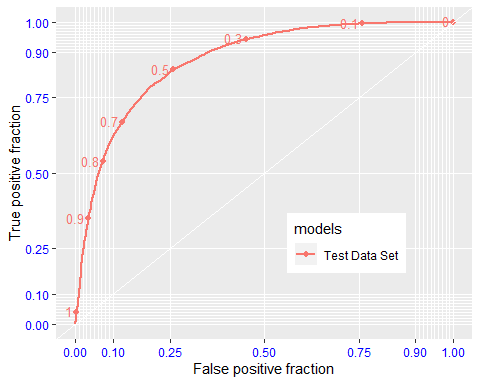
### ROC Curve train  
plot(TrainROC)

## Warning in verify\_d(data$d): D not labeled 0/1, assuming 1 = 0 and 2 = 1!



### ROC Curve test  
plot(TestROC)

## Warning in verify\_d(data$d): D not labeled 0/1, assuming 1 = 0 and 2 = 1!



### Area under the curve  
AUC\_results

## TrainAUC.PANEL TrainAUC.group TrainAUC.AUC TestAUC.PANEL TestAUC.group  
## 1 1 1 0.8748211 1 1  
## TestAUC.AUC  
## 1 0.8724196

# Slide 5: Logit Model Summary

## Both train and test set curves are above the diagonal chance-only line

## This means that determining whether wine is a “good value” is better than chance

## AUC values are high, above 80% or about 70% higher than chance

## AUC for train and test are nearly identical (0.868 vs 0.867) hence model neither over- or underfit

## Points, Specific Province, Specific Variety, Specific Winery, followed by Specific Taster explain the model best whereas Title information not so much