

# Homework 3

Rufus Petrie

Professors Magnolfi and Sullivan

Machine Learning

26 October, 2018

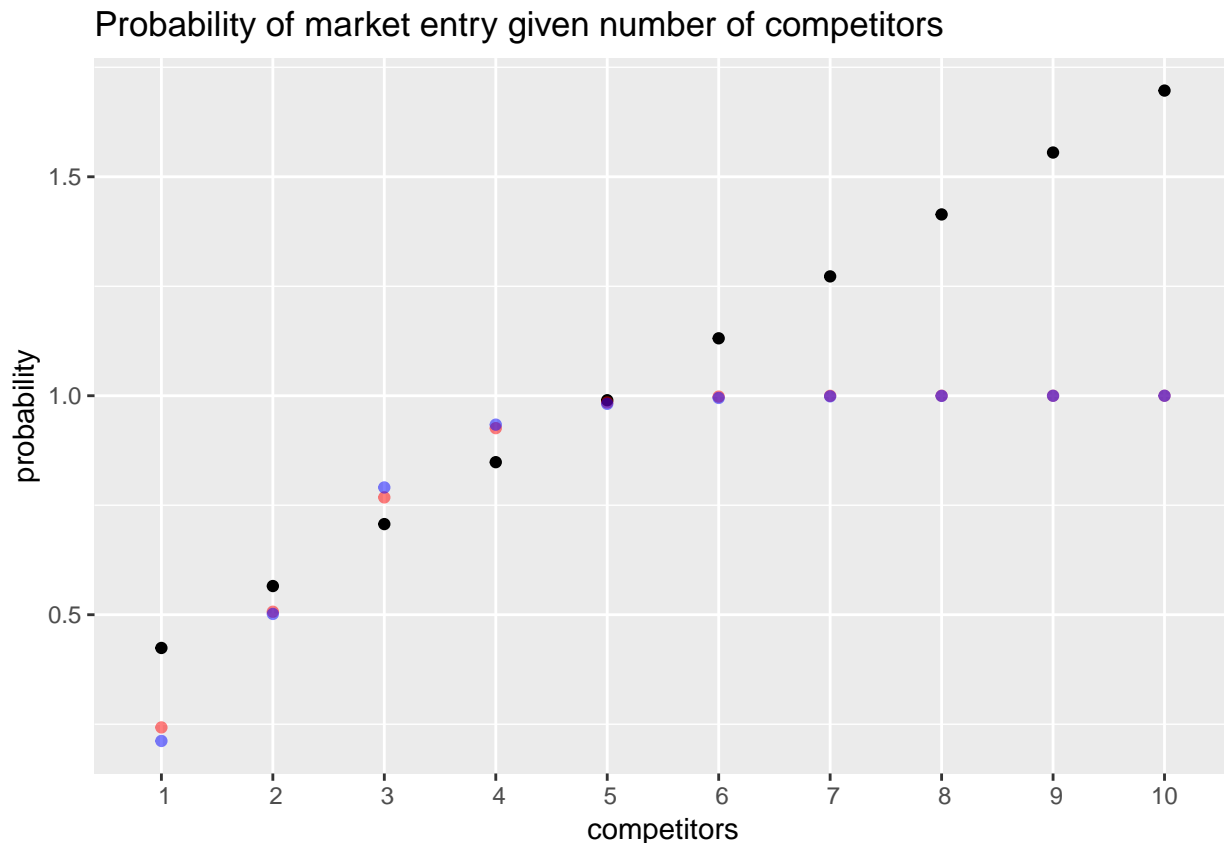
Note: I silenced the code to generate the data and regressions for this problem set. They are still visible in the included RMD and script file.

## Question 4

Below are the conditional probabilities of entering a market given a certain number of competitors observed in the training data. Note that American Airlines has no monopolies, so the probability estimates start from one competitor.

competitors	probability
1	0.1111111
2	0.5608466
3	0.8270469
4	0.9399885
5	0.9406675
6	0.9866221
7	1.0000000
8	1.0000000
9	1.0000000
10	1.0000000

## Question 5



In the graph above, the black dots are the linear model, the red dots are the probit model, and the blue dots are the logistic model. All three have fairly similar behavior. The probit and logistic models both have similar intercepts at around 0.2, but the linear model has a much higher intercept at about 0.4. Afterwards, all three estimates stay somewhat close to each other until after 5 competitors, at which point the probit and logistic models reach a horizontal asymptote, but the linear model estimates continue to increase. Notice that the blue logistic predictions begin underneath the probit predictions but then become higher. This makes sense because the logistic distribution has fatter tails than the normal distribution, so we should expect higher probabilities for extreme values ex ante.

We probably shouldn't interpret these coefficients causally. Because we only include the number of competitors as an estimator, there are a ton of other factors correlated with market size that could explain AA's entry decision. For instance, the profitability of a market is highly likely to induce AA's entry, but profitability is also likely correlated with the number of flights, competitors, etc.

## Question 6

```
## [1] "The optimal lambda value equals: 0.000281199085127456"
```

## Question 7

Sum of Squared Errors	
Linear	103.79419
Probit	83.56580
Logistic	82.01475
Bayesian	79.95883
Logistic Lasso	77.82641

From the table, we see that the linear model had the worst performance, followed by the probit, the logic, the bayesian classifier, and then finally the logistic lasso. Note that the bayesian classifier is produced by making predictions off of the observed probability in the training data. This implies that American Airlines probably has a highly nonlinear decision condition for entering a market. The probit, logic, and linear models all have a linear functional form, but the bayesian classifier (with no functional assumptions) and the logistic lasso (with a polynomial functional form) are both more flexible and perform better. However, we should note that although the bayesian and logistic lasso models perform better in terms of predicted probabilities, all models have similar performance in terms of predictions because they all basically assume that AA will enter a market if there are more than two competitors.