HW3

February 22, 2024

Problem 3.1 (Competitive Auctions on eBay.com)

The file `eBayAuctions.csv` contains information on 1972 auctions transacted on eBay.com during May–June 2004. The goal is to use these data to build a model that will distinguish competitive auctions from noncompetitive ones. A competitive auction is defined as an auction with at least two bids placed on the item being auctioned. The data include variables that describe the item (auction category), the seller (his or her eBay rating), and the auction terms that the seller selected (auction duration, opening price, currency, day of week of auction close). In addition, we have the price at which the auction closed. The goal is to predict whether or not an auction of interest will be competitive.

Data preprocessing. Create dummy variables for the categorical predictors. These include `Category` (18 categories), `currency` (USD, GBP, Euro), `endDay` (Monday–Sunday), and `Duration` (1, 3, 5, 7, or 10 days).

a. Create pivot tables for the mean of the binary outcome (`Competitive?`) as a function of the various categorical variables (use the original variables, not the dummies). Use the information in the tables to reduce the number of dummies that will be used in the model. For example, categories that appear most similar with respect to the distribution of competitive auctions could be combined.

b. Split the data into training (60%) and validation (40%) datasets. Run a logistic model with all predictors/features.

c. If we want to predict at the start of an auction whether it will be competitive, we cannot use the information on the closing price. Run a logistic model with all predictors as above, excluding price. How does this model compare to the full model with respect to predictive accuracy?

d. Interpret the meaning of the coefficient for closing price. Does closing price have a practical significance? Is it statistically significant for predicting competitivenessof auctions? (use a 10% significance level)

e. Use regularized logistic regression with L_1 penalty on the training data. Compare its selected predictors and classification performance to the unregularized models.

f. Based on these data, what auction settings set by the seller (duration, opening price, ending day, currency) would you recommend as being most likely to lead to a competitive auction?

```python
import pandas as pd
df = pd.read_csv('../data/eBayAuctions.csv')
df.sample(10)
```

```
[ ]:               Category currency  sellerRating  Duration endDay  \
     719          Home/Garden      GBP            53         7    Thu
     1713         Home/Garden       US          1304         7    Sun
     487         Toys/Hobbies       US          3385         7    Sat
     503          Collectibles     EUR           578        10    Mon
     434     Music/Movie/Game       US           841         5    Mon
     1636          Automotive       US         10067         3    Fri
     459          Collectibles      US           510         3    Sun
     224    Clothing/Accessories   EUR           279        10    Tue
     274                Books      EUR          2163        10    Tue
     1879         Collectibles      US          6945         7    Mon

           ClosePrice  OpenPrice  Competitive?
     719        17.40       3.55             1
     1713       23.50      23.50             0
     487         2.75       2.75             0
     503         4.93       2.45             1
     434         5.50       0.99             1
     1636       19.50      19.00             1
     459         2.50       2.50             0
     224         1.23       1.23             0
     274         1.84       1.23             1
     1879      392.51       9.99             1
```