## **APPLIED BUSINESS ANALYTICS**



## 2024-25, Spring Semester

## Exercises to Practice - Week 9

- 1. (Conceptual Question) What are the odds in logistic regression?
- 2. (Conceptual Question) How do we handle categorical variables in Logistic Regression?
- 3. (Conceptual Question) Can we solve the multiclass classification problems using Logistic Regression? If Yes then How?
- 4. (Conceptual Question) What is the main purpose of logistic regression, and how does it differ from the linear regression.
- 5. (Conceptual Question) Suppose that a term in a logistic regression equation is 0.687\*MallTrips. Explain, exactly what this means.
- 6. (Conceptual Question) What is the role of the Activation functions in Neural Networks?
- 7. (Conceptual Question) How to initialize Weights and Biases in Neural Networks?
- 8. (Conceptual Question) Explain the different types of Gradient Descent (Stochastic/Mini-batch/Batch Gradient Descent) in detail.
- 9. (Conceptual Question) How does the learning rate affect the training of the Neural Network?
- 10. (Conceptual Question) What do you mean by hyperparameters?
- 11. 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 with a cutoff of 0.5.
- 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?

- Interpret the meaning of the coefficient for closing price. Does closing price have a practical significance? Is it statistically significant for predicting competitiveness of auctions? (Use a 10% significance level.)
- e. Use stepwise regression as described in Section 6.4 to find the model with the best fit to the training data (highest accuracy). Which predictors are used?
- f. Use stepwise regression to find the model with the highest accuracy on the validation data. Which predictors are used?
- g. What is the danger of using the best predictive model that you found?
- h. Explain how and why the best-fitting model and the best predictive models are the same or different.
- i. Use regularized logistic regression with L1 penalty on the training data. Compare its selected predictors and classification performance to the best-fitting and best predictive models.
- j. If the major objective is accurate classification, what cutoff value should be used?
- k. 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?