Exam 1

Buan 6340

Spring 2019

Answer all three questions. Submit your code on the eLearning page.

Question 1

Analyze the data in MROZ.csv. The data contains information on married women in the US in 1976. The variable of interest is the first variable, `inlf`, which is an indicator for being in the labor force. A complete answer will include summary statistics, groupby statements, and some sort of models (linear or otherwise) for the target variable.

Question 2

Following the sample jupyter notebook, extend the functionality of the `logit` class we developed to include the following alternative R^2 statistics:

$$\begin{aligned}
R^2_{BL} &= \frac{1}{n}\sum_{i=1}^{n} y_i\hat{F}_i + (1-y_i)\left(1-\hat{F}_i\right) \\
R^2_{Ef} &= 1 - \frac{\sum_{i=1}^{n}\left(y_i-\hat{F}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2} \\
R^2_{VZ} &= \frac{\delta-1}{\delta - R^2_{McF}} R^2_{McF} \\
R^2_{MZ} &= \frac{\sum_{i=1}^{n}\left(x_i'\hat{\beta}-\bar{x}'\hat{\beta}\right)^2}{n+\sum_{i=1}^{n}\left(x_i'\hat{\beta}-\bar{x}'\hat{\beta}\right)^2}
\end{aligned}$$

where

$$\delta = \frac{n}{2\ln L_0}$$

and where R^2_{McF} is the McFadden's R^2 already in the code.

Question 3

Extend the `logit` class by making a `logitclassifier` class. This class takes a probability *prob* and classifies the data by comparing the predicted logit probability to this *prob* value. Make this *prob* value default to 0.5 so that we can use this class in our `kfold` function. Give this class functionality for mean-squared prediction error (hint: change the predict function), accuracy, and F_1 scores.