Question 14.1

The breast cancer data set breast-cancer-wisconsin.data.txt from
http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/ (description at
http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29) has missing values.
1. Use the mean/mode imputation method to impute values for the missing data.
2. Use regression to impute values for the missing data.
3. Use regression with perturbation to impute values for the missing data.
4. (Optional) Compare the results and quality of classification models (e.g., SVM, KNN) build
   using
       (1) the data sets from questions 1,2,3;
       (2) the data that remains after data points with missing values are removed; and (3) the
           data set when a binary variable is introduced to indicate missing values.

Question 15.1

Describe a situation or problem from your job, everyday life, current events, etc., for which optimization
would be appropriate. What data would you need?