DATS 6202 | **Machine Learning I** | **Quiz 1**
Term 2018-Fall | **September 18, 2018**

Quiz 1: Solutions

Full Name: _____

GWID: _____

- DATS 6202, Instructor: Yuxiao Huang

Material Covered

- Data preprocessing
- Linear regression

Note

- The quiz has 100 points.
- The quiz period is 20 minutes.
- The quiz is closed book and closed notes.
- The quiz is closed electronics (e.g., no laptops, netbooks, OLPCs, tablets, iPads, calculators, cellular phones, iPhones, Nexi, iPods, Zunes, Kindles, Nooks).
- There is only one correct answer for each `Multiple Choice Question`.
- For each `Calculation` question (if there is any), you must show the essential steps. **No mark will be given if only the result is provided**.

1 Multiple Choice Questions (40 points)

1. When imputing the missing values in a dataframe (where the rows are the samples and column variables), which of the following claim is correct?

   (a) the mean (or mode) of the row should be used for imputing the missing value

   (b) the mean (or mode) of the column should be used for imputing the missing value

   (c) the mean (or mode) of the dataframe should be used for imputing the missing value

   b

2. In data preprocessing, which is the correct order for splitting the data (into training and testing) and standardization?

   (a) split first, standardize second

   (b) standardize first, split second

   (c) both of the above orders are correct

   a

3. Which are the functions that should be used for standardizing the training and testing data?

   (a) `fit` for training, `transform` for testing

   (b) `fit_transform` for training, `transform` for testing

   (c) `fit_transform` for training, `fit_transform` for testing

   (d) `fit_transform` for training, `fit` for testing

   b

4. Which of the following claim is correct?

   (a) One-hot-encoding should be used when encoding categorical target

   (b) One-hot-encoding should be used when encoding ordinal features

   (c) One-hot-encoding should be used when encoding nominal features

   c

2 Description (60 points)

The linear regression model can be written as

$$\hat{y} = \sum_{j=1}^{d} \mathrm{w}_j \cdot \mathbf{x}_j, \tag{1}$$

where,

- $\hat{y}$ is the predicted value of the target
- d is the number of features
- $\mathbf{x}_j$ is the jth feature
- w_j is the weight of $\mathbf{x}_j$

The rule for updating w_j (where $1 \leq j \leq d$) can be written as

$$\mathrm{w}_j = \mathrm{w}_j + \Delta\mathrm{w}_j \quad \text{where} \quad \Delta\mathrm{w}_j = \eta \cdot (y - \hat{y}) \cdot \mathbf{x}_j. \tag{2}$$

Here,

- $\Delta\mathrm{w}_j$ is the update of w_j
- η is the learning rate
- y is the target
- $y - \hat{y}$ is the error

1. Explain why η cannot be zero.

   If η were zero, $\Delta\mathrm{w}_j$ would always be zero. As a result, w_j would not be updated.

2. Explain why η cannot be negative. You should demonstrate that, if η were negative then the updating rule would increase (rather than decrease) the error. You should rely on the following assumption when making your argument.

   - $y - \hat{y} > 0$ (i.e., the error is positive)
   - $\mathbf{x}_j > 0$ (i.e., the feature is always positive)

   If η were negative, $\Delta\mathrm{w}_j$ would be negative (since the assumption says that both the error and the feature is positive). In turn, w_j would be decreased (based on eq. (2)), $\hat{y}$ would be decreased (based on eq. (1)), and finally the error would be increased.

THIS PAGE INTENTIONALLY LEFT BLANK
(You may use it as scratch paper, but *do* submit it as part of your completed exam.)