CST 383: Intro to Data Science

Dr. Glenn Bruns

# Reading: Linear Regression

**Instructions**. Read the section 'In Depth: Linear Regression' in the Python Data Science Handbook up to, but not including, the subsection on 'Regularization'. This involves only a few pages of reading. Answer the questions below by downloading and editing [lin-regr.txt](https://drive.google.com/file/d/1R6-QdTGFpVARCHvvpKR4-IhMXaC9-dzO/view?usp=sharing). Some of these questions are based on class lectures.

1. Suppose I have a linear model that uses 3 predictors. During the training of my model, how many model parameters will be set?
   1. 2
   2. 3
   3. 4
2. What kind of machine learning does linear regression represent?
   1. supervised learning
   2. unsupervised learning
3. "When using linear regression, you will obtain a lower RMSE on test data if you scale your predictors first." Is this statement:
   1. always true
   2. sometimes true (depending on the data)
   3. never true
4. Which takes longer to train, KNN regression or linear regression? (Consider time to perform arithmetic operations only.)
   1. KNN regression
   2. linear regression
   3. it depends on the number of rows of training data
   4. it depends on the number of predictors
5. I have a linear model that predicts the cost of tuition at a university based on the number of students at the university. My model predicts a yearly tuition of $22,000 dollars if there are 12,000 students at a university, and $35,000 dollars if there are 2,000 students at a university. According to this model, what would my model predict for a university with 6,500 students?
6. Suppose I want to fit a linear model to some training data. Which parts of the training data do I need?
   1. predictor values only
   2. target values only
   3. both predictor and target values
7. With linear regression, which quantity are we attempting to minimize during the training process?
   1. the average of the squared differences between predicted target values and actual target values
   2. the average of the absolute value of the differences between predicted target values and actual target values
   3. the average of the differences between the predicted target values and actual target values
8. Here is a model that estimates the monthly rental price of an office in Cork, Ireland (in dollars):

The units for size are square feet, and the units for internetSpeed is Mb/sec. What is the predicted rental price for a 850 square foot office on the 15th floor of a building with an internet speed of 100 Mb/sec?

1. Referring to the model of the last question, what would be the difference in price if the office were on the 18th floor instead of the 15th floor? In other words, what is the new price minus the old price? Be sure to provide the minus sign if needed.
2. (Yes/No) Referring again the office model, would the price change associated with a change from the 15th to the 18th floor be the same as the price change associated with a change from the 18th to the 21st floor?

**Submission**. Submit your edited lin-regr.txt on iLearn.

**Grading**. 10 points/question.