Handout 02: Simple Linear Regression

Last time, I said that I would be using note from my textbook this semester to teach the course. Today is one exception because I think it is useful to start with a bit of review that is not included in the text. For this, I think it is best to just jump right into the notes rather than explaining this to you in detail. Hopefully the material today will come quickly. If this seems overwhelming, today would be a good time to talk with me about your background and the course going forward.

LAB QUESTIONS

1. To start, download and open the `class02.Rmd` file in RStudio. Follow the script until you get to the section that asks you to return to these notes.

2. Last time we started with the basic idea of statistical learning. We observe pairs (x_i, y_i) and want to construct a function $f(x)$ from this training data that does a good job of predicting future values of y_i given new values of x_i. One of the simplest such models for predicting a continuous response y is simple linear regression. Visually this corresponds to fitting a linear function f to the data such that:

$$f(x_i) = a + b \cdot x_i. \tag{2.1}$$

   Where the parameters a (the intercept) and b (the intercept) are *learned* from the data. Write down, symbolically, what the mean squared loss function is of using the above f to predict the values y_i.

3. We are going to simplify things further by removing the intercept term a from the model and assuming that we have only:

$$f(x_i) = b \cdot x_i. \tag{2.2}$$

   Taking the equation you had from the previous question, write down the loss function for the new value of f. Take the derivative with respect to b and set it equal to zero. Can you find a formula for b that minimizes the loss function?

4. Taking the second derivative of the loss function, prove that you found a global minimizer in the previous question rather than a saddle point or maximum.

5. We typically write the learned parameters in a model with a 'hat'. So the slope you computed above becomes $\widehat{b}$. Can you re-write $\widehat{b}$ such that the estimator is written a weighted sum of the values y_i?

6. So far, we have made no assumptions about the 'true' nature of the relationship between x and y. Assume that we can write:

$$y_i = b \cdot x_i + \epsilon_i \tag{2.3}$$

   For some term ϵ_i known as the *error term*. Plugging this into your equation for $\widehat{b}$, can you argue that $\widehat{b}$ will be close to b if the error terms are small?

7. Return to the R code to complete today's lab.