Homework 3 - Due 2/8/2017

Matthew Pancia

1. Do the following course: Datacamp R course
2. Include a PDF in your homework folder named `homework_3_username_datacamp.pdf` that proves your completion in some way.

Problems

3. Let X be a 1-dimensional feature space, with associated training set $\{(x_i, y_i)\}$. Suppose that we try to fit a linear hypothesis to this training set of the form

$$f(x) = \beta \cdot (1, x) = \beta_0 + \beta_1 x \quad \text{for } \beta = (\beta_0, \beta_1) \in \mathbb{R}^2$$

.

Using calculus, show that the minimizer of the loss function $L = \frac{1}{2}\sum_i (\beta \cdot (1, x_i) - y_i)^2$ is given by

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Where $\bar{y}$ denotes taking the average.

4. Using the formula from 3, write a function that performs linear regression for 1-dimensional feature spaces. This should be a function

```
1_dim_lm <- function(x, y){
...
}
```

That returns a named list of:

1. Fitted values (a numeric vector the same length of x);
2. Coefficient estimates (a named vector with names `beta_0`, `beta_1`);
3. Residual errors (the differences (y - predicted values).
4. Use the `lm` function on the iris data set to verify that your function from 4 produces the same coefficient estimates.
5. Problem 3.7.14 in *Introduction to Statistical Learning*.