# Lecture 13: Linear Regression II

COSC 480 Data Science, Spring 2017

Michael Hay

# Exercise

**Instructions:** ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

- When moving from simple linear regression, what needs to change? For each one, explain how it changes.
  - Hypothesis function $h_\beta(x\_i)$? [Answer is **yes** for this one!]
  - Cost function $J(\beta_0,\beta_1)$?
  - Gradient $\nabla J$?
  - Gradient descent?

# Exercise

**Instructions:** ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

- (Example credit card dataset described on board)
- Continuing with the credit card example… suppose we instead used the encoding $x_i=1$ if $i^{th}$ person is female and -1 if $i^{th}$ person is male
- How would you interpret coefficients? What values would $\beta_0$ and $\beta_1$ take on?

# Exercise

**Instructions:** ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

For each of the following scenarios, suppose the real-world operated as described. For each, if we obtain a random sample of data and fit a multiple linear regression model with a coefficient for *each predictor variable*, can we accurately model the data?

A. Males have *higher balances* than females

B. Females have *higher balances* and balances *increase* with age

C. Males have *higher balances*, balance *increases* with age, but *decreases* with "number of jobs" (that the person works)

D. Balance *increases* with age, more rapidly for males than females

# Exercise

**Instructions:** ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

Suppose we have a simple linear regression model, predicting credit card balance as a function of age. We fit a model and estimate that $\beta_0 = 0$ and $\beta_1 = 10$.

Now suppose that I create a new dataset with columns (balance, age, age2) where age2 is simply a copy of age. And I want to fit a multiple regression model with coefficients $\beta_0$, $\beta_1$, and $\beta_2$.

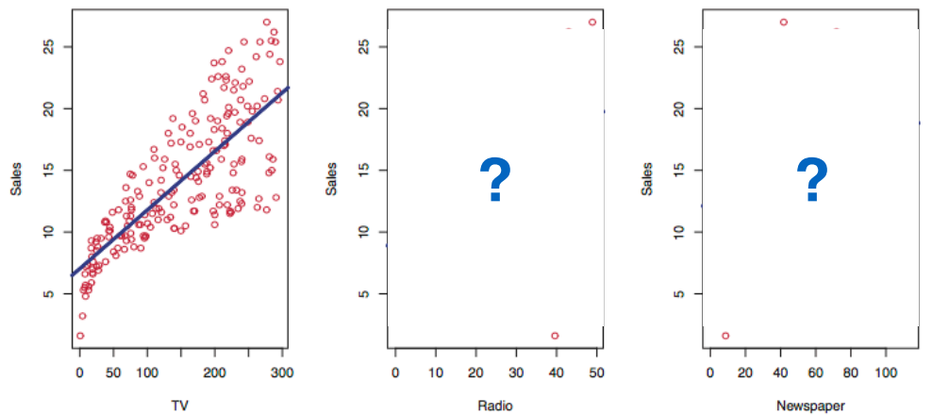
A. Where is $J(\beta_0,\beta_1,\beta_2)$ minimized?

B. How might the duplicate predictor variable affect gradient descent?

# Exercise

**Instructions:** ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

Suppose I have a dataset of Sales and budgets spent on advertising on various platforms (TV, Radio, Newspape etc.)




First, I fit a simple linear regression model on one predictor variable: TV. Second, I fit a multiple linear regression model with three predictor variables: TV, Radio, Newspaper. Third, I fit a multiple linear regression model with 10 predictor variables: previous three plus 7 random variables ("noise"). Which model will likely have the lowest sum of squared error?

A. First model

B. Second model

C. Third model

D. We don't have enough information: it depends on which predictor variable is most correlated with Sales

# Exercise

**Instructions:** ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

Suppose I have a dataset on credit balances with predictor variables such as gender, age, number of jobs working, education (in years), number of years employed.

I fit a multiple linear regression model. The predictor variable with the largest coefficient is the one variable most strongly correlated with the target variable.

A. True

B. False

Whichever choice you make, be prepared to explain your answer, perhaps referencing the credit balance dataset.

# Exercise

**Instructions:** ~3 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

Suppose I add fake ("pseudo") records to my dataset. I will add one pseudo-record for each predictor variable. The pseudo-record $p_j$ for variable j is

$$p_j = (0, 0, \ldots, \text{sqrt}(r), 0, \ldots, 0)$$

where the constant sqrt(r) is in the $j^{th}$ entry and the observed y value for this pseudo-record is 0.

Write down the cost function J(β) for this augmented dataset. Simplify the equation into two terms: one for the "real" data and one for the "fake" data.