Introduction to Data Science
SKKU University, Summer 2015

Prof. Jevin D. West

University of Washington

Lecture 9 – July 9, 2015

Agenda

- 9:30 – 10:30   Regression
- 10:30 – 10:40   Break
- 10:40 – 11:50   Preliminary Analysis
- 11:50 – 12:00   Questions

Week 2

- **DUE** Friday, July 10 at the beginning of class: <u>Preliminary Analysis</u>

Group Project, Preliminary Analysis

- **DUE** Friday, July 10 at the beginning of class: Preliminary Analysis and summary statistics (several tables and graphs). Conduct a first-pass descriptive analysis of your dataset. Produce tables and graphs that show exactly what data you have, and that contain summary statistics about the data. Questions to answer for each data source include:
  - How many unique observations to you have?
  - What information/features/characteristics do you have for each observation?
  - What are the min/max/mean/median/sd values for each of these features? What is the distribution of the core features (show a histogram)?
  - Are there obvious trends in the data (over time, across subgroups, etc.), and are the differences statistically significant?
  - What are the other salient aspects of the data (e.g. geospatial factors, text content, etc.)
  - Provide a bullet-list of the next 5-10 tasks you will perform in analyzing your dataset.
- Submit to your team repository on github
- Please include pictures and names of team members in the report

Regression