Linear Regression

- Goal: What is the relationship that ties all points together?
- Increase x by 1, expect a linear increase in y.
- Trying to study what the randomness is
- Ex. x = age; y = income
  - Lots of incomes for one age
  - Most of the time, income will fall in a Gaussian range
    - However, there is also randomness to the income
    - Different distributions for different values of x
    - For each value of x, y is random and follows a distribution around some mean
- Assumptions:
  - The data we observed was generated by some linear function plus some noise
- Cost function:
  - Try to minimize to find the optimal solution
  - Square Euclidean distance so that it's a convex function (there is one

minimum)

- Will you converge if you keep increasing the sample size?
- If we all conducted the same study, on average, will we be close to the true relationship?
- Linear regression- Need to specify which features are linearly related