Cross validation

AKA most important lecture of your time here

what are we doing here?

- lets talk about the process of data science

A. define a business problem

1. make tesla cars the most dependable cars around

B. collect some relevant data

2. car event logs, repair/service data, driver habits

C. train a model

3. features: event statistics, target: time to failure

D. deploy model

4. predict time to fail on parts, send notifications/technicians out to parts with low time

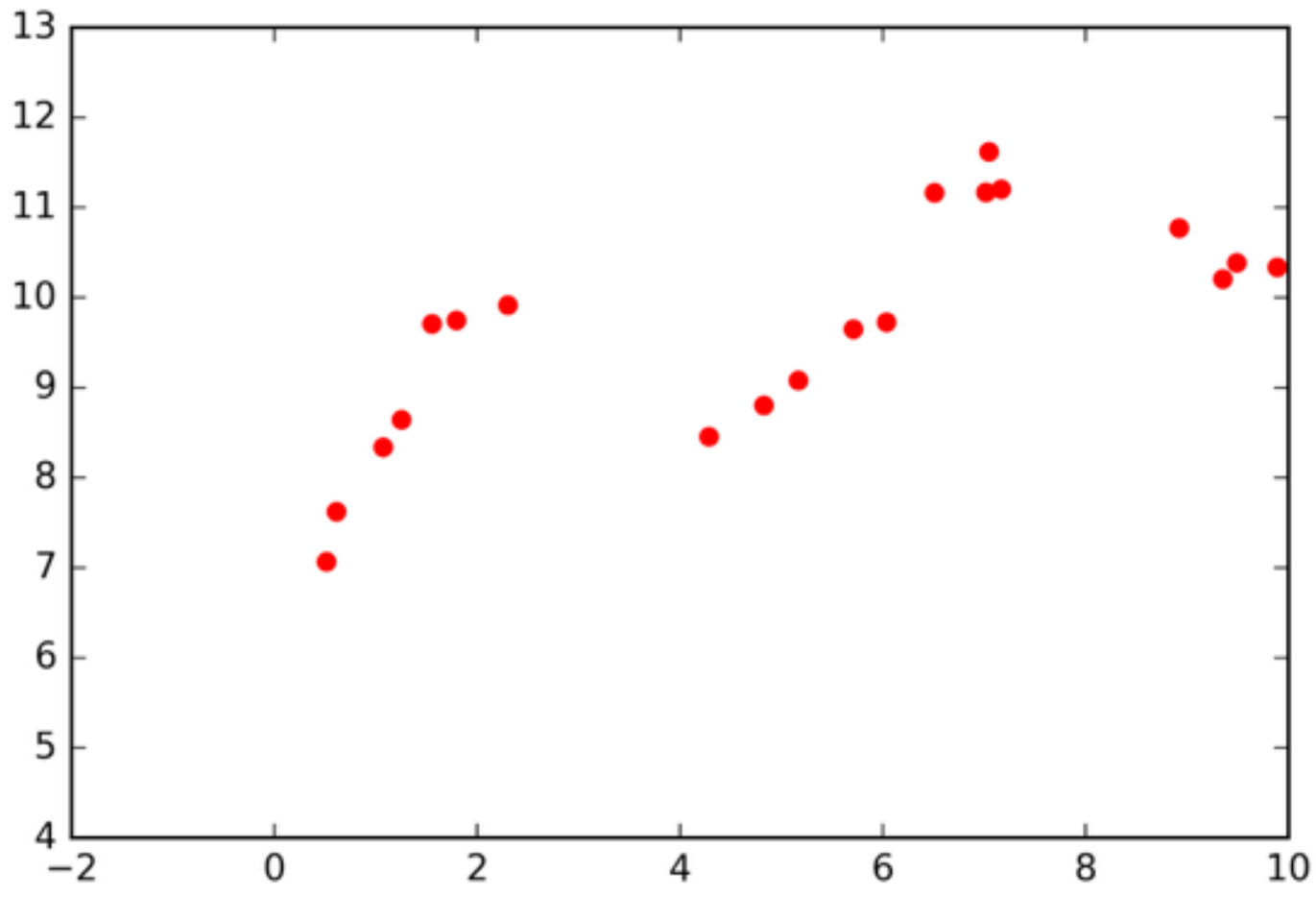
how do models work?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon$$

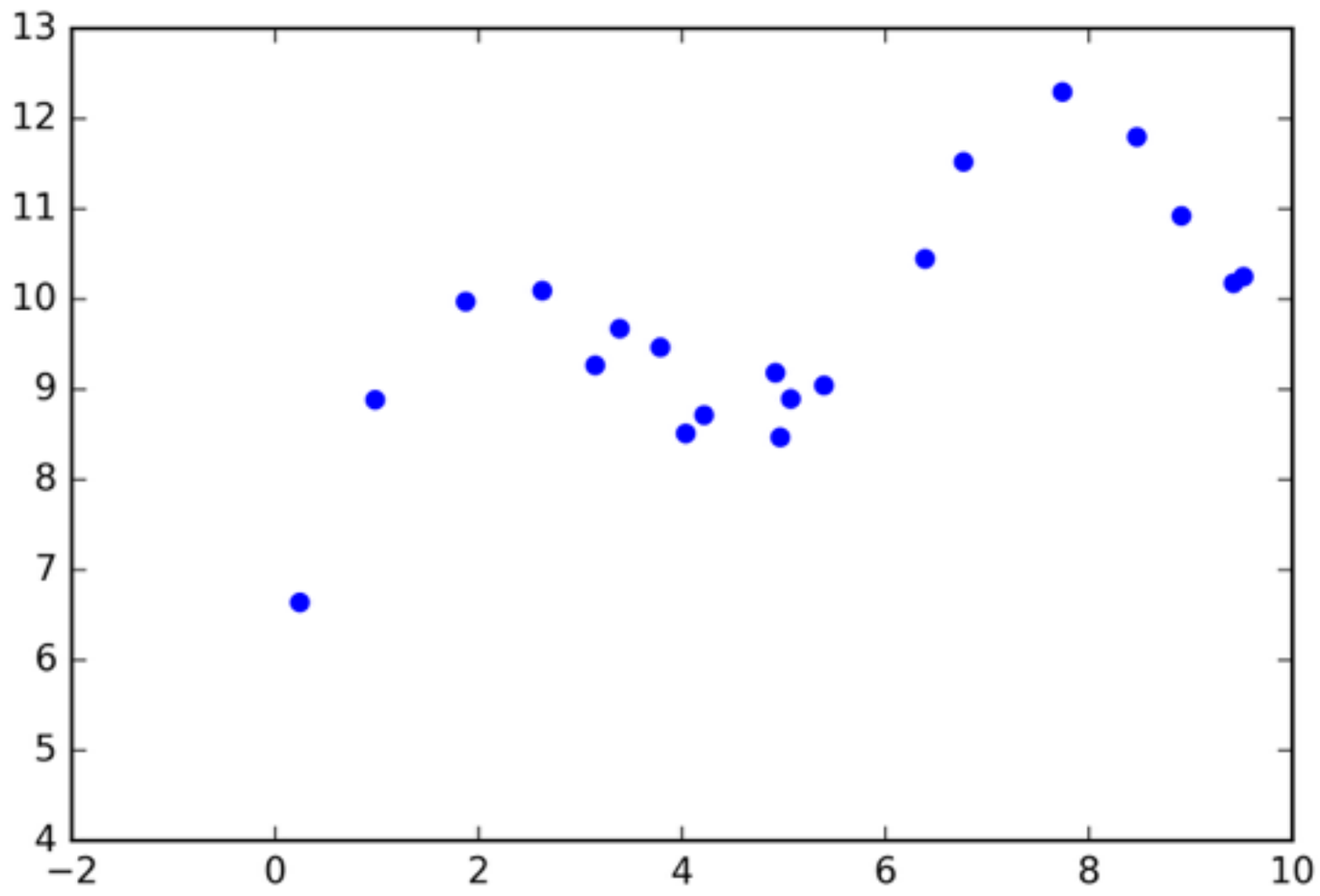
how do models work?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon$$

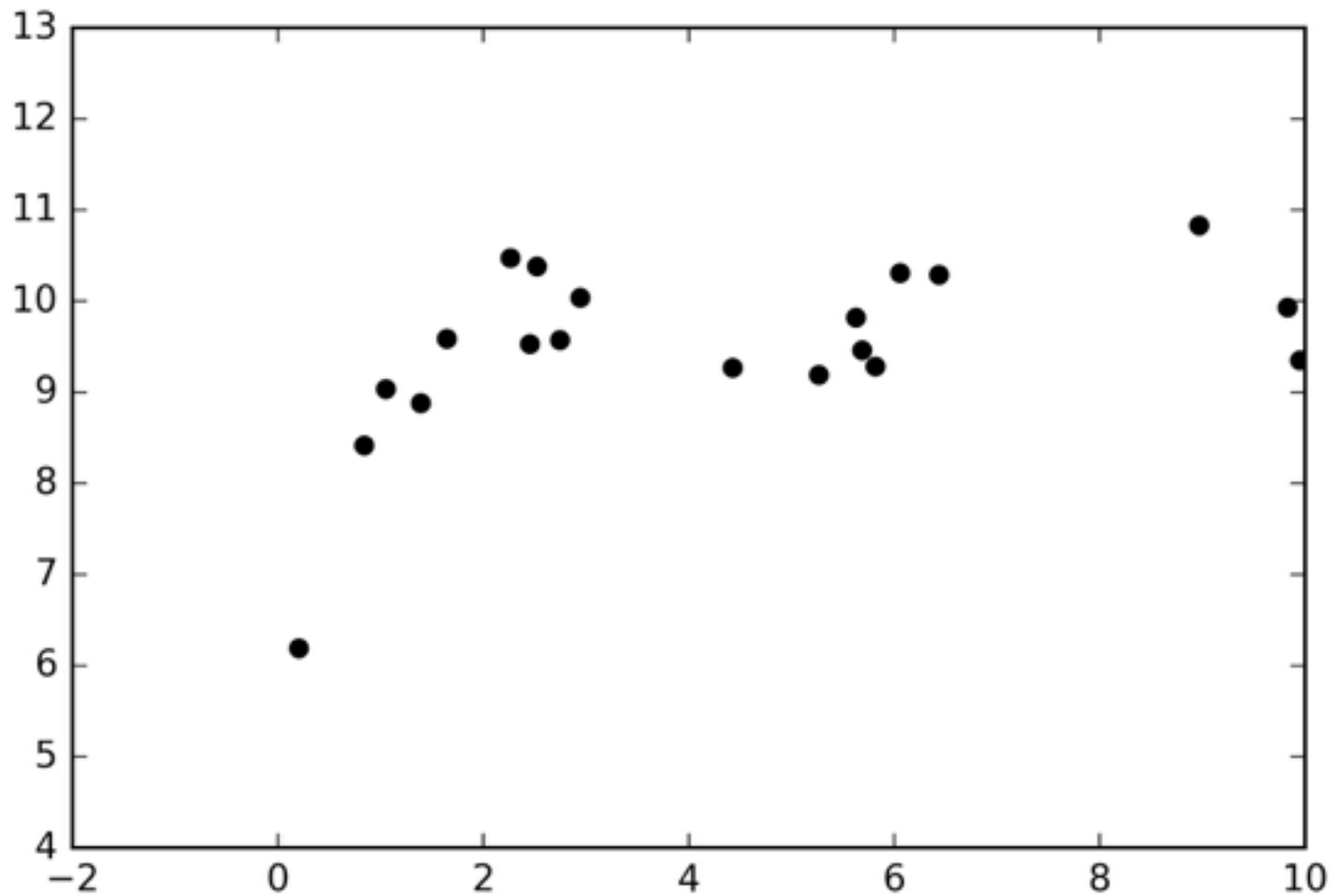
lets fit some data



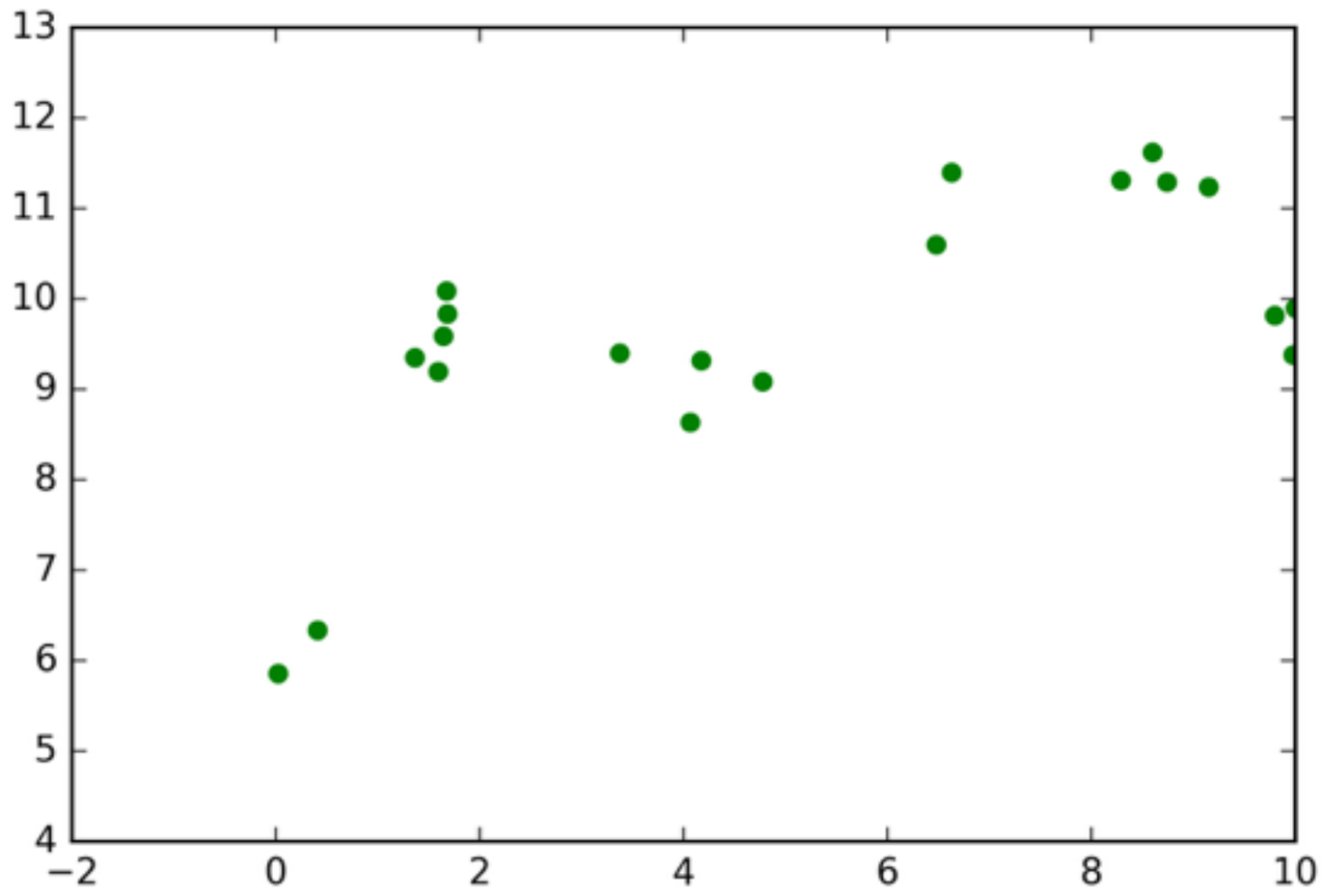
lets fit some data



lets fit some data



lets fit some data



but what is going on behind the sampling?