

Problem Set 1

Seth Harrison

1/17/2020

Statistical and Machine Learning

1. Describe in 500-800 words the difference between supervised and unsupervised learning.

What is the relationship between the X 's and Y ?

In supervised learning, the outputs (Y) are based on expected values based on prior knowledge. By contrast, in unsupervised learning, the outputs (Y) are not based on any external knowledge and are based only on the structure of the data.

What is the target we are interested in?

For supervised learning, because the outputs are based off of expected values, common targets include both classification and regression.

How do we think about data generating processes?

What are our goals in approaching data?

How is learning conceptualized?

Linear Regression

- a. Predict miles per gallon (mpg) as a function of cylinders (cyl). What is the output and parameter values for your model?