IDS 702

Prediction

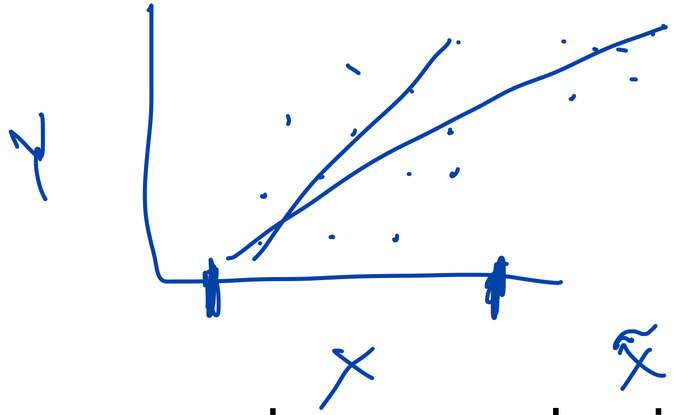
2 Main Goals of Regression Models

- Sometimes we want to use a regression model to conduct **inference**
- Inference is mainly concerned with assessing the relationship between predictors and the outcome
- “Are X and Y associated?”
- “What is the effect of X on Y?”

Prediction

- Sometimes we want to use the regression model to make predictions
- Given a set of characteristics ($\tilde{X}$), what is the predicted value of y?
- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_1 + \ldots + \hat{\beta}_p \tilde{x}_p$

Be wary of extrapolation!



- **Extrapolation**: using a regression line to predict a response whose x-value is outside of the initial range of X

- Note that with multiple linear regression (several predictors), it can be easy to fall into the trap of extrapolation

Inference or prediction?

Key questions:

- Are you more interested in the predicted values ($\hat{Y}$) or the model output (estimates, p-values, confidence intervals)?
- Are you interested in **relationships** between the predictors X and outcome Y? Or simply predicting an accurate value of Y?
- Are you interested in the past or the future?
- Are you interested in the data-generating process?

↳ inference ↳ prediction

Inference

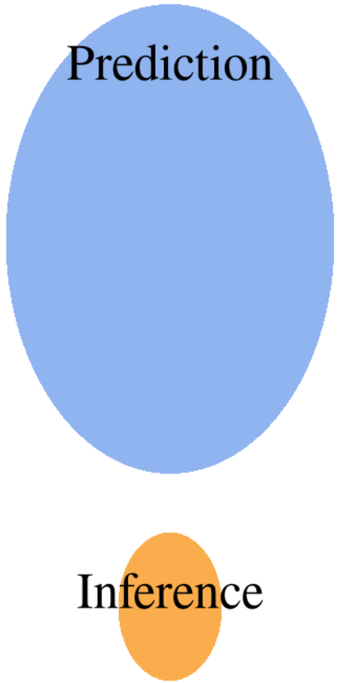
Inference or prediction?

Research questions:

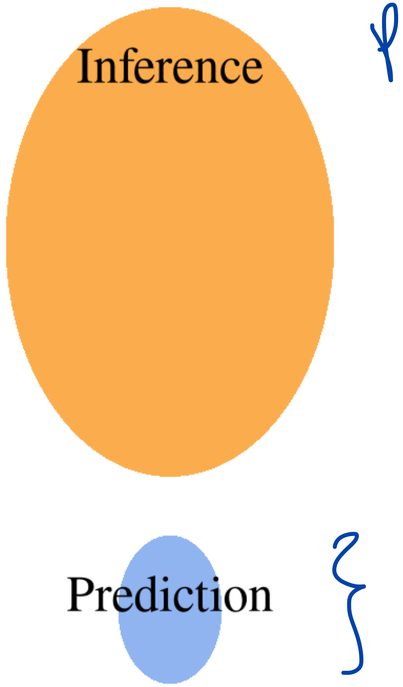
(ISLR 2.1): Model regressing house values on predictors including school zone, distance to a river, crime rate, etc

- *How much extra will a home be worth if it has a view of the river?*
- *Is a particular home with certain characteristics under- or over-valued?*

Machine Learning



Statistics



p-values
confidence intervals

Quantifying uncertainty

https://www.datascienceblog.net/post/commentary/inference-vs-prediction/