Introduction to Classification

Matt Brems

GENERAL ASSEMBLY

Data Science Process

1. Define problem.
2. Gather data.
3. Explore data.
4. Model with data.
5. Evaluate model.
6. Answer problem.

Modeling

- Modeling is something that we naturally do.
- A **model** is a simplification of reality.
- How do we simplify?
  - Taking into account only really important factors.
  - Making assumptions about how things behave.

Modeling Techniques

Introduction to Classification

- If our **Y variable is continuous**, then we refer to this problem as a **regression problem.**

- If our **Y variable is discrete**, then we refer to this problem as a **classification problem.**

We can turn our continuous Y into a discrete Y!

- Voting Example:

- Lyft Example:

What is the same about classification and regression?

- The bias-variance trade-off is still **extremely** relevant!
  - We will always do a train-test split.
  - We can still cross-validate.
  - We can still do feature selection and feature engineering!

- We're still simplifying reality with a model... our Y just looks a little different.

What is different about classification?

- Our observed values may or may not be ordered.

- I will evaluate my model differently.

