



# Annotation and Feature Engineering

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

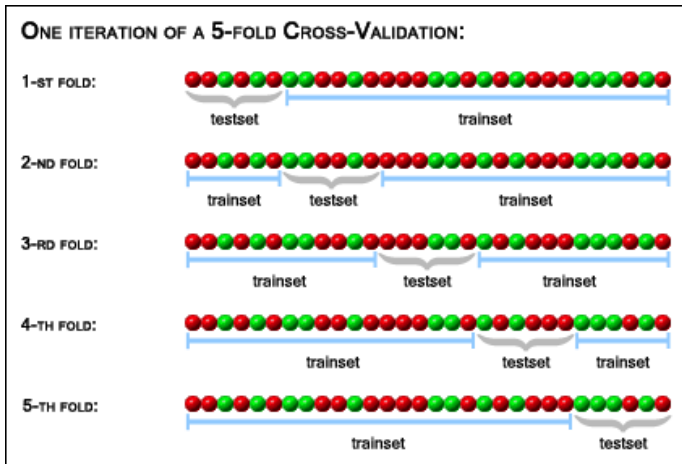
HOUSES, SPOILERS, AND TRIVIA

## Meta-Lessons

---

- How to see how you're doing
- How to deal with a lack of features

## Divide Your Data



## Only works if there aren't ordering effects!

---

- Question Answering: Askable things change
- Elections: You only care about predicting the future
- Thus, you must create validation sets that best mimic your final evaluation

## No Features, No Problem

---

- If you have a clear signal: deep learning
- No clear signal: unsupervised training
- Evaluation is hard

## “Eyeball” test

---

- Talk more about clustering later in class
- Each cluster defined by set of words

## This is silly, but ...

---

- Using this as a feature improves MT
- Using this as a feature improves classification on limited training data