

CPSC 330 Lecture 3: ML fundamentals

Varada Kolhatkar

Announcements

- Homework 2 (hw2) has been released (Due: Sept 16, 11:59pm)
 - You are welcome to broadly discuss it with your classmates but final answers and submissions must be your own.
 - Group submissions are not allowed for this assignment.
- Advice on keeping up with the material
 - Practice!
 - Make sure you run the lecture notes on your laptop and experiment with the code.
 - Start early on homework assignments.
- If you are still on the waitlist, it's your responsibility to keep up with the material and submit assignments.
- Last day to drop without a W standing: Sept 16, 2023

Recap

- Importance of generalization in supervised machine learning
- Data splitting as a way to approximate generalization error
- Train, test, validation, deployment data
- Overfitting, underfitting, the fundamental tradeoff, and the golden rule.
- Cross-validation

iClicker 3.1

Clicker cloud join link: <https://join.iclicker.com/VYFJ>

Select all of the following statements which are TRUE.

- a. A decision tree model with no depth (the default `max_depth` in `sklearn`) is likely to perform very well on the deployment data.
- b. Data splitting helps us assess how well our model would generalize.
- c. Deployment data is scored only once.
- d. Validation data could be used for hyperparameter optimization.
- e. It's recommended that data be shuffled before splitting it into train and test sets.

iClicker 3.2

Clicker cloud join link: <https://join.iclicker.com/VYFJ>

Select all of the following statements which are TRUE.

- a. k -fold cross-validation calls fit k times
- b. We use cross-validation to get a more robust estimate of model performance.
- c. If the mean train accuracy is much higher than the mean cross-validation accuracy it's likely to be a case of overfitting.
- d. The fundamental tradeoff of ML states that as training error goes down, validation error goes up.
- e. A decision stump on a complicated classification problem is likely to underfit.

Class demo