

MA477: Data Science
Lesson 5 Board Sheet — 16 January 2026
 United States Military Academy, West Point
 Instructor: MAJ Patrick Kuiper

1 Lesson 5 Resampling, CV, Bootstrapping - Lesson Objectives

- Understand how resampling methods are used in model assessment.
- Know how to use cross-validation to estimate the test error of a statistical model.
- Know the differences between leave-one-out cross-validation and k-fold cross-validation.
- Understand how the bias-variance tradeoff applies when deciding on the number of folds.

2 Limited data discussion

1. With limited data, what types of error can we reduce?
2. Given we are limited to the data available what can we do to improve our model?
3. What are the costs of this method?

3 A simple visual (5-fold CV)

Legend: **T** = training fold, **V** = validation fold

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Round 1:	V	T	T	T	T
Round 2:	T	V	T	T	T
Round 3:	T	T	V	T	T
Round 4:	T	T	T	V	T
Round 5:	T	T	T	T	V

4 Coding Exercise

Pair Up and Think of a problem that we can define a hypothetical distribution from. Each pair will work on the Colab notebook template provided and write your to the following points on the board.

You may reference the Lesson 5 Colab Notebook for help.

1. What problem are you answering?
2. How are you generating your data?
3. What does your data look like?
4. What statistics are you looking at? What conclusions do you have?

At the end I want several groups to present their results to the class.