

MA477: Data Science
Lesson 22 Board Sheet — 10 March 2026
United States Military Academy, West Point
Instructor: MAJ Patrick Kuiper

1 Random Forest Lesson Objectives

- Understand how the random forest algorithm differs from bagging.
- Use functions in Python's sklearn to fit random forest models.
- Understand how you can extract feature importance using Random Forests

Discussion Questions

1. Why does bagging fail to substantially reduce variance when one predictor is much stronger than all others?
2. How does randomly selecting m predictors at each split help reduce correlation between trees?
3. What happens if we set $m = p$ in a random forest?
4. Why can we increase the number of trees B in a random forest without causing overfitting?