1 - Summary

This week's lecture focused on decision trees and ensemble methods like bagging, random forests, and boosting. We learned how trees partition the predictor space into simple regions and how ensemble methods improve prediction accuracy by combining multiple models.

2 - Concepts

- **Decision trees**: Segment the predictor space into disjoint regions using recursive binary splits, creating interpretable, rule-based models.
- **Regression vs. classification trees**: In regression, each region predicts the mean response; in classification, it predicts the majority class.
- **Tree pruning**: A way to combat overfitting by growing a large tree and then trimming it back using cost-complexity pruning, guided by cross-validation.
- **Bagging**: Uses bootstrap samples to build many trees in parallel and averages their predictions to reduce variance.
- **Random forests**: An extension of bagging where each split considers a random subset of predictors, further reducing variance and decorrelating trees.
- **Boosting**: Builds trees sequentially, where each new tree improves on the residual errors of the previous ones. It introduces hyperparameters like learning rate (shrinkage), number of trees, and interaction depth.
- **Hyperparameters**: Include tree depth, minimum node size, learning rate, and number of trees; each affects the bias-variance trade-off.

3 - Uncertainties

I'm still working to understand when to choose between bagging, random forests, and boosting. While all reduce variance, boosting also reduces bias, but it seems more sensitive to hyperparameters. I'm also unsure how to balance interpretability with performance. Decision trees are easy to interpret, but ensemble methods often are not. Lastly, I'd like more clarity on how hyperparameters interact, such as how tree depth and learning rate affect overfitting in boosting, or how the number of predictors considered at each split changes the behavior of random forests.