Week 4 Learning Reflection:

Summary:

This week, we explored resampling methods, focusing on how they can help us estimate model accuracy and select models. We specifically looked at cross-validation and the bootstrap technique, which are powerful tools when we lack access to large independent test sets. These methods help us understand the performance of statistical learning procedures and assess their prediction error more reliably.

Concepts:

- **Resampling Methods:** Techniques used to repeatedly draw samples from a training set and refit a model of interest to each sample in order to estimate prediction error.
- **Cross-Validation:** A model evaluation method that splits the data into training and validation subsets multiple times to measure how the results of a statistical analysis will generalize to an independent dataset.
- **Bootstrap Method:** A powerful statistical tool that involves repeatedly sampling with replacement from the dataset and computing a statistic (e.g., mean, regression coefficient) for each sample. Useful for estimating the **accuracy (variance)** of a statistic. Allows for estimation of standard errors and confidence intervals. Often used when theoretical estimation is complex or unknown.

Uncertainties:

- How do I choose between using cross-validation and the bootstrap in practice? Are there situations where one clearly performs better than the other?