EE5373: Data Modeling Using R

Fall 2025

LAB 5 REPORT

Training, Testing, and Predicting

Talha Hamza

October 30, 2025

1 Part 1 Results

1.1 Benchmark int95

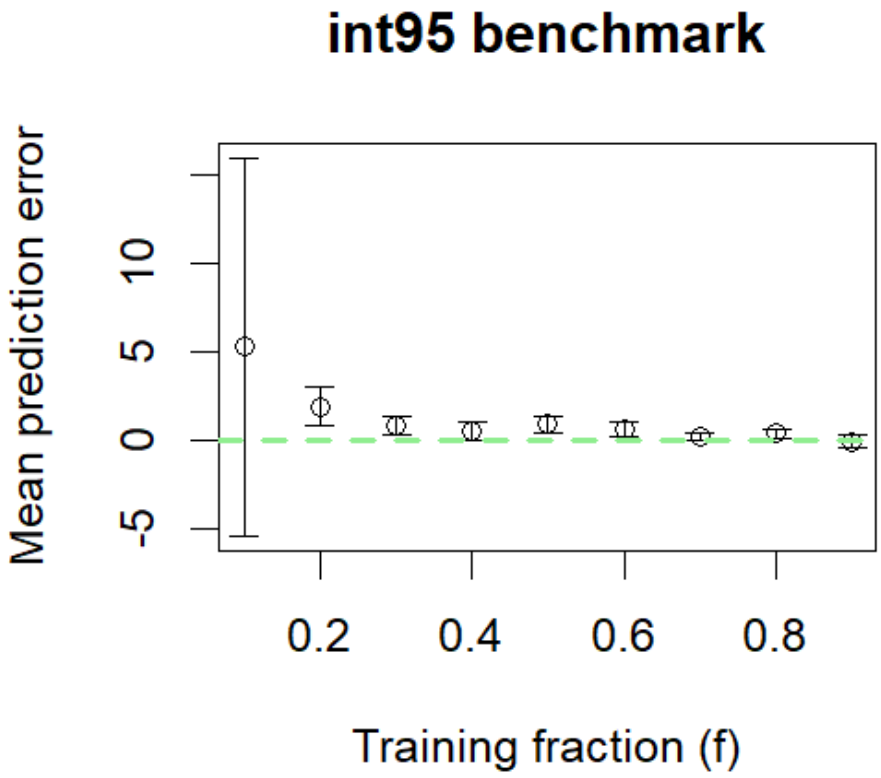


Figure 1: Mean and 95% confidence interval for D_f vs training fraction f for int95 benchmark

Discussion:

- Optimal training fraction f for int95 is **0.7**
- Behavior at extreme values:
  - At $f = 0.1$: We get a mean of about +5.5, and a very wide margin of error. The model is over predicting very frequently
  - At $f = 0.9$: The mean is centered at 0 and there is almost no margin of error.

1.2 Benchmark int06

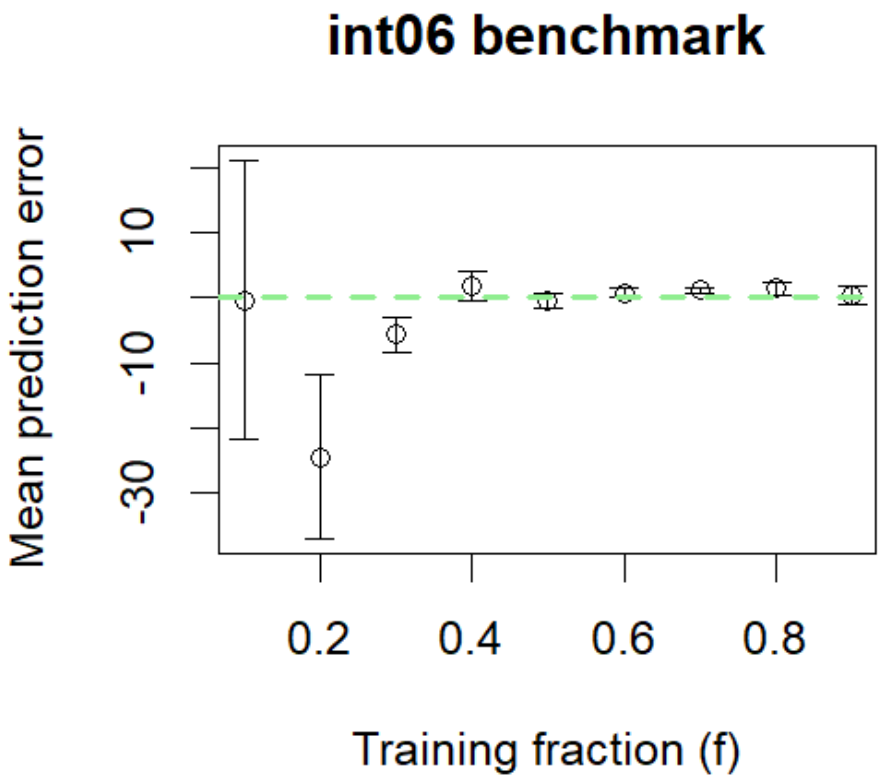


Figure 2: Mean and 95% confidence interval for D_f vs training fraction f for int06 benchmark

Discussion:

- Optimal training fraction f for int06 is **0.6** as its mean is the closest to 0 with a low margin of error.
- Behavior at extreme values:
  - At $f = 0.1$: The mean is at 0 which show that the model overestimated exactly as mush as it underestimated. The margin of error is quite large.
  - At $f = 0.9$: Also centered around 0 but with very margin little of error.

1.3 Benchmark fp95

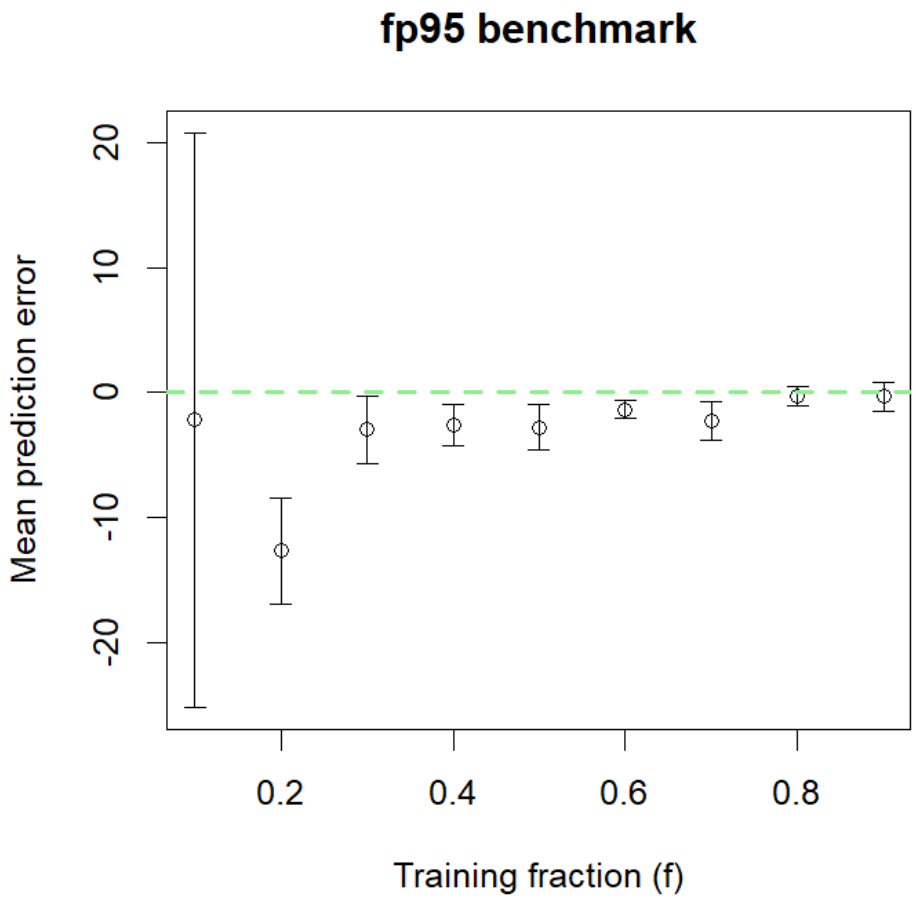


Figure 3: Mean and 95% confidence interval for D_f vs training fraction f for fp95 benchmark

Discussion:

- Optimal training fraction f for fp95 is **0.8** because it's mean is around 0 and has the smallest margin of error.
- Behavior at extreme values:
  - At $f = 0.1$: Has the greatest margin of error and is centered around -2. The model predictions are consistently inaccurate.
  - At $f = 0.9$: Has a mean of 0 but surprisingly has a higher margin of error when compared to lower training fractions, such as 0.8.

1.4 Benchmark fp06

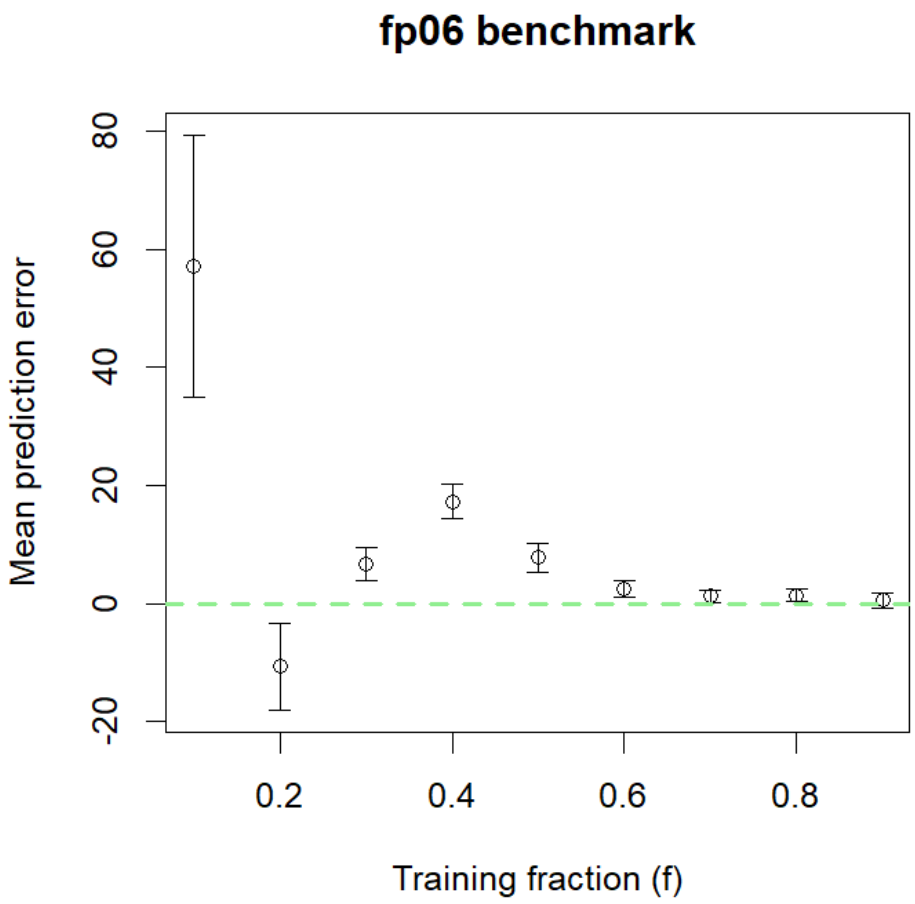


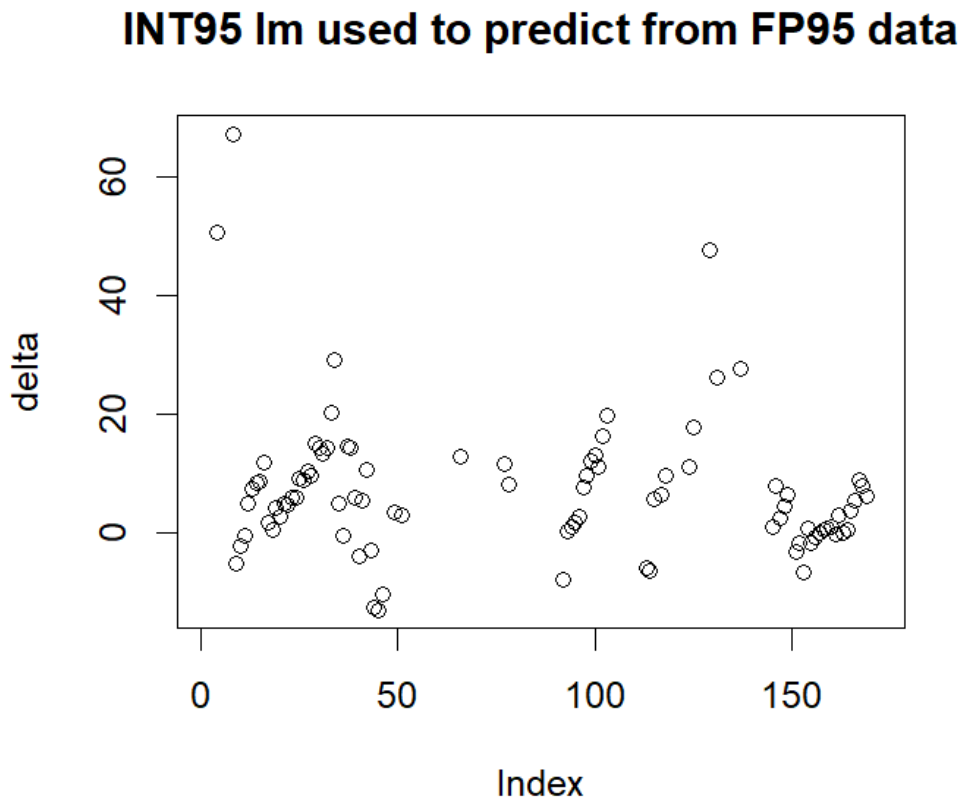
Figure 4: Mean and 95% confidence interval for D_f vs training fraction f for fp06 benchmark

Discussion:

- Optimal training fraction f for fp06: **0.9**
- Behavior at extreme values:
  - At $f = 0.1$: The model grossly over predicts at a training fraction of 0.1
  - At $f = 0.9$: The average of difference in predictions is 0 and there is very little deviation at this training fraction.
- Mean error patterns: Interestingly, the majority of mean errors are non-zero. This was not the case for previous simulations.
- Notable observations: The predictions vary in an odd, alternating pattern. At f = 0.1, the model over predicts, and at f = 0.2 it under-predicts, and again at f = 0.3 it over-predicts.

2 Part 2 Results

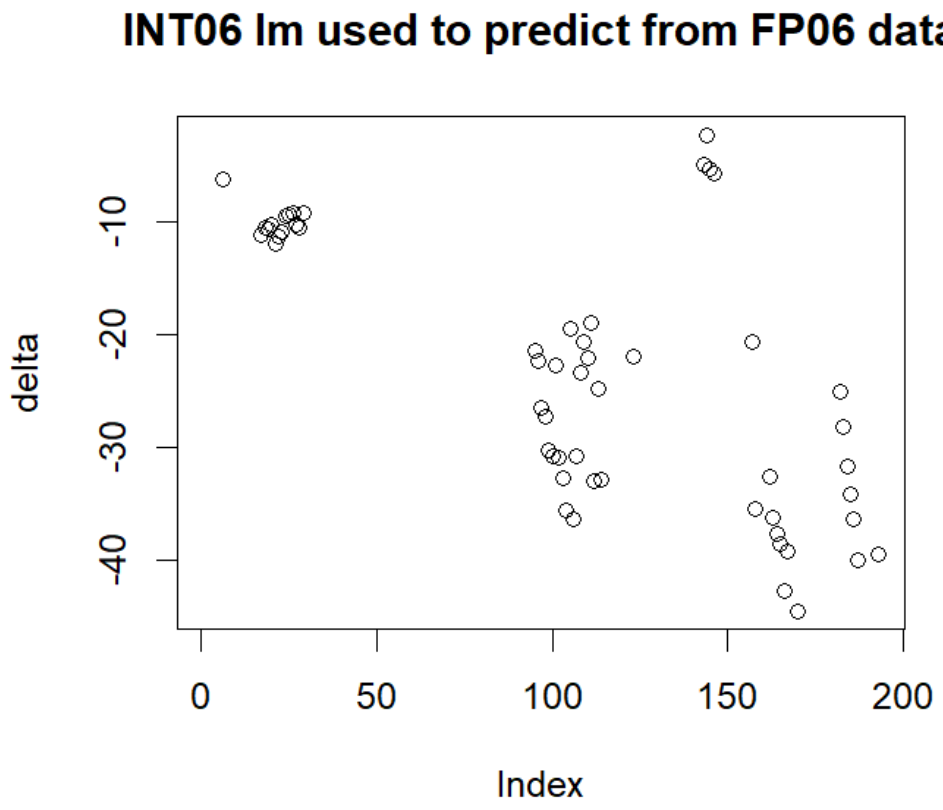
2.1 Model int95 used to predict values for fp95



Discussion:

- **Mean of deviation from actual value** = 7.081812
- **95% confidence interval**: $4.579693 < \delta < 9.583931$
- **Conclusion**: The INT95 model mostly over-predicts the FP95 data. Since the confidence interval **does not** contain 0, this would be a bad model to predict FP95 data.

2.2 Model int06 used to predict values for fp06



Discussion:

- **Mean of deviation from actual value** = -23.34504
- **95% confidence interval**: $-26.55332 < \delta < -20.13676$
- **Conclusion**: The INT06 model always under-predicts the FP06 data. Since the confidence interval **does not** contain 0, this would be a bad model to predict FP06 data.