Resampling Techniques - 101

Bert Gollnick

Resampling Techniques - 101

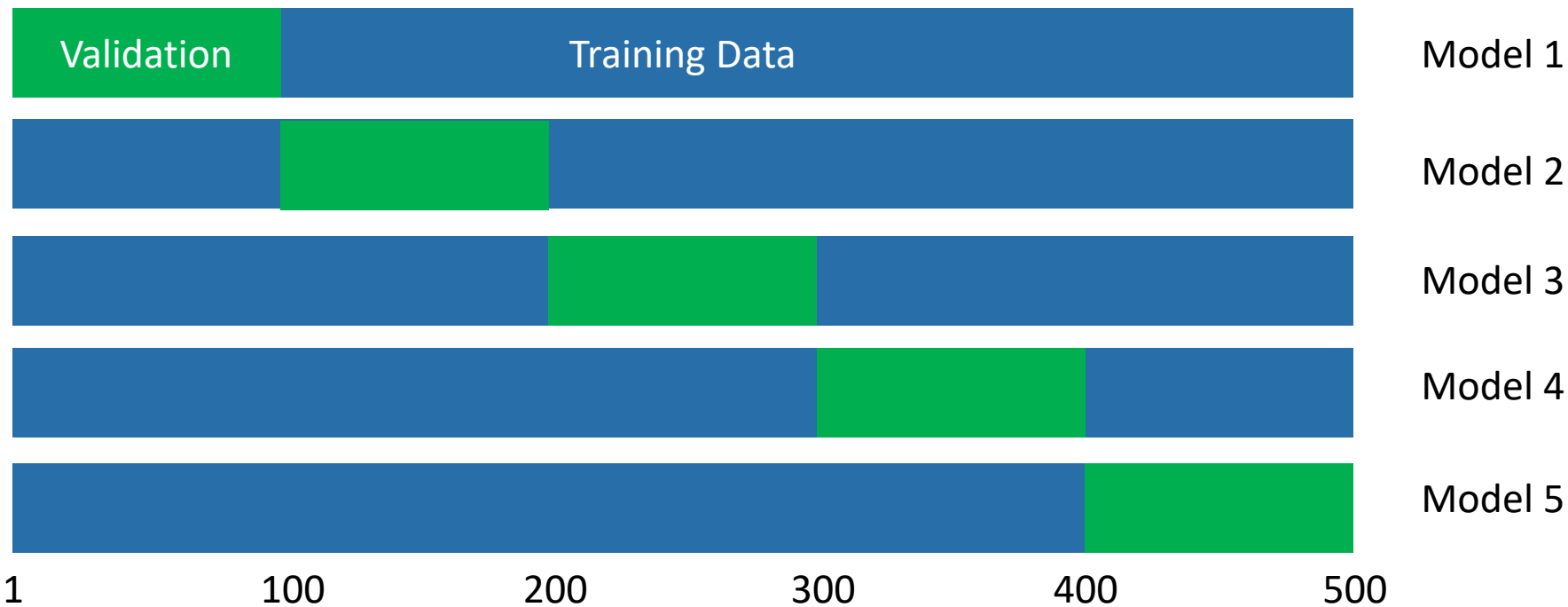Problem of Train / Validation / Test Split

All Data

Data

Best training
→ Max!

Best validation
→ Max!

Resampling Techniques - 101

Solution: k-fold Cross-Validation

- Create k different training/validation sets
- Average validation set performance

| Validation | Training Data | | Model 1 |

Model 2

Model 3

Model 4

Model 5

1   100   200   300   400   500

Resampling Techniques - 101

Example: 5-fold Cross Validation

- Averaged Performance = 0.27

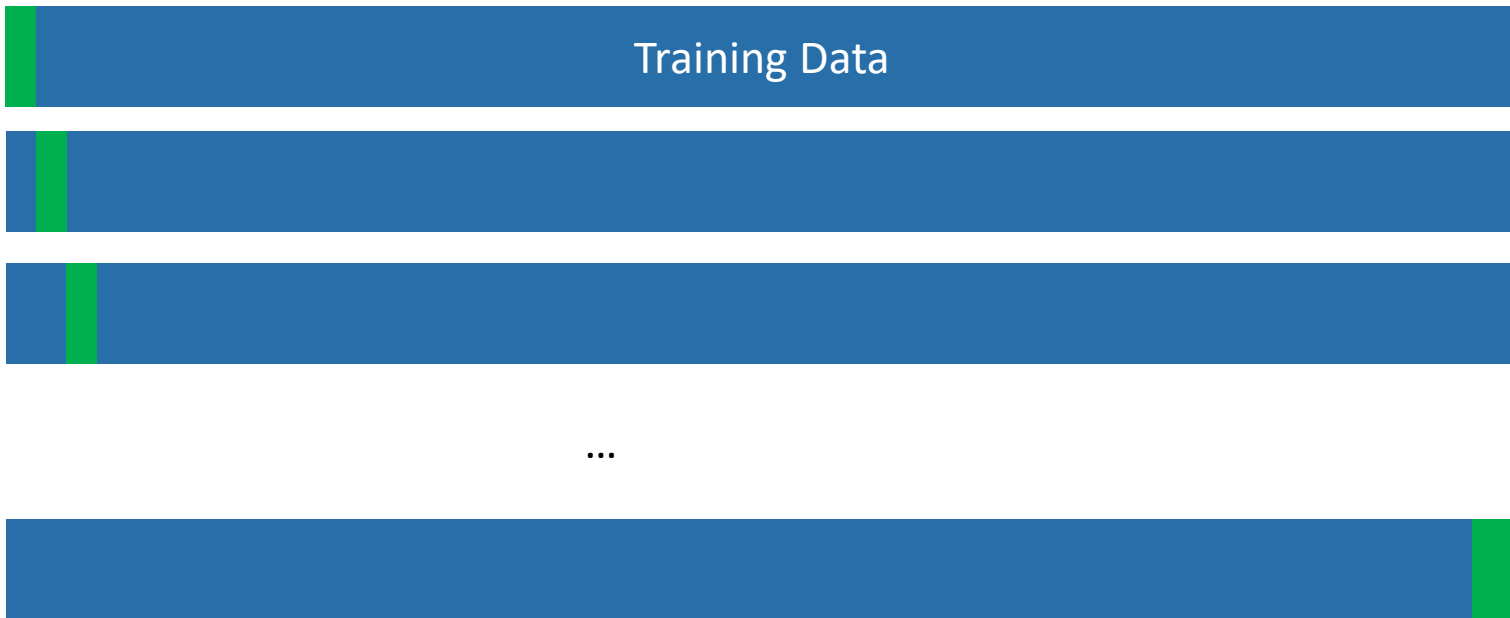| Validation | Training Data | | Val R2 = 0.25 |

Val R2 = 0.2

Val R2 = 0.3

Val R2 = 0.2

Val R2 = 0.4
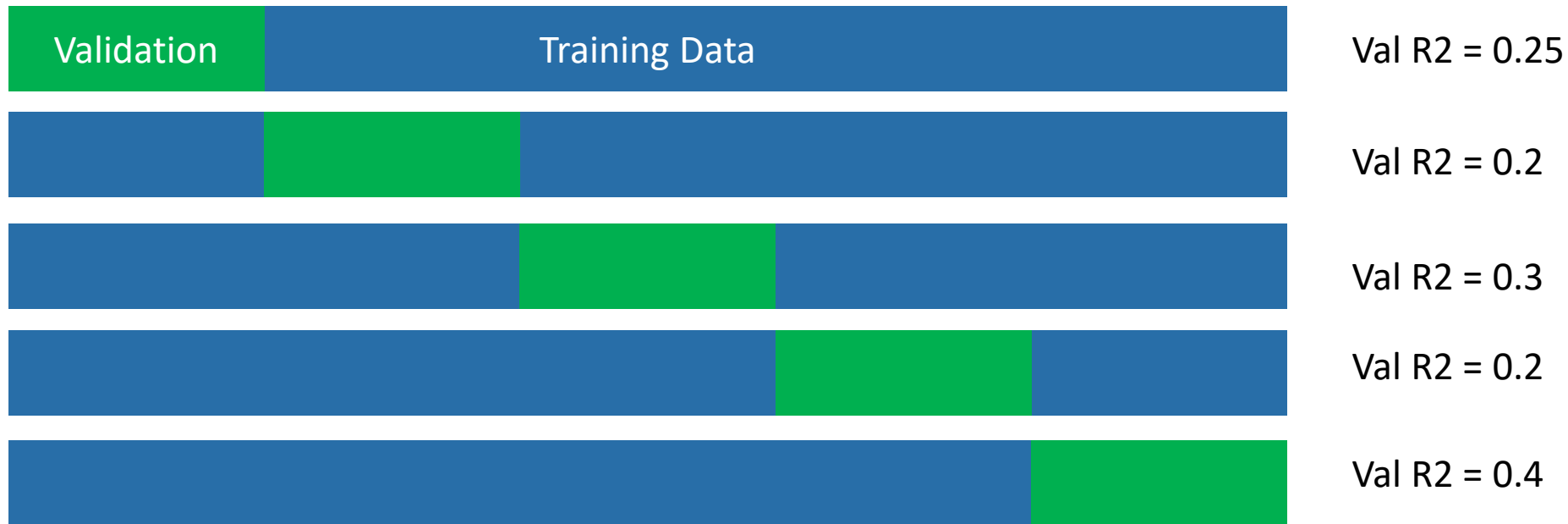
Resampling Techniques - 101

Leave One Out Cross Validation (LOOCV)

- Use all data for training, except one dataset
- One dataset used for validation

Training Data

...

Resampling Techniques - 101

Advantages / Disadvantages

+

- better / more stable results
- good to compare different ML algorithms
- good for analysing hyperparameter tuning

–

- Higher computational effort