

PSY9511: Seminar 4

Testing, resampling, and splitting

Esten H. Leonardsen

26.10.23



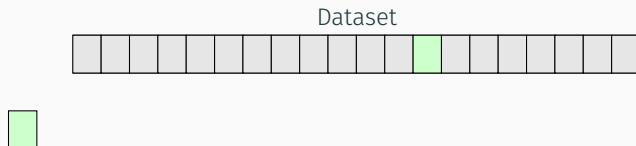
**UNIVERSITETET
I OSLO**

Model assessment: The bootstrap

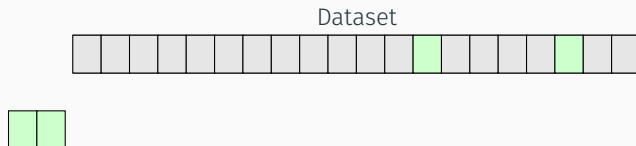
Dataset



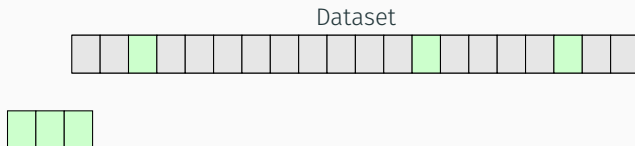
Model assessment: The bootstrap



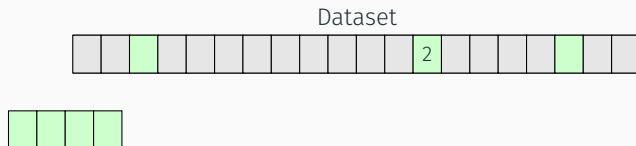
Model assessment: The bootstrap



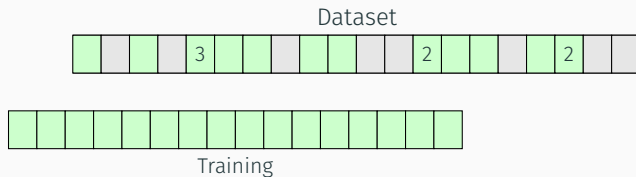
Model assessment: The bootstrap



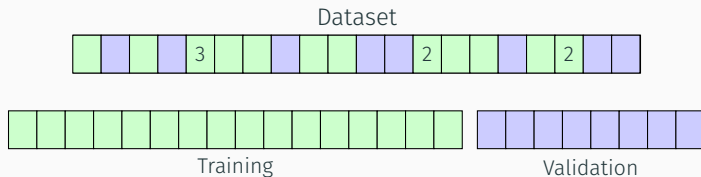
Model assessment: The bootstrap



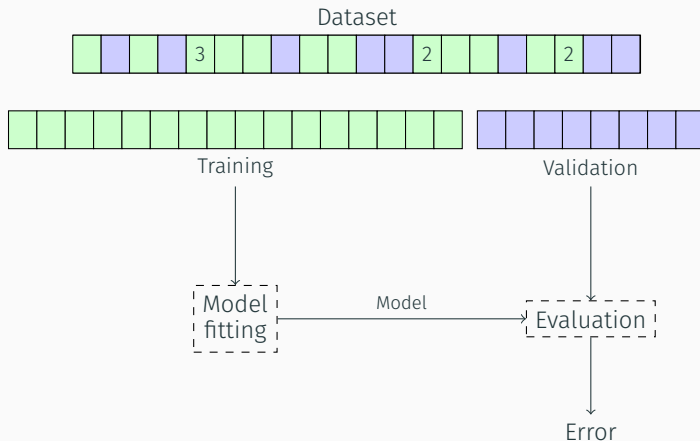
Model assessment: The bootstrap



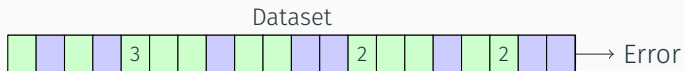
Model assessment: The bootstrap



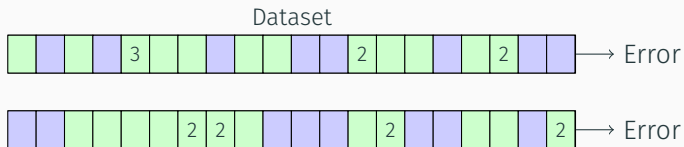
Model assessment: The bootstrap



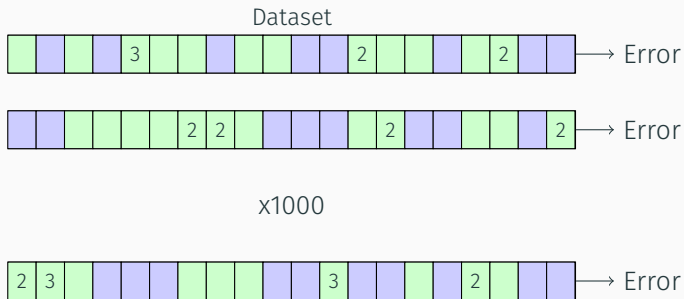
Model assessment: The bootstrap



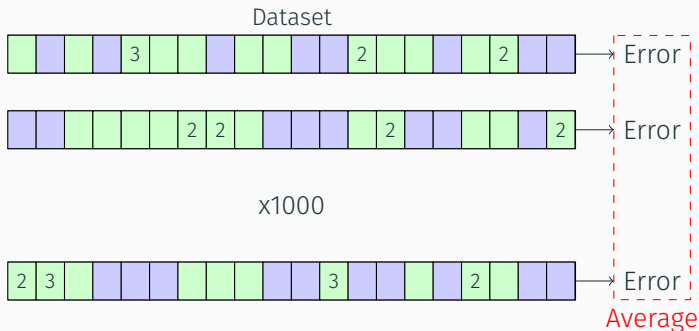
Model assessment: The bootstrap



Model assessment: The bootstrap



Model assessment: The bootstrap



Model assessment: The bootstrap

Fits x models with m datapoints each, sampled from the original dataset with replacement.

- + Uses all data to train models
- + Provides a smooth distribution of model performance
- + **Versatile: Can be used for other things, e.g. getting a confidence interval for model parameters**
- Different choices of k (and exact splits) yields different results



Model assessment: Summary

- Model assessment should **always** happen out-of-sample
- If n is big (≥ 10000), a single train/validation split is often sufficient
- For smaller samples, k -fold cross-validation with $5 \leq k \leq 10$ is a good trade-off between bias and variance
- The bootstrap is an effective way of getting confidence intervals for model parameters



Model selection and assessment



UNIVERSITETET
I OSLO

Model selection and assessment

- Model assessment via cross-validation is sufficient if we want to estimate the out-of-sample error of a known model.
- Very often we want to know whether a set of predictors are informative for an outcome *given the best possible model*
- In that case, we have to both choose the best model, and estimate its performance
- If we choose the model based on regular cross-validation, the performance estimate will likely be inflated



Model selection and assessment

- Model assessment via cross-validation is sufficient if we want to estimate the out-of-sample error of a known model.
 - Very often we want to know whether a set of predictors are informative for an outcome *given the best possible model*
 - In that case, we have to both choose the best model, and estimate its performance
 - If we choose the model based on regular cross-validation, the performance estimate will likely be inflated
- We need a more advanced strategy

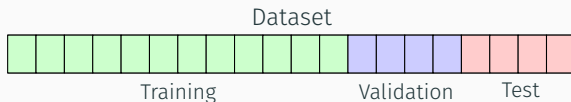


Model selection and assessment: Train/validation/test

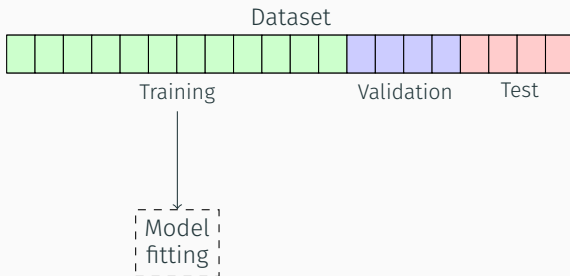
Dataset



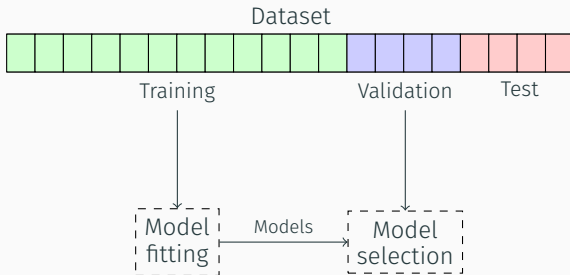
Model selection and assessment: Train/validation/test



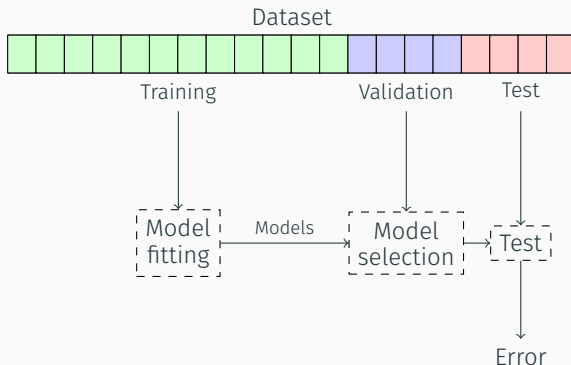
Model selection and assessment: Train/validation/test



Model selection and assessment: Train/validation/test



Model selection and assessment: Train/validation/test



Model selection and assessment: Nested cross-validation



Model selection and assessment: Summary

- Whenever a choice is made on the basis of performance in a dataset, the performance of the chosen model on that dataset is going to be biased.
- If n is big (≥ 10000), a single train/validation/test split is often sufficient
- If possible, use nested cross-validation to select the best model *and* estimate the out-of-sample error

