# PSY9511: Seminar 4

## Model selection, validation and testing

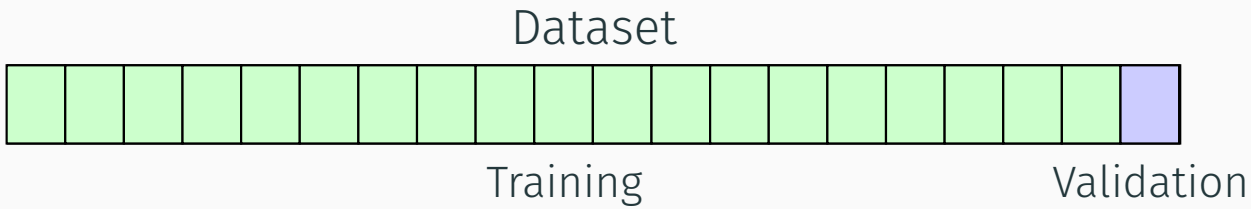
Esten H. Leonardsen

23.09.24

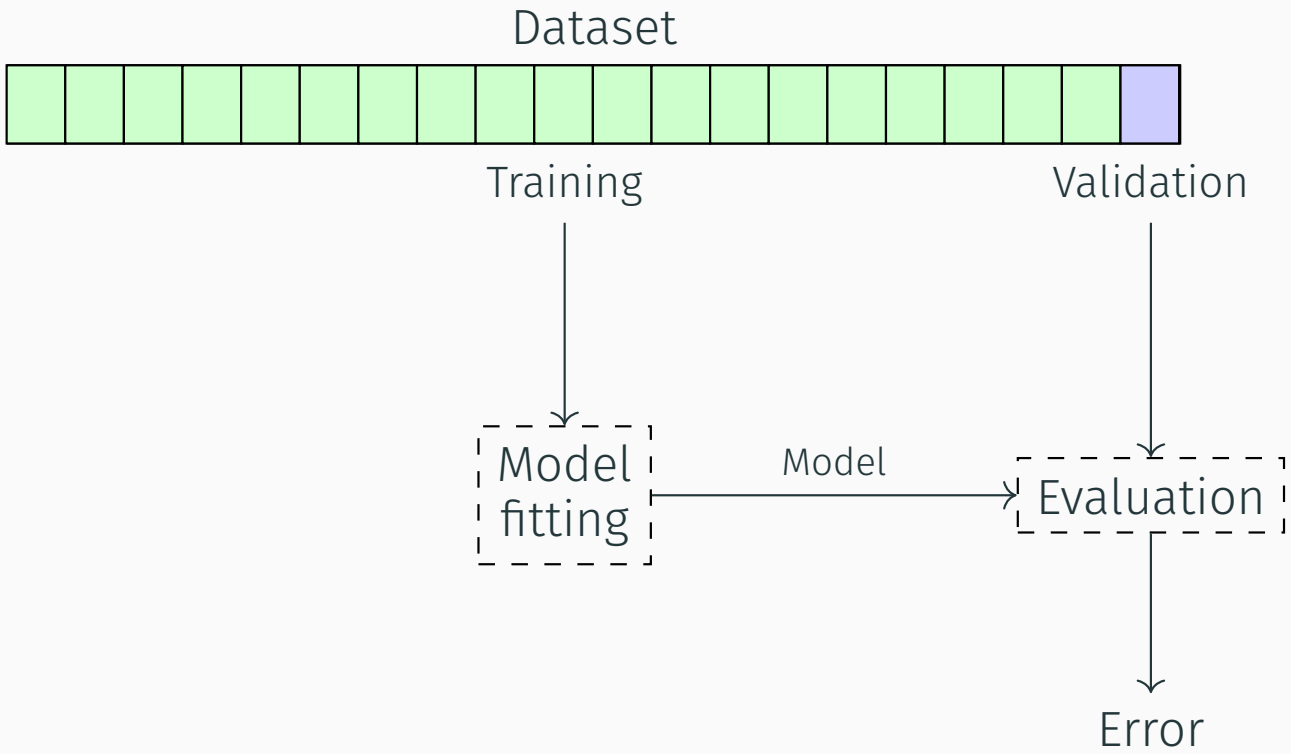

UNIVERSITETET
I OSLO

# Model evaluation: Leave-one-out cross-validation

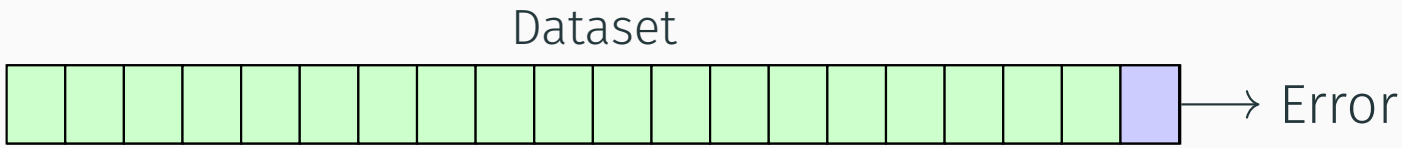
# Model evaluation: Leave-one-out cross-validation



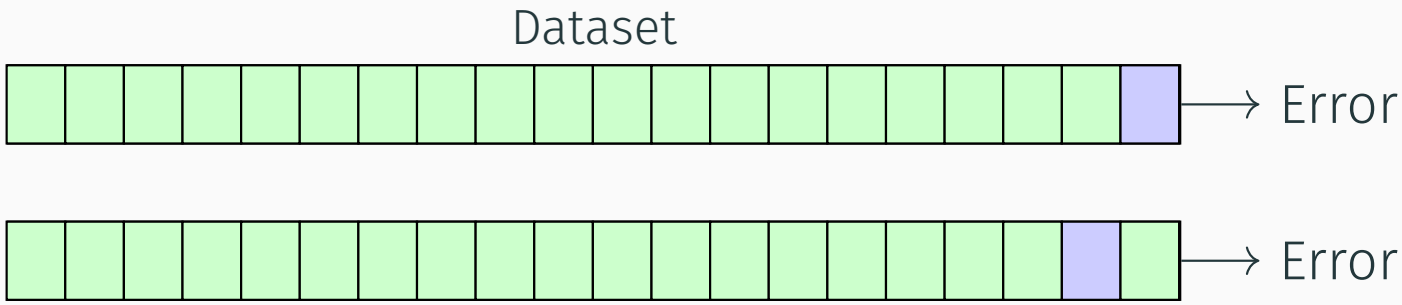
# Model evaluation: Leave-one-out cross-validation



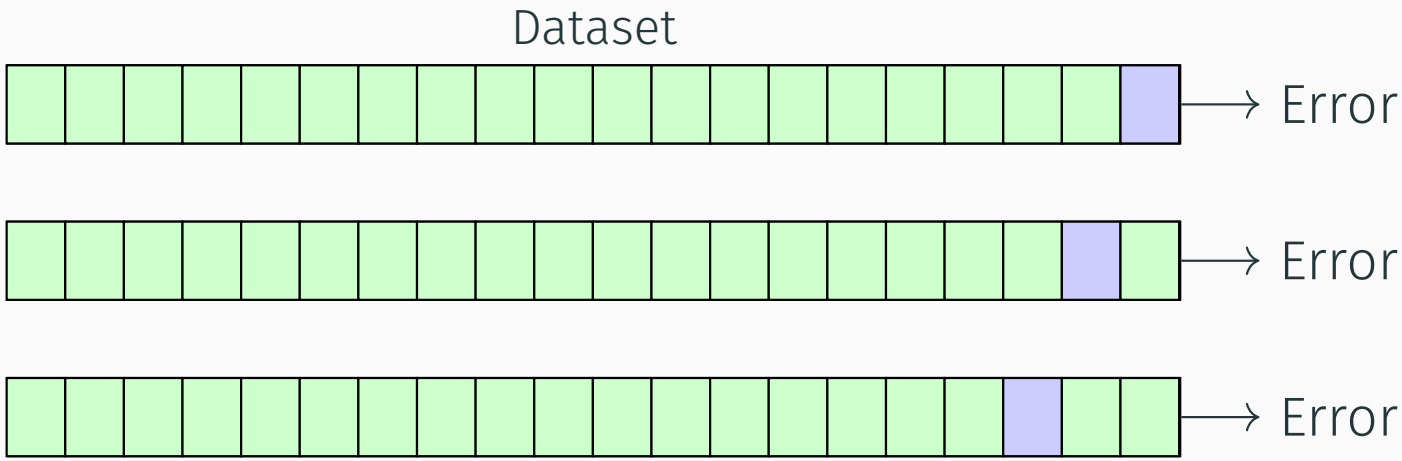
# Model evaluation: Leave-one-out cross-validation



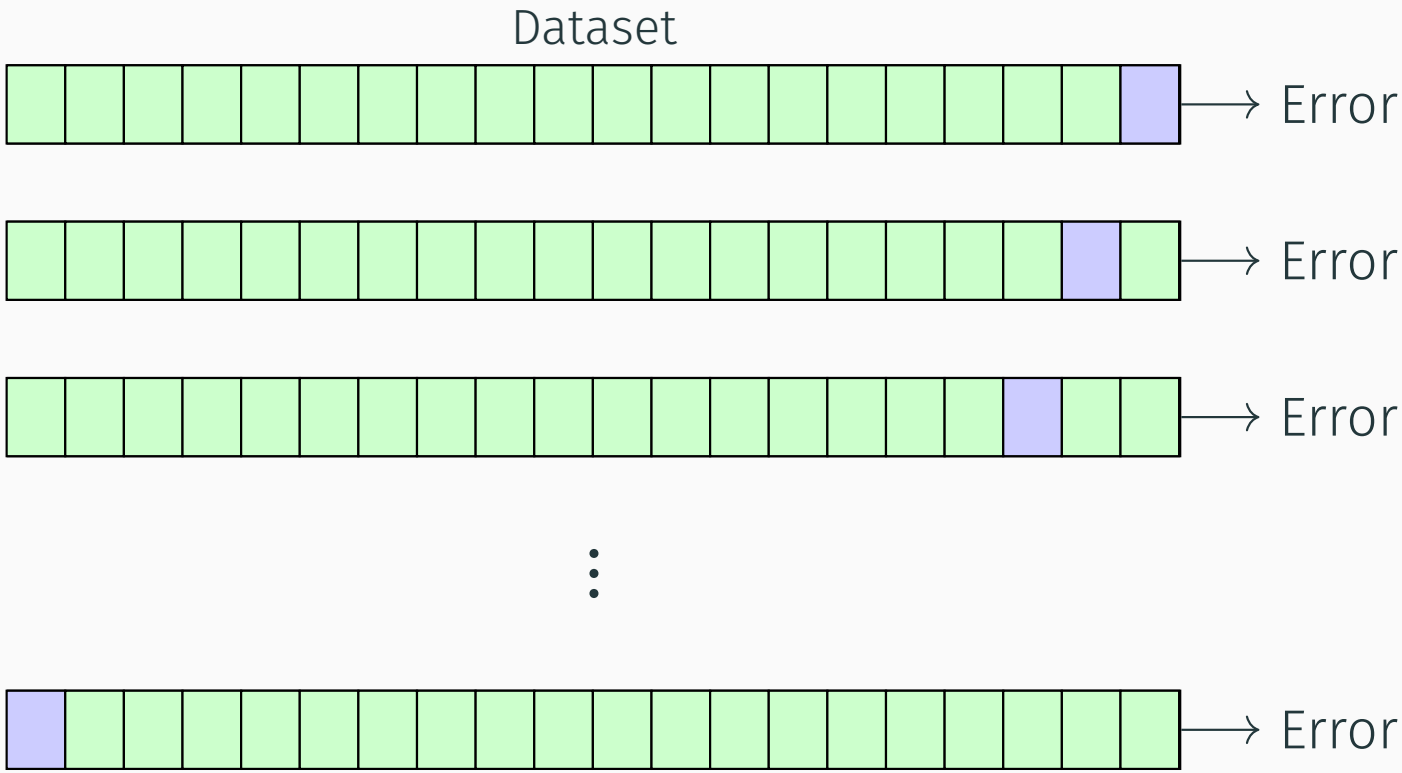
# Model evaluation: Leave-one-out cross-validation



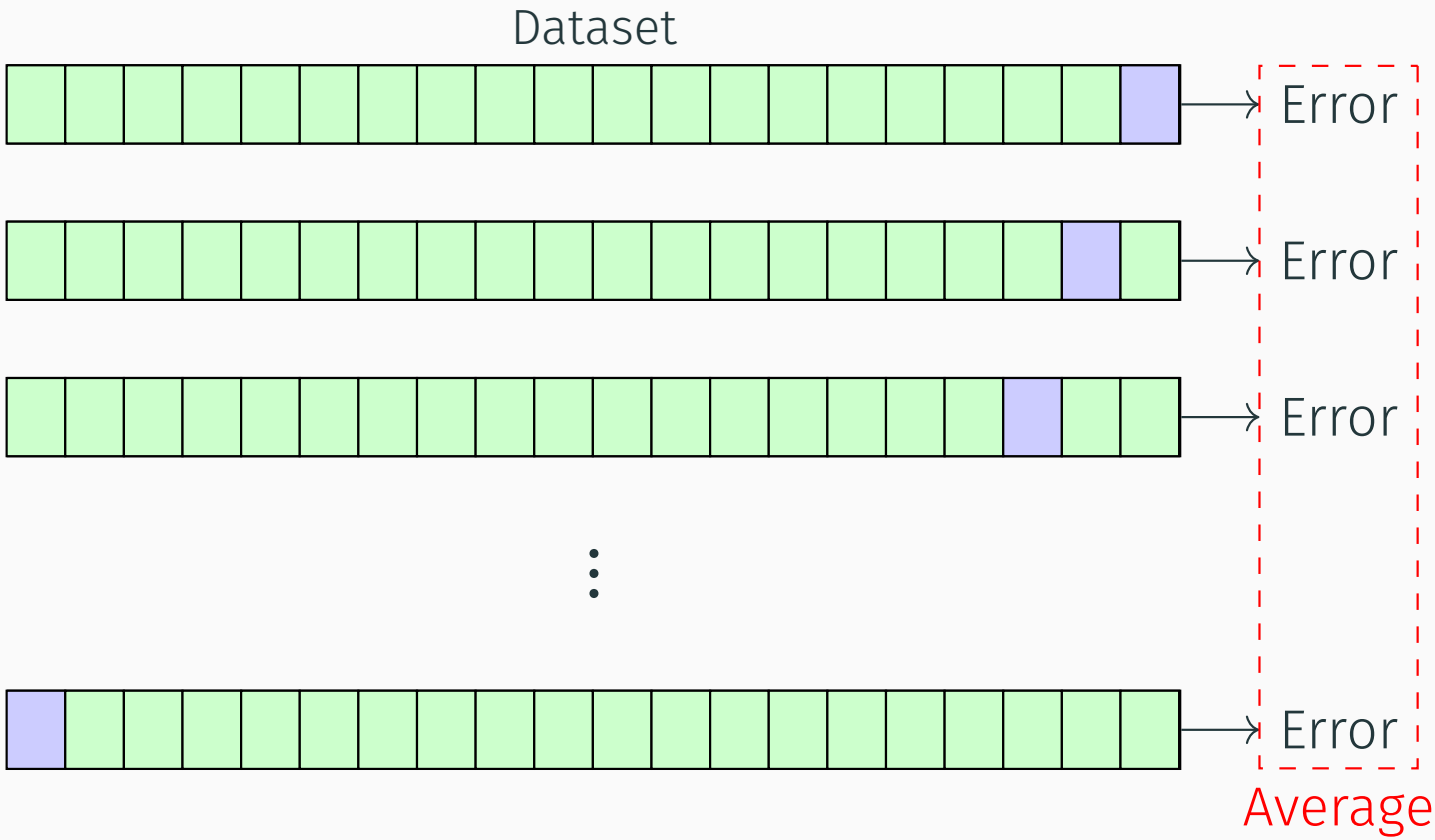
# Model evaluation: Leave-one-out cross-validation



# Model evaluation: Leave-one-out cross-validation



# Model evaluation: Leave-one-out cross-validation



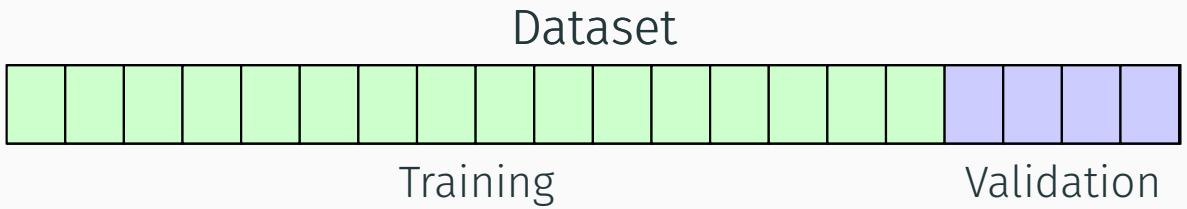
# Model evaluation: Leave-one-out cross-validation

Fits $n$ models for $n$ datapoints, each time leaving a single datapoint out for testing.

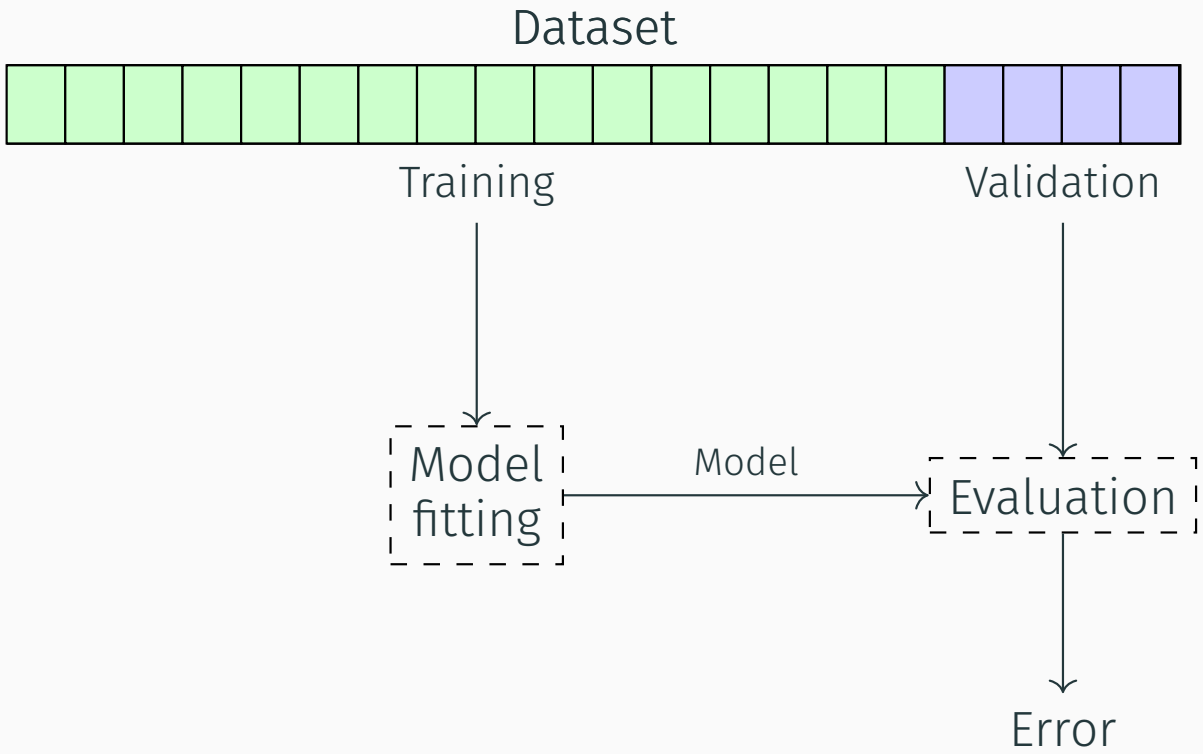
+ Uses all data to train models
+ Not dependent on arbitrary data splits
+ Unbiased (with regards to the full dataset)
- Computationally expensive
- Effectively gives a point estimate of the error
- All models are going to be trained on $>$ 99% overlapping data $\rightarrow$ highly correlated

# Model evaluation: Cross-validation

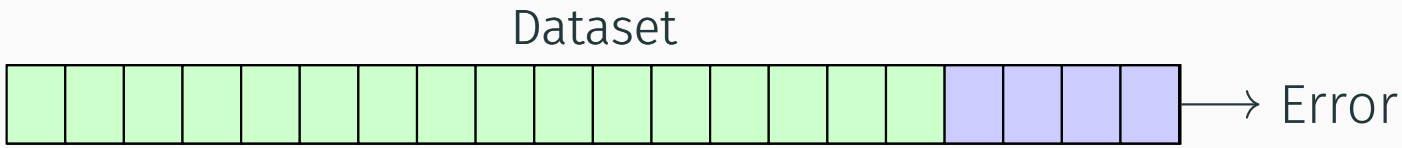
# Model evaluation: Cross-validation



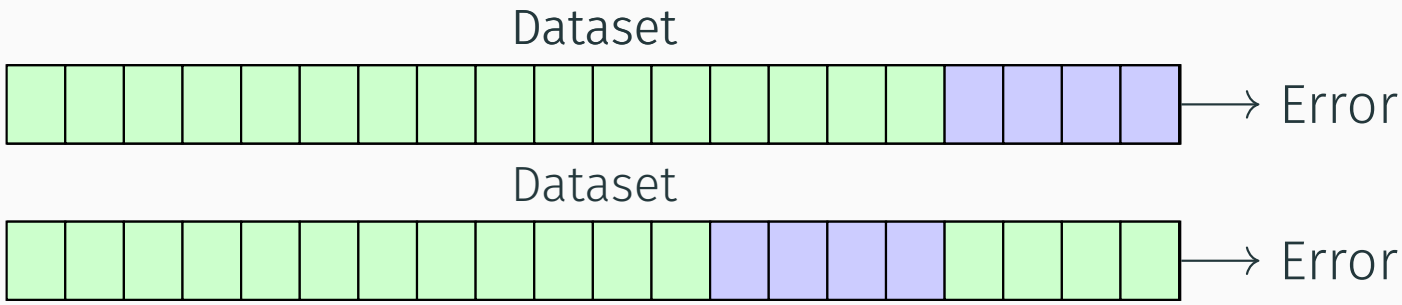
# Model evaluation: Cross-validation



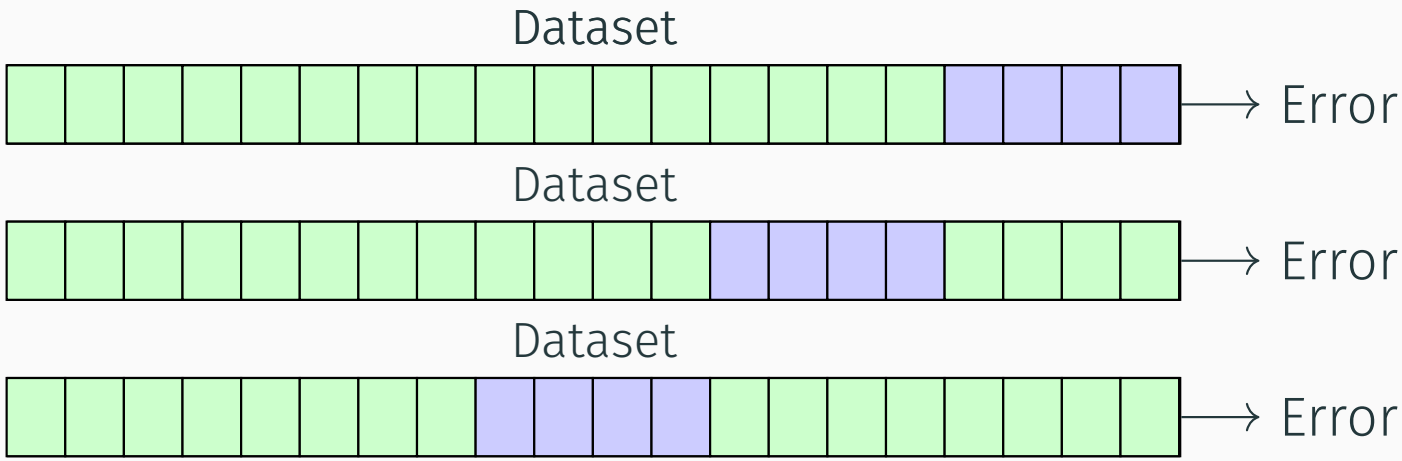
# Model evaluation: Cross-validation



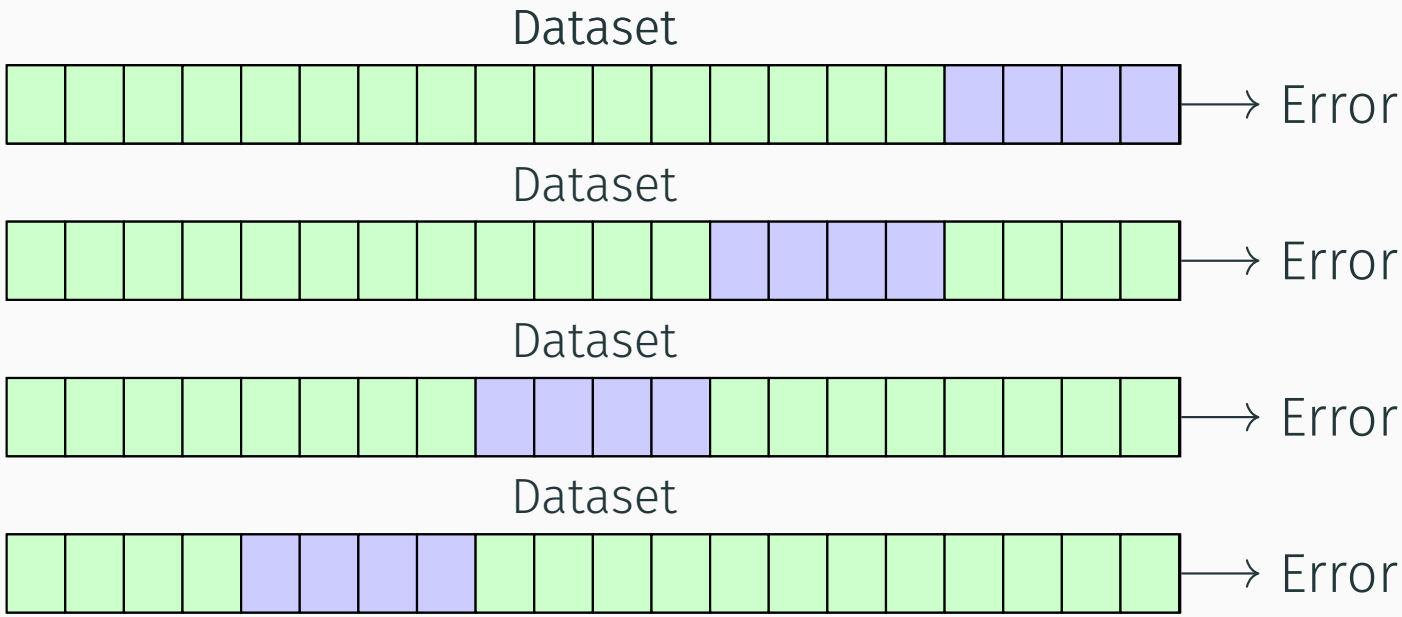
# Model evaluation: Cross-validation



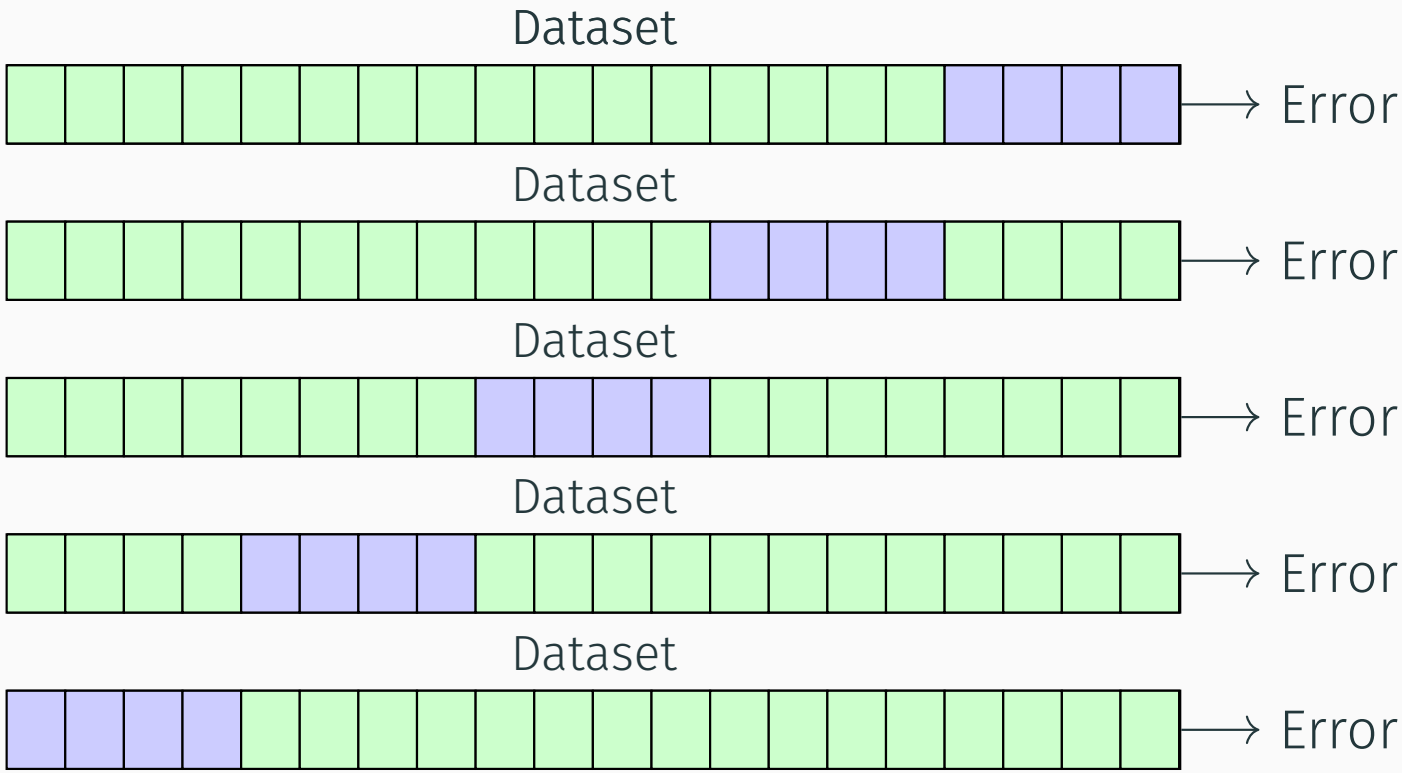
# Model evaluation: Cross-validation



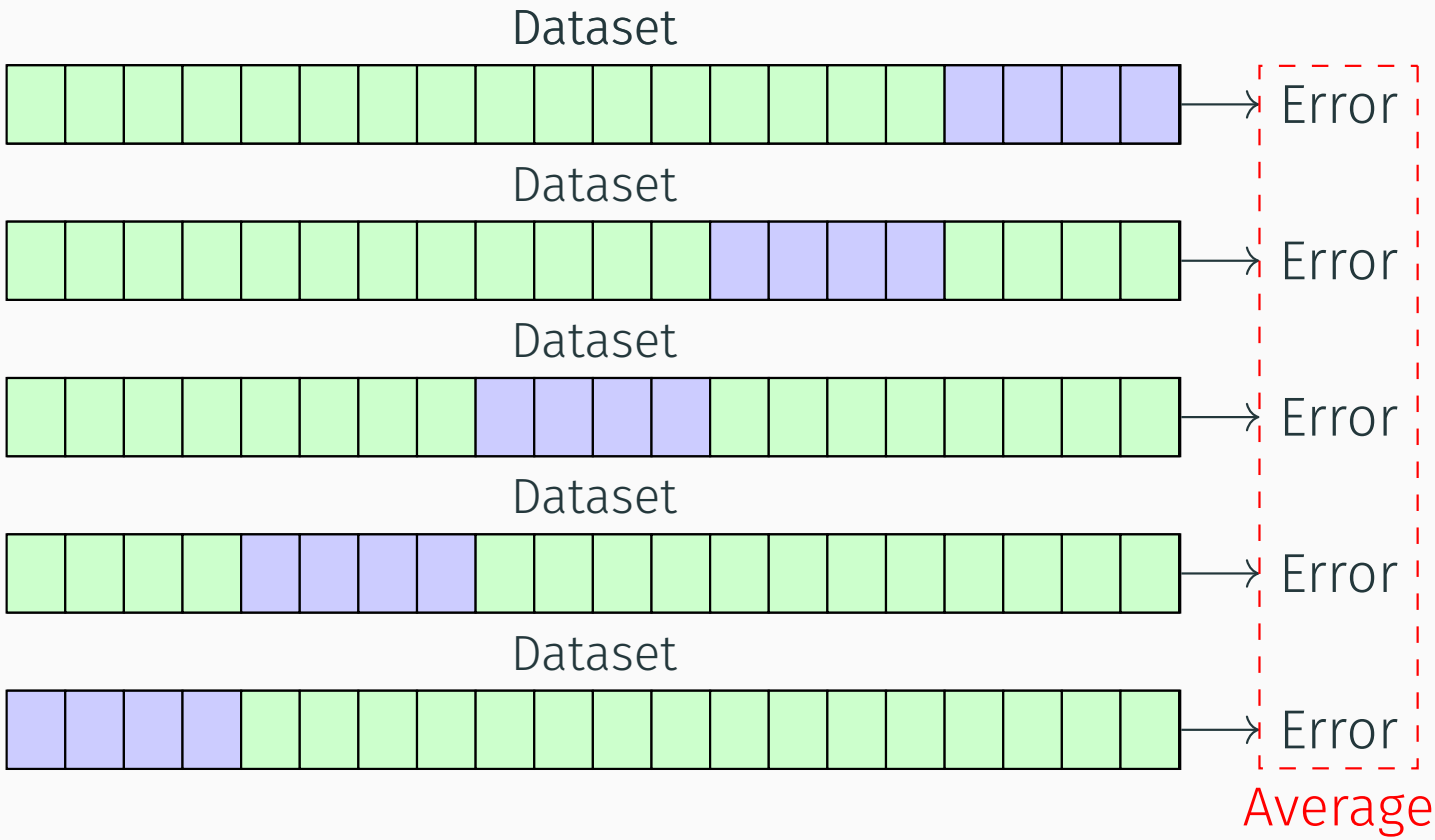
# Model evaluation: Cross-validation



# Model evaluation: Cross-validation



# Model evaluation: Cross-validation



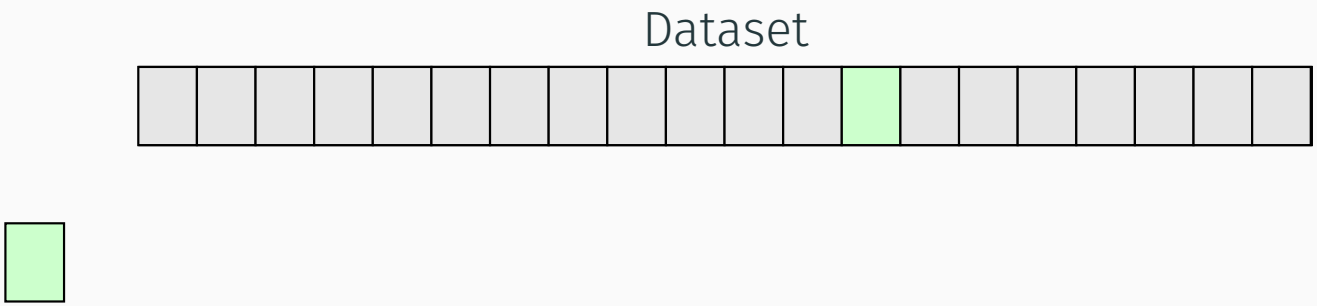
# Model evaluation: Cross-validation

Fits $k$ (usually $k \in \{5, 10\}$) models for $n > k$ datapoints, each leaving $n/k$ datapoints for out-of-sample testing.

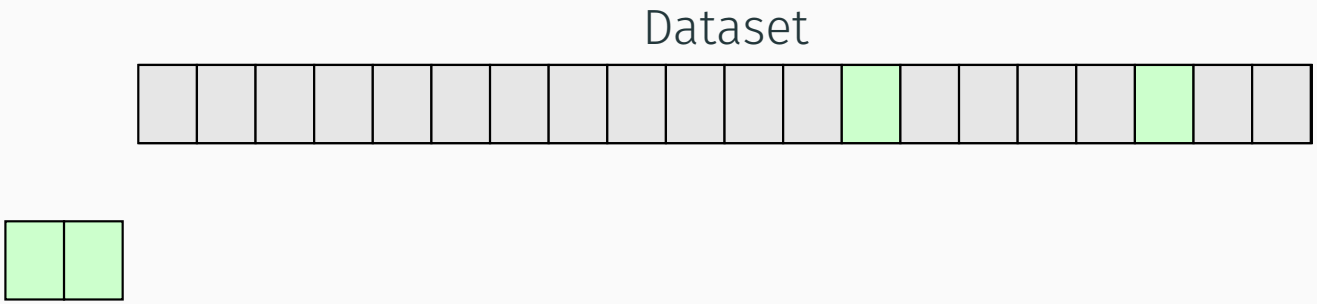
+ Uses all data to train models
+ Yields multiple estimates of out-of-sample error
- Different choices of $k$ (and exact splits) yields different results
- **No longer a single model from which information (e.g. parameter estimates and p-values) can be derived**

# Model evaluation: Bootstrapping

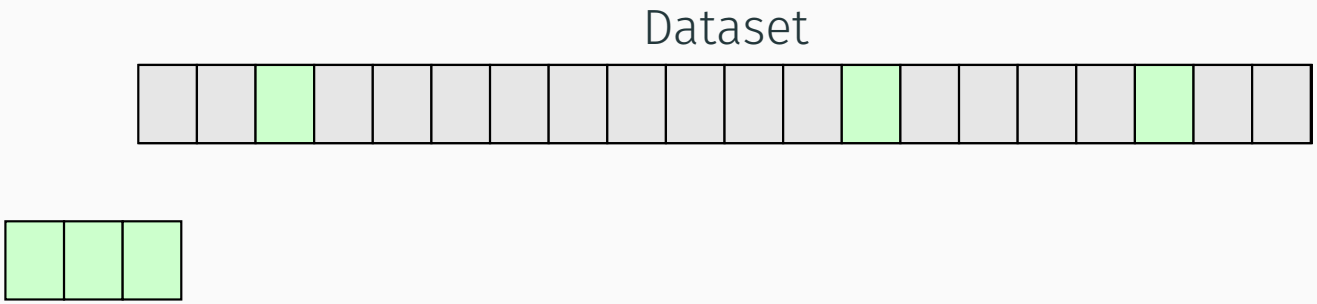
# Model evaluation: Bootstrapping



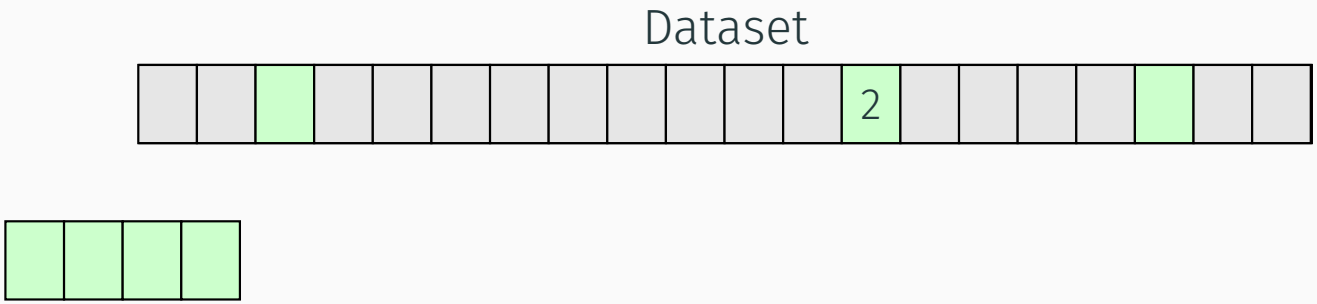
# Model evaluation: Bootstrapping



# Model evaluation: Bootstrapping

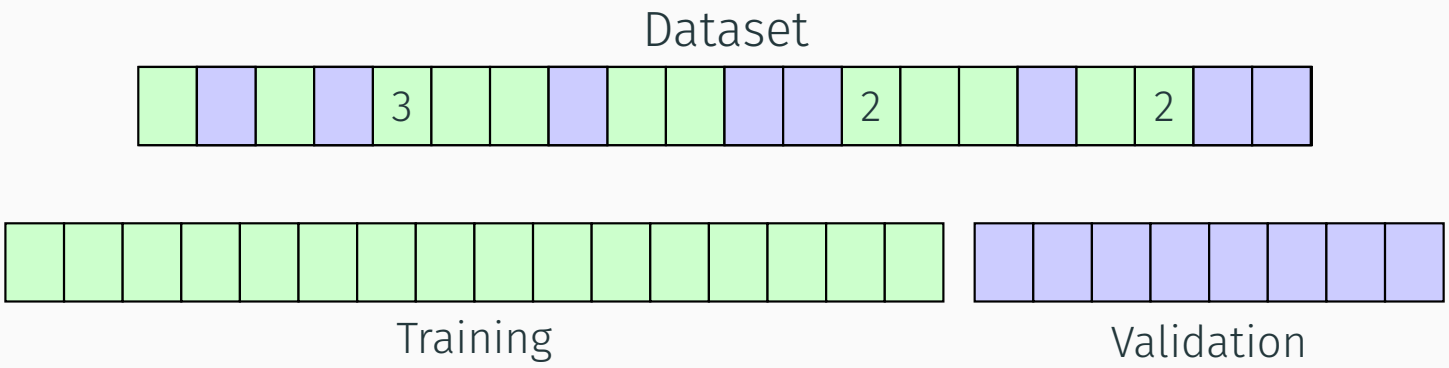


# Model evaluation: Bootstrapping

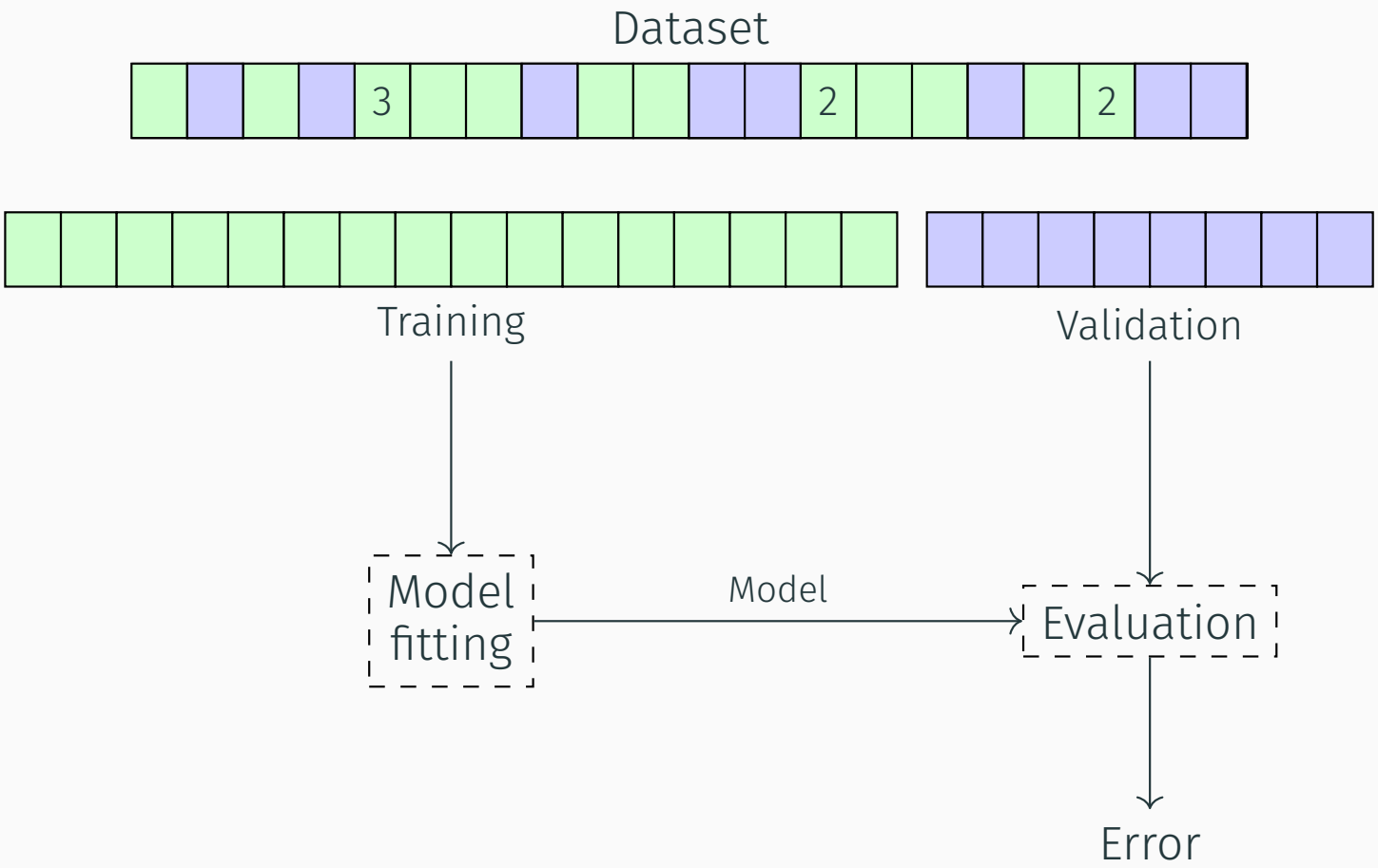


# Model evaluation: Bootstrapping

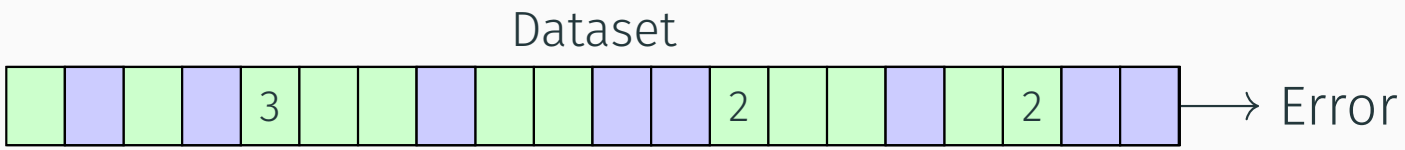
# Model evaluation: Bootstrapping



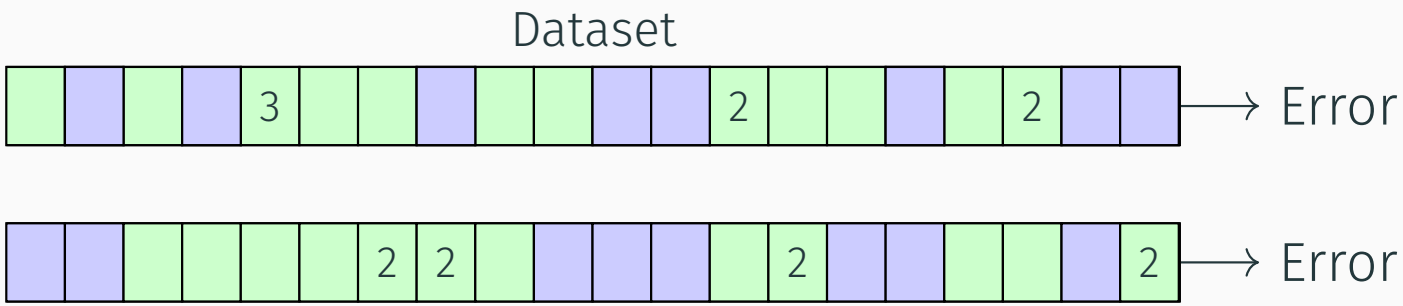
# Model evaluation: Bootstrapping



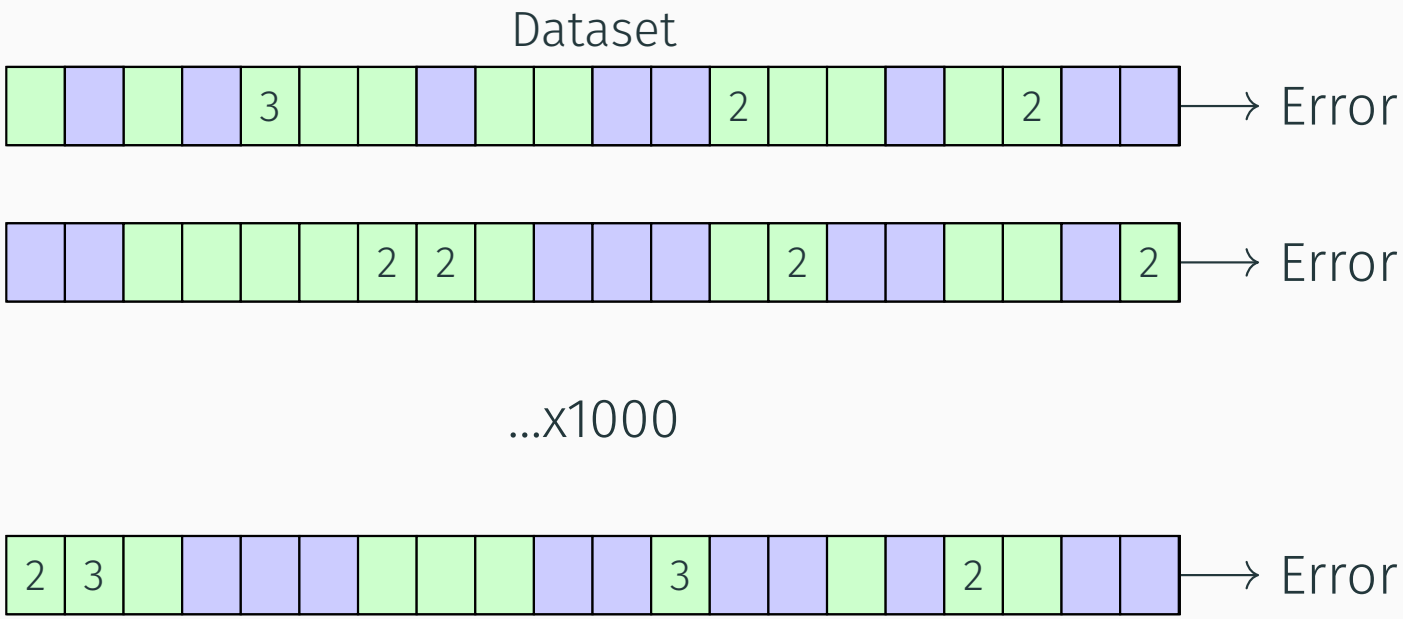
# Model evaluation: Bootstrapping



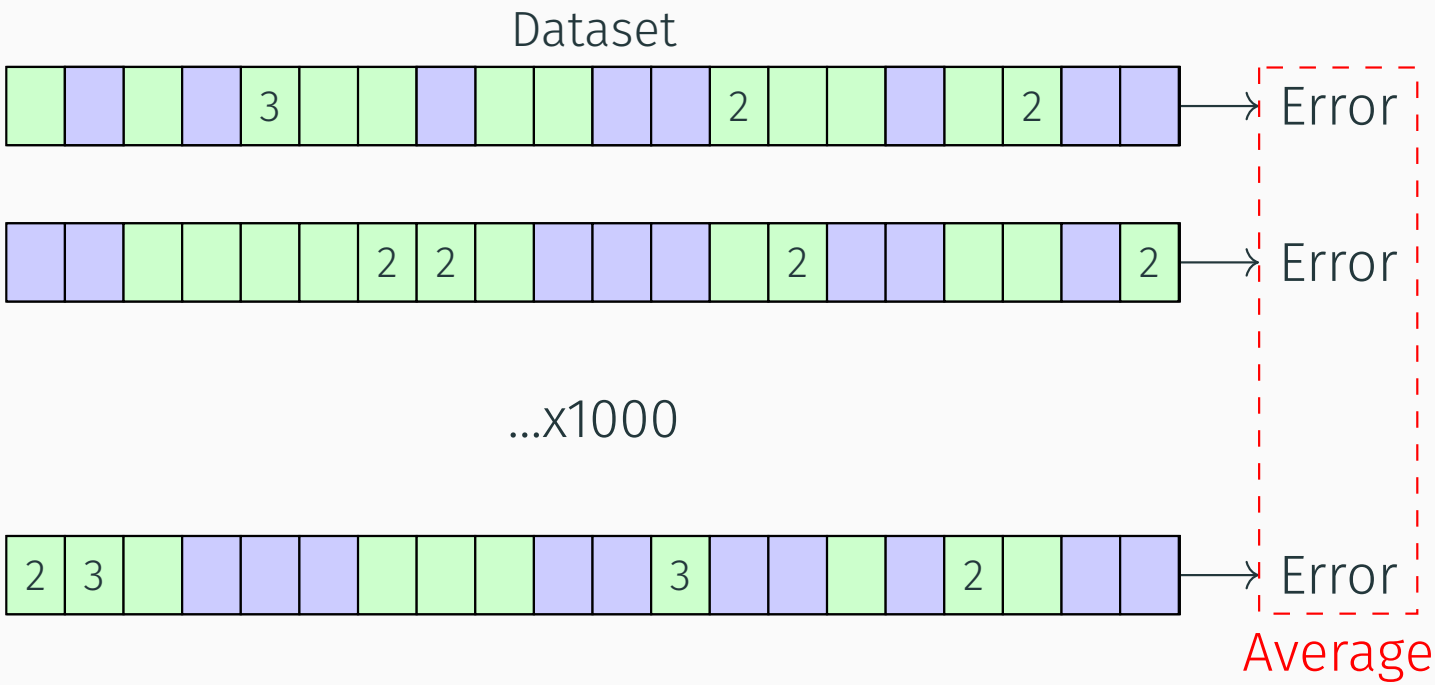
# Model evaluation: Bootstrapping



# Model evaluation: Bootstrapping



# Model evaluation: Bootstrapping



# Model evaluation: Bootstrapping

Fits $b$ models with $m$ datapoints (typically $m < n$), sampled from the original dataset **with replacement**.

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

# Model evaluation: Bootstrapping