



Machine Learning 1

Lecture 4.1 - Supervised Learning
Model Selection

Erik Bekkers

(Bishop 1.3)



Supervised Learning: Evaluating Errors

Question 1:

How can we estimate the model performance properly for unknown data?

Question 2:

How can we choose the optimal hyperparameters?

Supervised Learning: Evaluating Errors

Solution

Divide data $D = \{(\mathbf{x}_1, t_1), \dots, (\mathbf{x}_N, t_N)\}$ in 3 groups:

- ▶ Training set D_{train} (± 80 % of D):

- ▶ Minimize the error $E(y(\mathbf{x}), t)$ for $(\mathbf{x}, t) \in D_{\text{train}} \rightarrow \mathbf{w}^*$

- ▶ Validation set D_{val} (± 10 % of D):

- ▶ Used to estimate test error $E(y(\mathbf{x}_{\text{val}}, \mathbf{w}^*), t_{\text{val}})$

- ▶ Test set D_{test} (± 10 % of D):

- ▶ final test/generalization error estimate $E(y(\mathbf{x}_{\text{test}}, \mathbf{w}^*), t_{\text{test}})$

Can never be part of model selection!!!

repeat for different

hyperparameter settings

model selection part

Supervised Learning: Small Datasets

- ▶ Small dataset → small validation and test set
- ▶ Approximate validation step!

Cross-validation

- Split data: $D = \{(x_1, t_1), \dots, (x_N, t_N)\}$ into K -folds
- Train y on $K-1$ folds $\hat{y}^{-k}(x)$

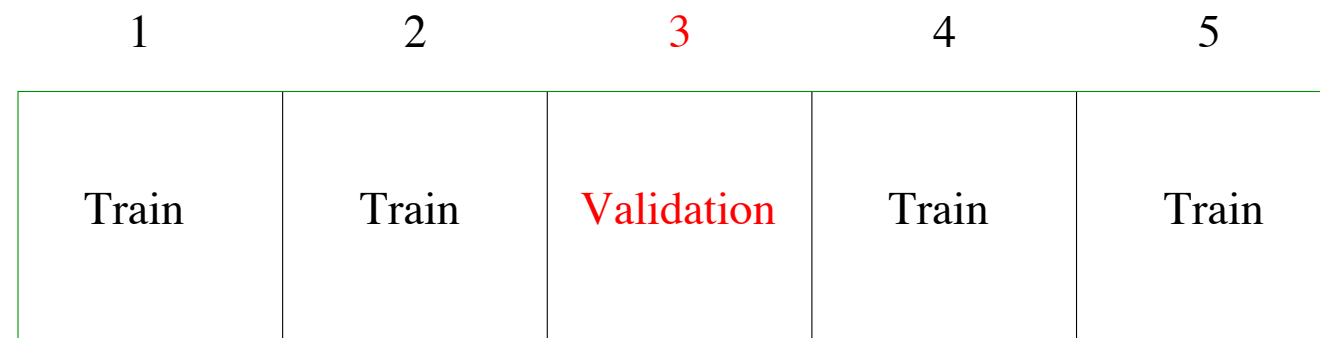


Figure: K-fold splitting of dataset (ESL 7.10)

- $K = N$: leave-one-out cross validation

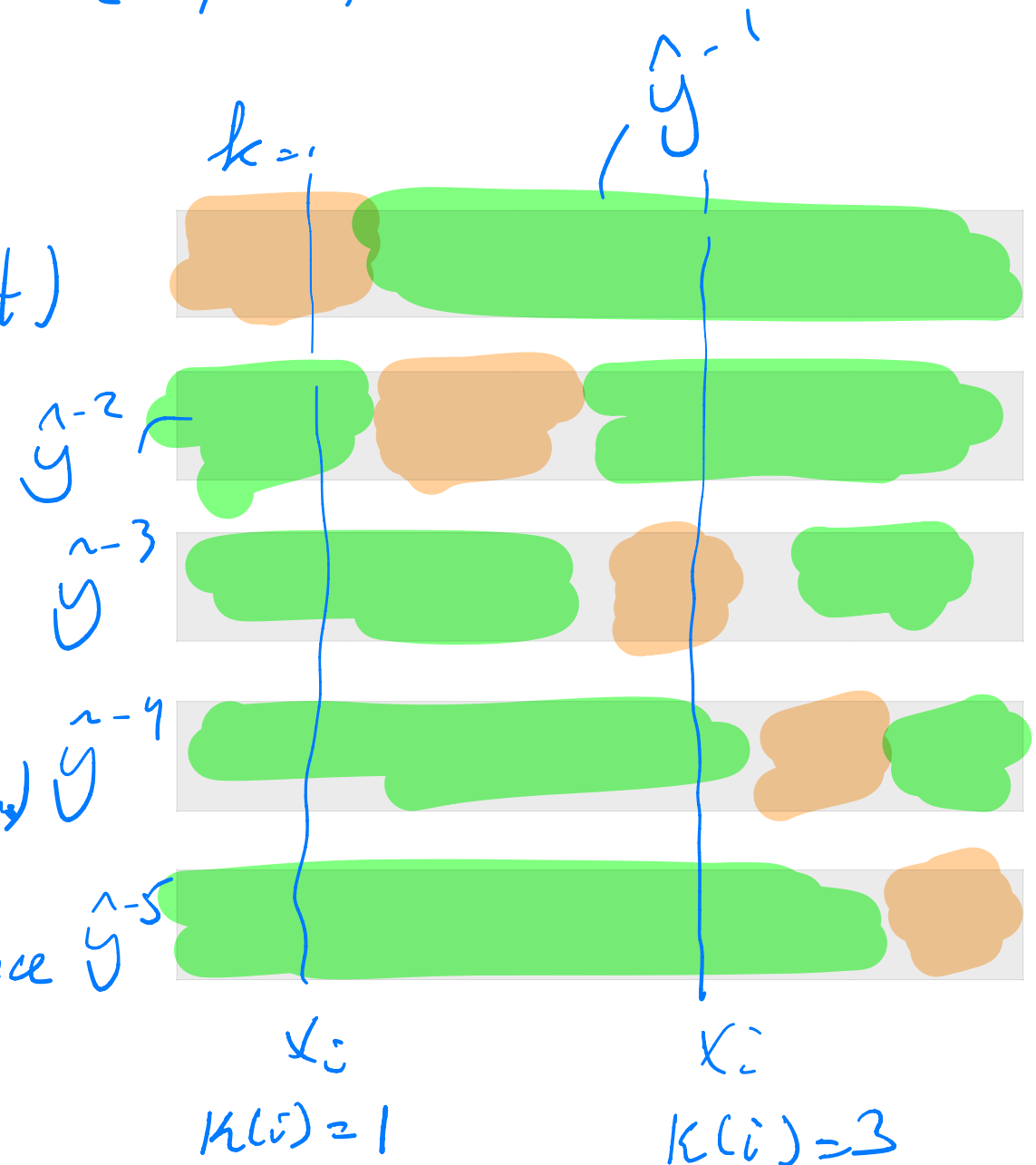
Cross-Validation

- K trained functions $\hat{y}^{-k}(x)$
- Indexing function $\kappa : \{1, \dots, N\} \mapsto \{1, \dots, K\}$
- Estimate of prediction error

$$CV(\hat{y}) = \frac{1}{N} \sum_{i=1}^N E(\hat{y}^{-\kappa(i)}(x_i), t)$$

2 tasks

1. model selection
(optimal hyperparameters)
2. estimate model performance



Cross-Validation: Model Selection

- Hyperparameter selection α

- $CV(\hat{y}_\alpha) = \frac{1}{N} \sum_{n=1}^N E(\hat{y}_\alpha(x_n), t_n)$

- Optimal $\alpha^* = \arg \min_{\alpha} CV(\hat{y}_\alpha)$

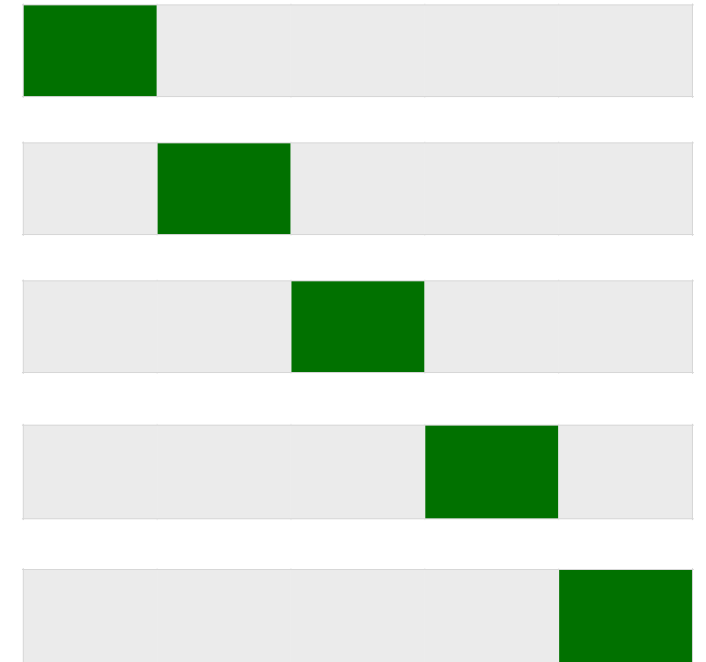
- Multiple hyperparameters: $\alpha \in \{\alpha_1, \alpha_2\}, \beta \in \{\beta_1, \beta_2, \beta_3\}$

- How many times should CV be performed?

$$2 \times 3$$

- Total number of training runs?

$$6K$$



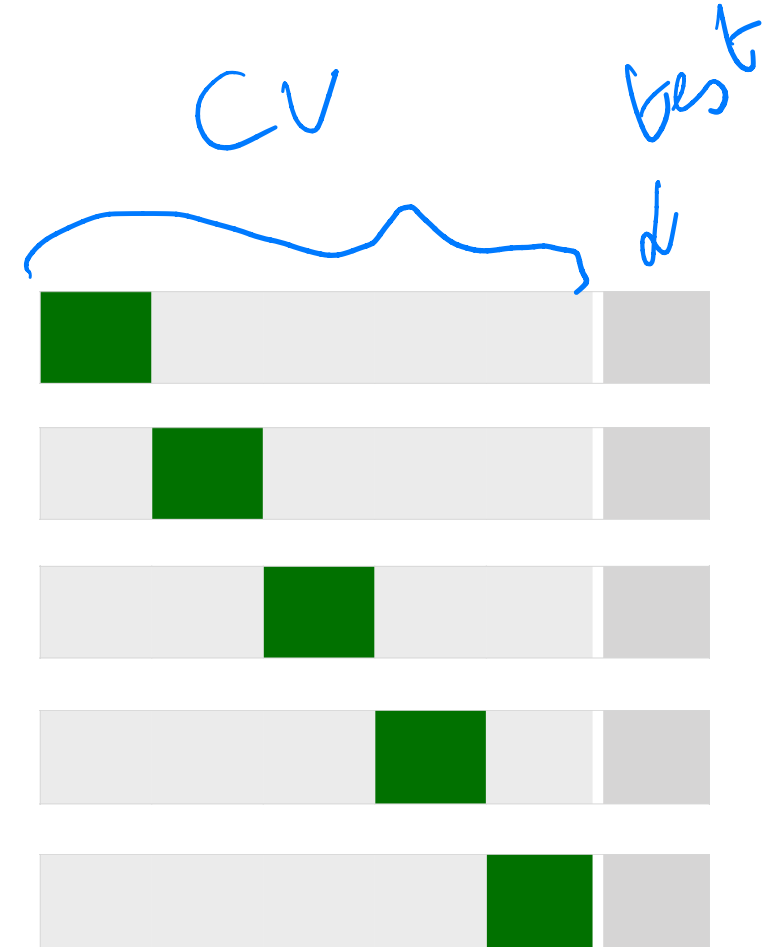
Cross-Validation: Test Error Estimation

- ▶ After Model selection

$$\alpha^*, \beta^*$$

- ▶ Retrain f on all K folds with α^*, β^*
- ▶ Evaluate model on held-out test set

- ▶ Nested cross validation!



Nested Cross-Validation

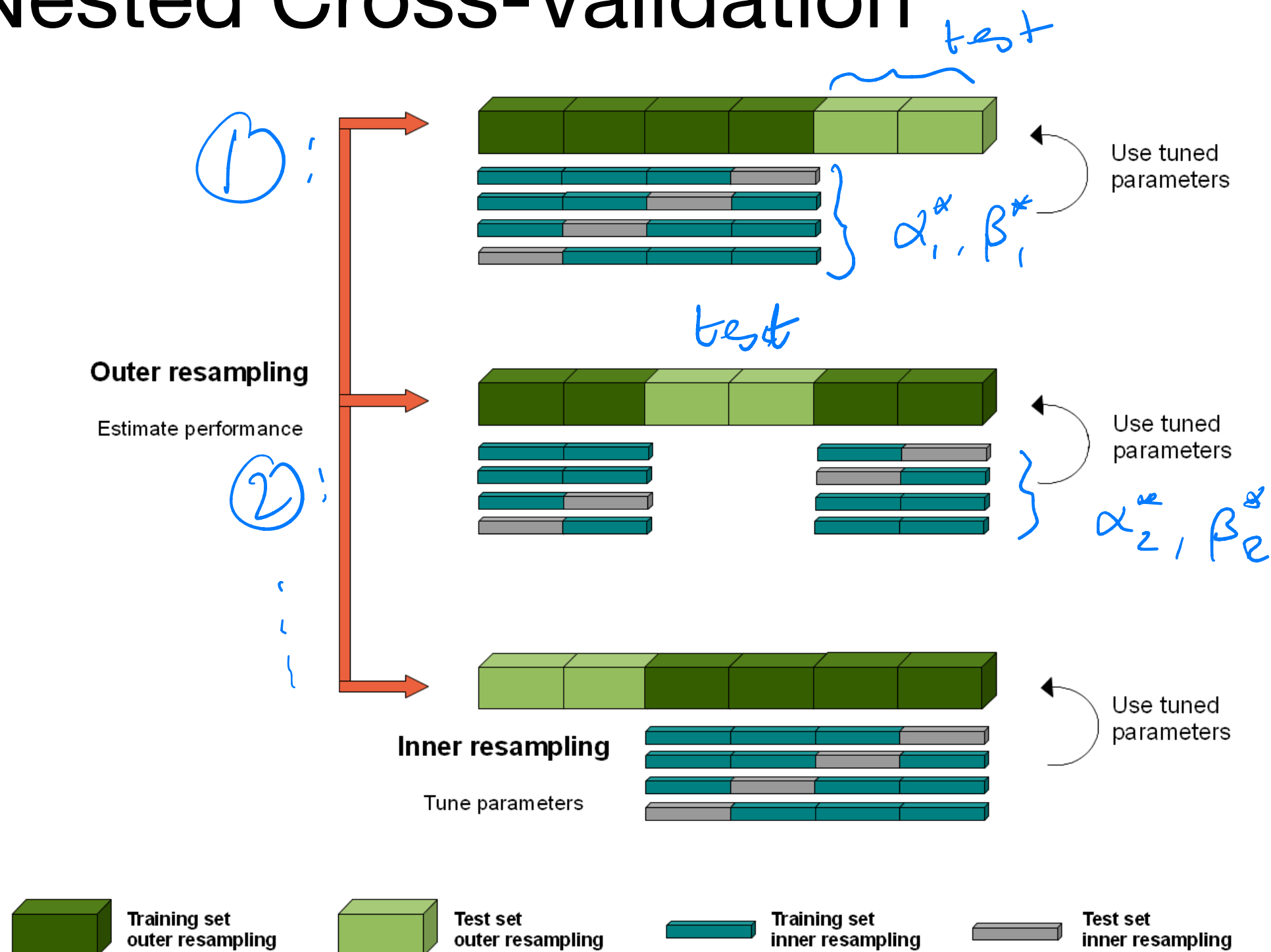


Figure: Nested cross-validation

https://mlr-org.github.io/mlr-tutorial/devel/html/nested_resampling/index.html