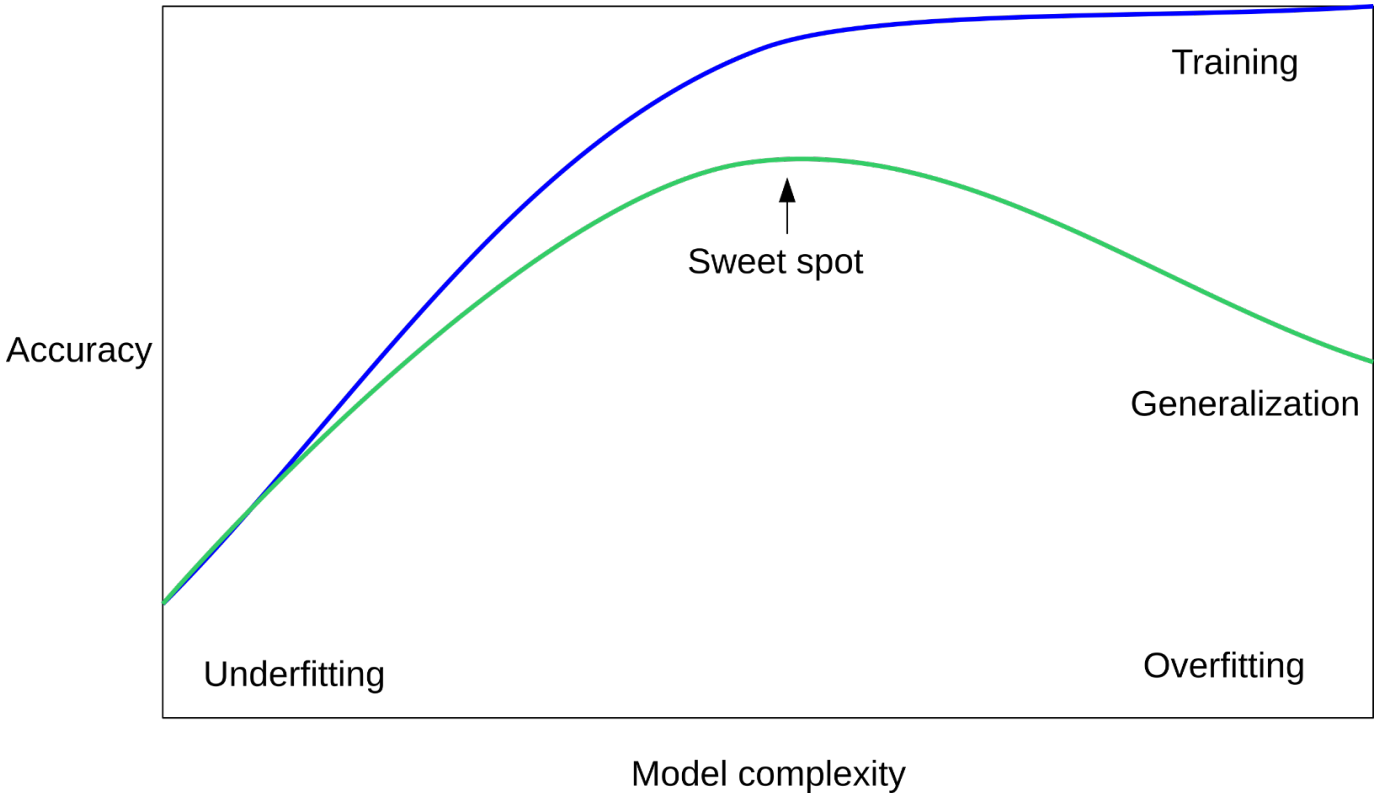
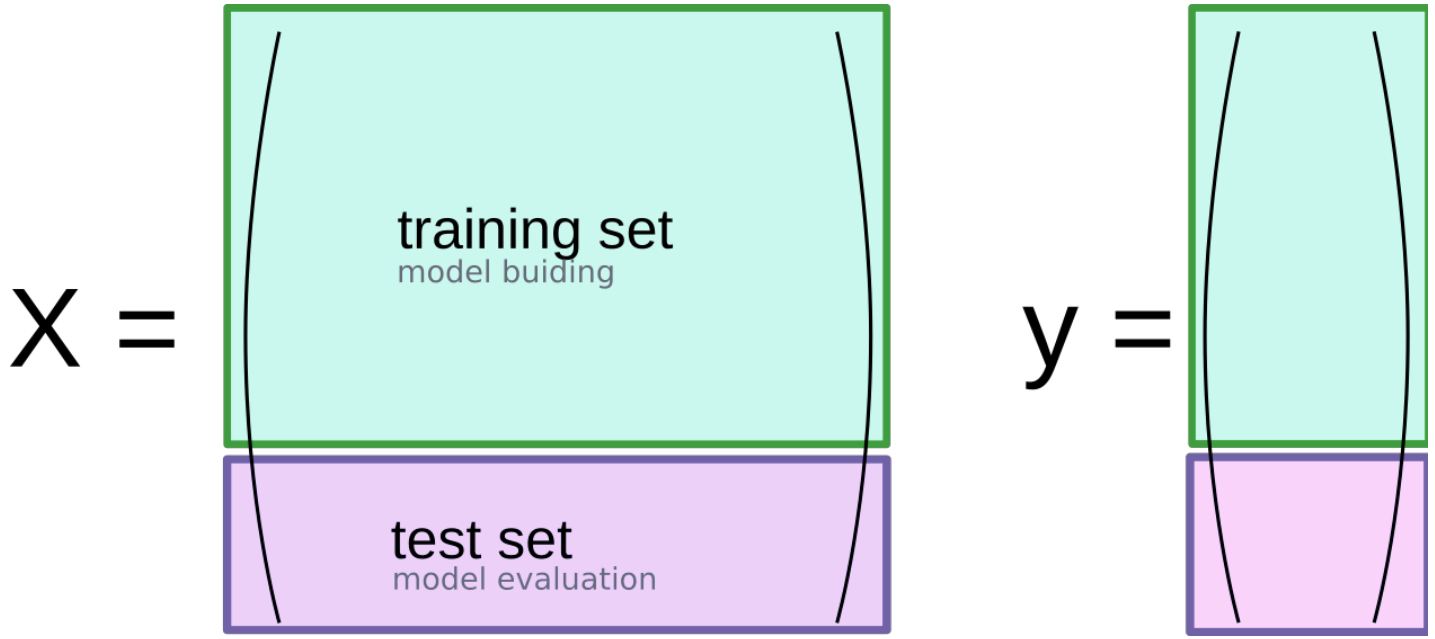
Train – Validate – Test

Part I

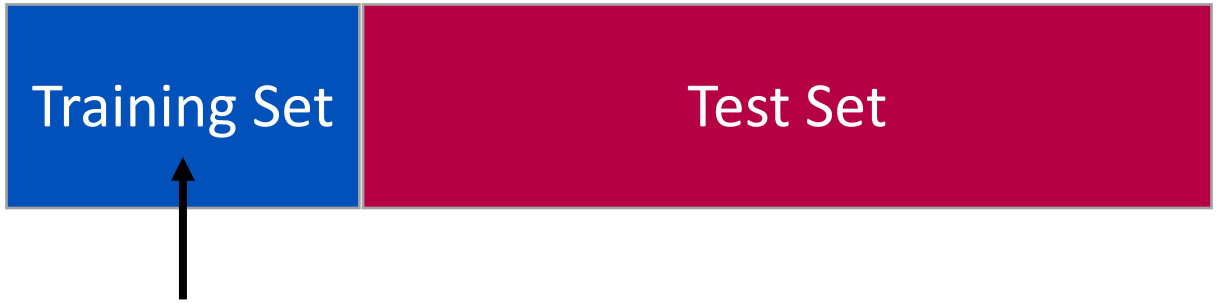
Overfitting and Underfitting



Train – Test – Split: Part I

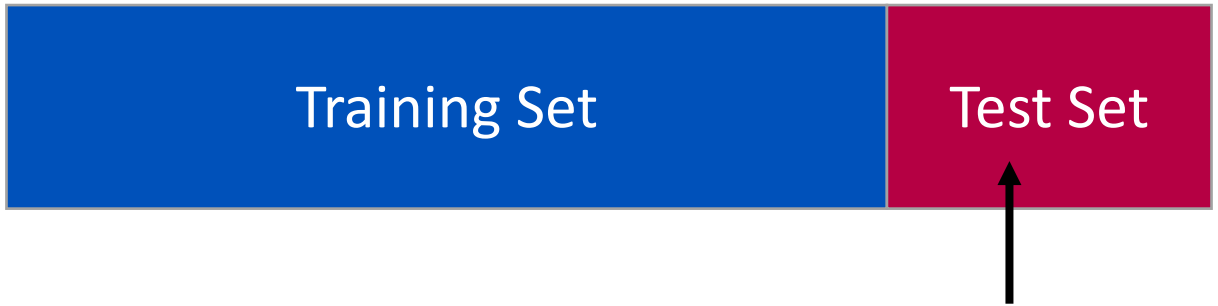


Train – Test – Split: Part II



Too few → $\hat{w}$ poorly estimated

Train – Test – Split: Part III



Too few → test error is a bad approximation of generalization of errors.

Train – Test – Split: Part IV

Typically, just enough test points to form a reasonable estimate of generalization of errors.

Machine Learning Workflow: Part I

1. **Model selection**
   - Often need to **choose tuning parameters** λ controlling model complexity (e.g., degree of polynomial)
2. **Model assessment**
   - Having selected a model, **assess the generalization of error**.

Hypothetical Implementation: Part I

1. Model selection

   For each considered model complexity λ:

   i. Estimate parameters $\widehat{w}_\lambda$ on **training data**

   ii. Assess performance of $\widehat{w}_\lambda$ on **test data**

   iii. Choose λ^* to be λ with lowest test error

2. Model assessment

   Compute test error of $\widehat{w}_{\lambda^*}$ (fitted model for selected complexity λ^*) to approximate generalization error.

Hypothetical Implementation: Part II

1. Model selection

   For each considered model complexity λ:

   i. Estimate parameters $\widehat{w}_\lambda$ on **training data**

   ii. Assess performance of $\widehat{w}_\lambda$ on **test data**

   iii. Choose λ^* to be λ with lowest test error

2. Model assessment

   Overly optimistic!

   Compute test error of $\widehat{w}_{\lambda^*}$ (fitted model for selected complexity λ^*) to approximate generalization error.

Hypothetical Implementation: Part III

Issue: Just like fitting $\widehat{w}$ and assessing its performance on training data.

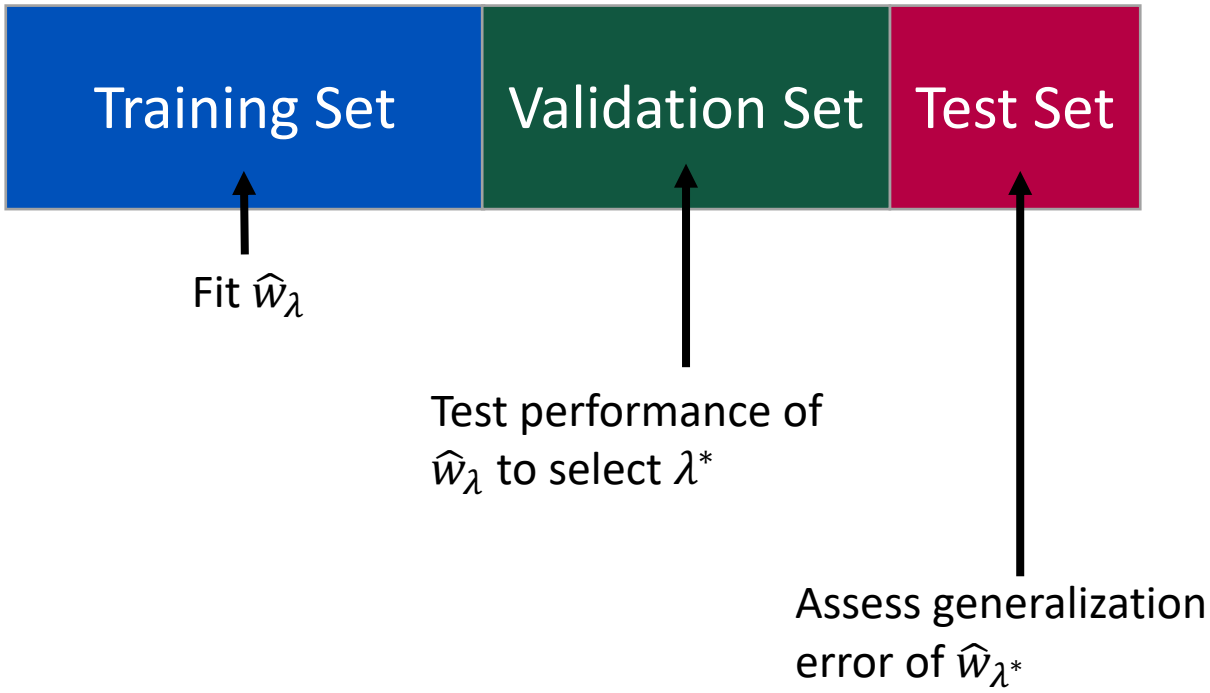
- λ^* was selected to minimize test error (i.e., λ^* was fit on test data)
- If test data is not representative of the whole world, then $\widehat{w}_{\lambda^*}$ will typically perform worse than test error indicates.

Practical Implementation

Solution: Create two "test" sets!

1. Select λ^* such that $\widehat{w}_{\lambda^*}$ minimizes error on validation set
2. Approximate generalization error of $\widehat{w}_{\lambda^*}$ using test set

Practical Implementation, Continued



Typical Splits

80%
50%

10%
25%

10%
25%