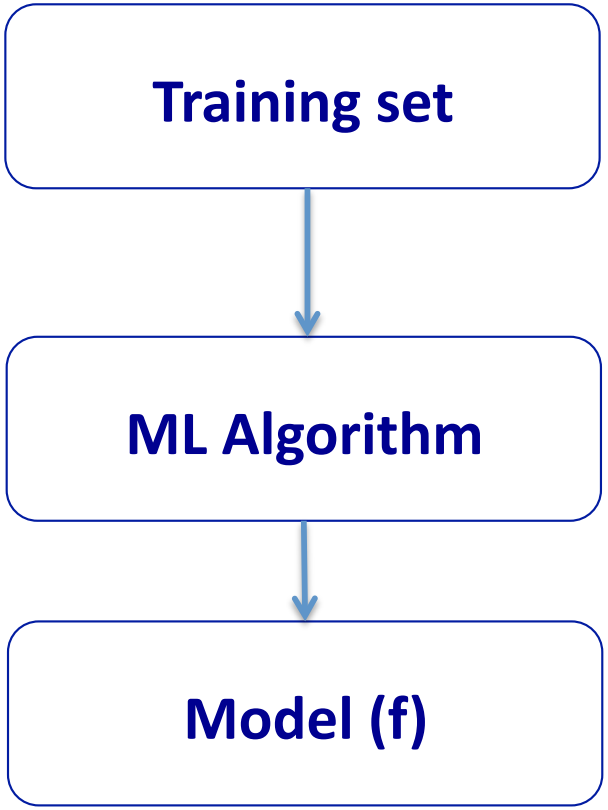
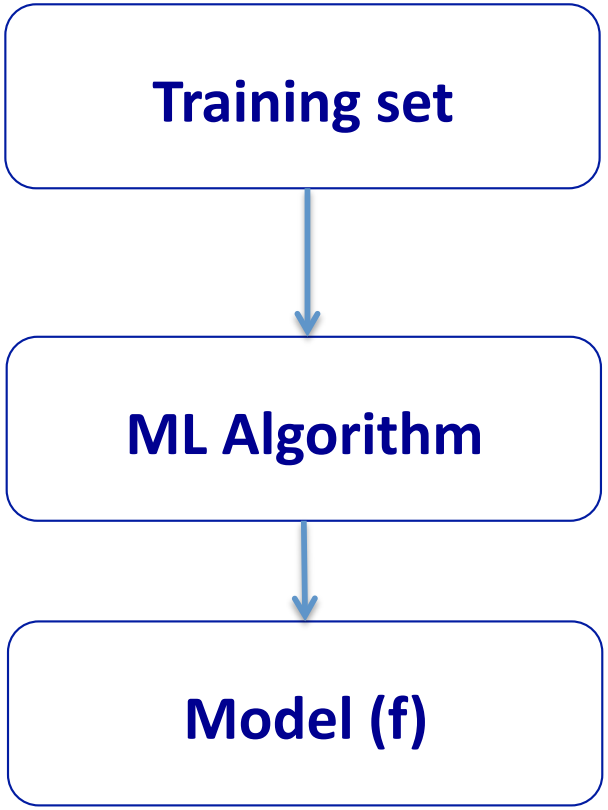
Training and Testing



Question: How can we test f?

Training and Testing



Question: How can we be confident about f?

Training and Testing

- We calculate E^{train}, the in-sample error (training error or empirical error/risk)

$$E^{train}(f) = \sum_{i=1}^{n} \ell oss(y_i, f(\vec{x_i}))$$

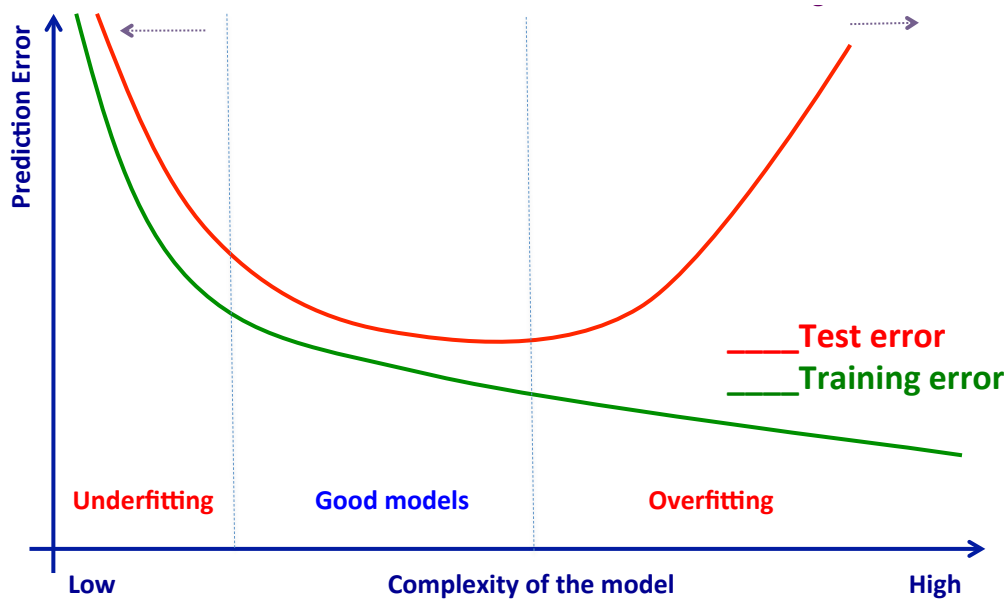
Training and Testing

- We calculate E^{train}, the in-sample error (training error or empirical error/risk)

$$E^{train}(f) = \sum_{i=1}^{n} \ell oss(y_i, f(\vec{x_i}))$$

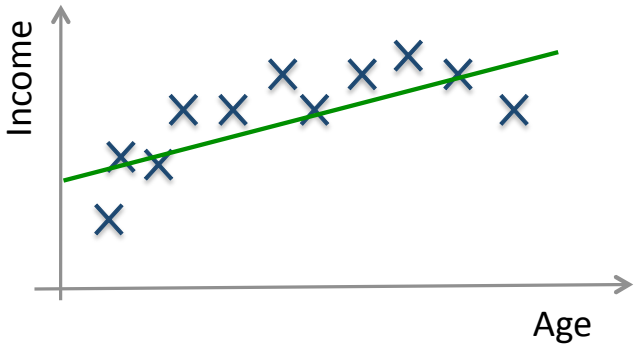
- We aim to have $E^{train}(f)$ small (i.e., minimize $E^{train}(f)$)
- We hope that $E^{test}(f)$, the out-sample error (test/true error), will be small, too.

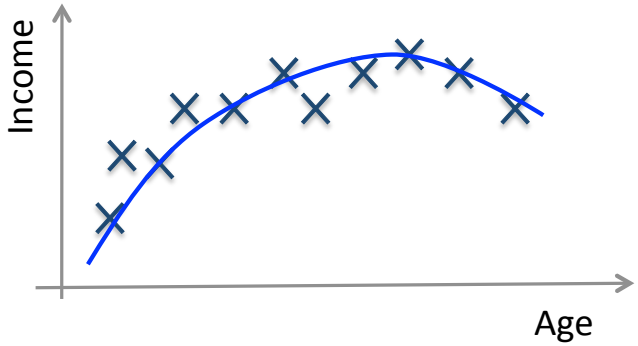
Overfitting and Underfitting



Balancing the fit to training data against the complexity of the model is key in machine learning (ML).

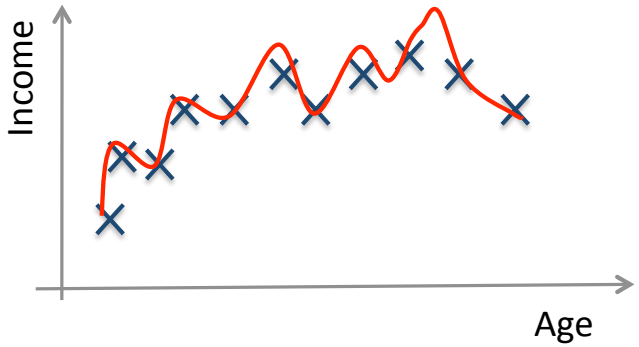
Overfitting and Underfitting





High bias (underfitting)

Just right!



High variance (overfitting)

In general, to avoid overfitting, use simple models.

Train, Validation, and Test

1. The training set is a set of examples used for learning a model (e.g., a classification model).
2. The validation set is a set of examples that cannot be used for learning the model but can help tune model parameters
3. The test set is used to assess the performance of the final model and provide an estimation of the test error.

Train, Validation, and Test

1. The training set is a set of examples used for learning a model (e.g., a classification model).
2. The validation set is a set of examples that cannot be used for learning the model but can help tune model parameters
3. The test set is used to assess the performance of the final model and provide an estimation of the test error.

   Note: Never use the test set in any way to further tune the parameters or revise the model.

Confusion Matrix

		Actual Label	
		Positive	Negative
Predicted Label	Positive	**True Positive (TP)**	**False Positive (FP)**
	Negative	**False Negative (FN)**	**True Negative (TN)**

Evaluation Metrics

		Actual Label	
		Positive	Negative
Predicted Label	Positive	**True Positive (TP)**	**False Positive (FP)**
	Negative	**False Negative (FN)**	**True Negative (TN)**

Accuracy	(TP + TN) / (TP + TN + FP + FN)	The percentage of predictions that are correct
Precision	TP / (TP + FP)	The percentage of positive predictions that are correct
Sensitivity (Recall)	TP / (TP + FN)	The percentage of positive cases that were predicted as positive
Specificity	TN / (TN + FP)	The percentage of negative cases that were predicted as negative