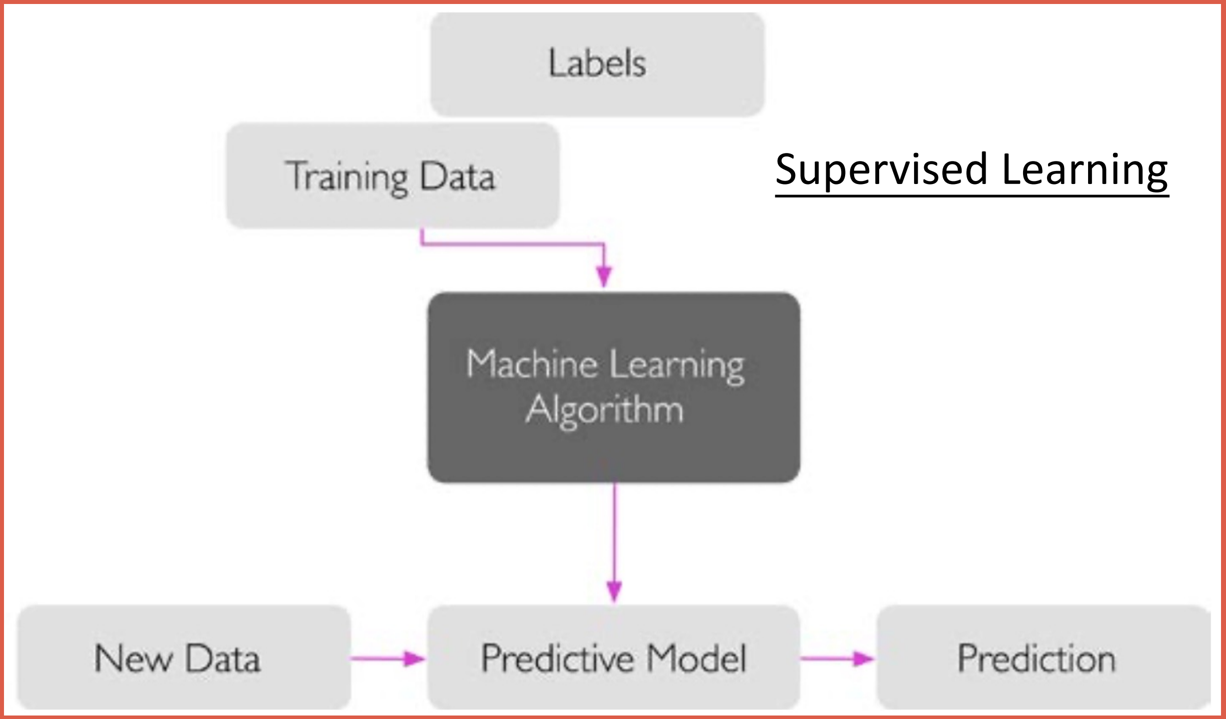
Fundamentals and Landscape of Classical Machine Learning (III)

- Learn the sources of errors when you apply your ML model for new data
- Learn the gold standard: evaluate the performance of machine learning algorithms with cross validation

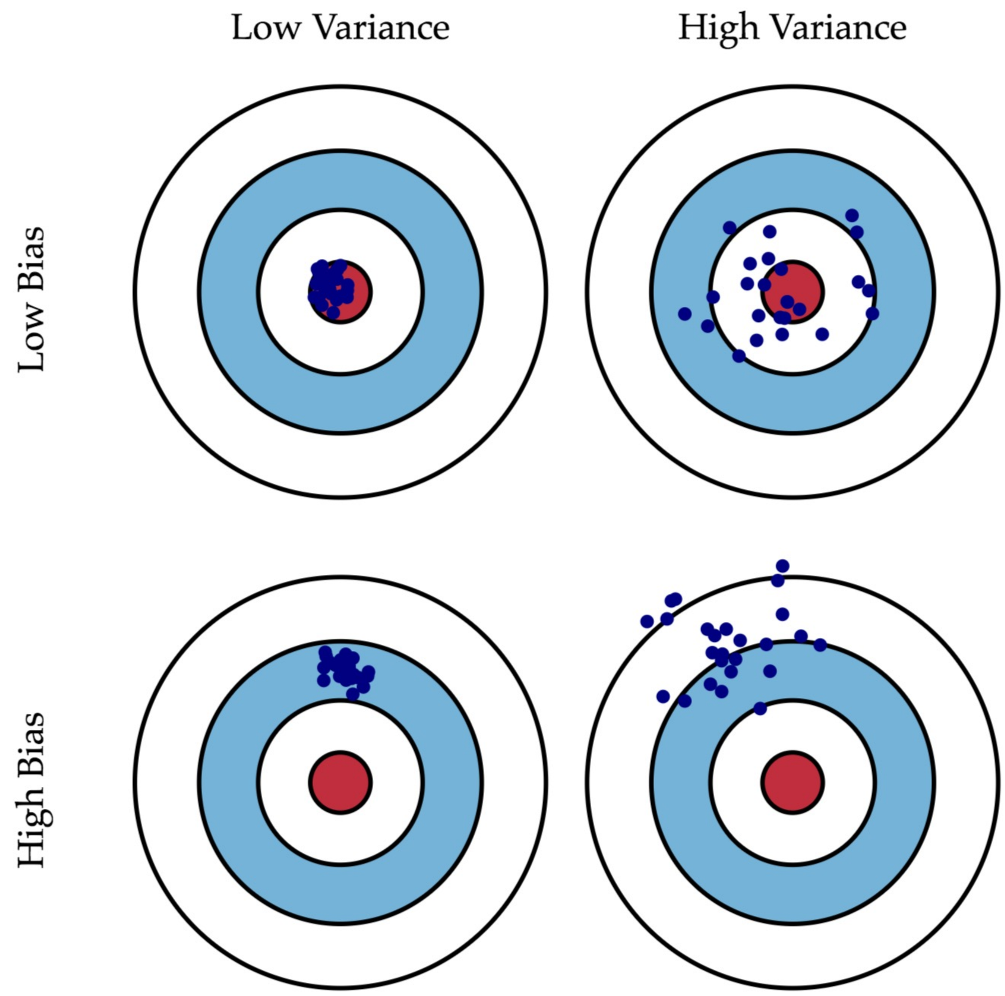
https://www.sli.do/
#073374

Learn the sources of errors when you apply your ML model for new data



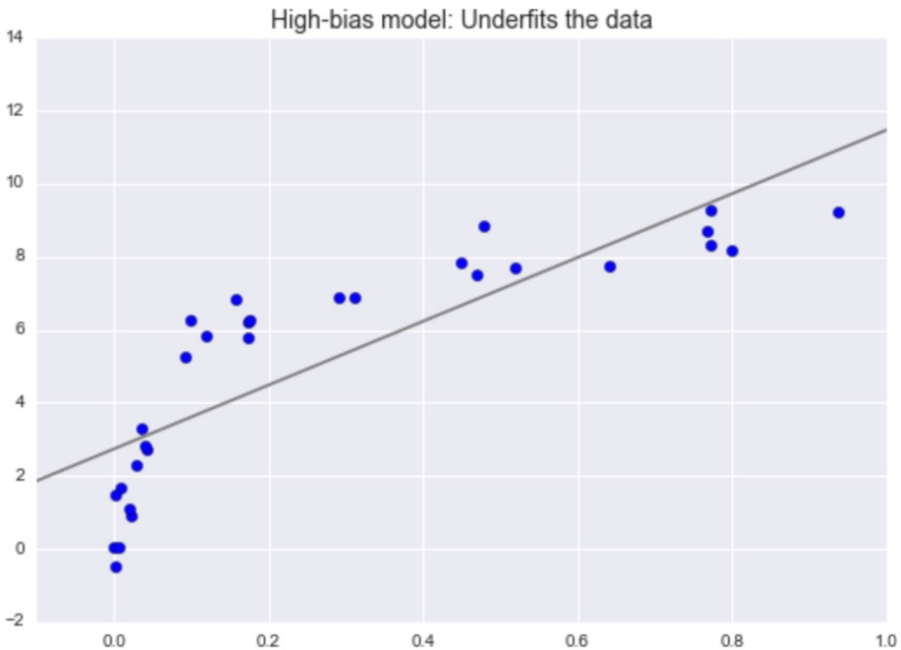
Supervised Learning

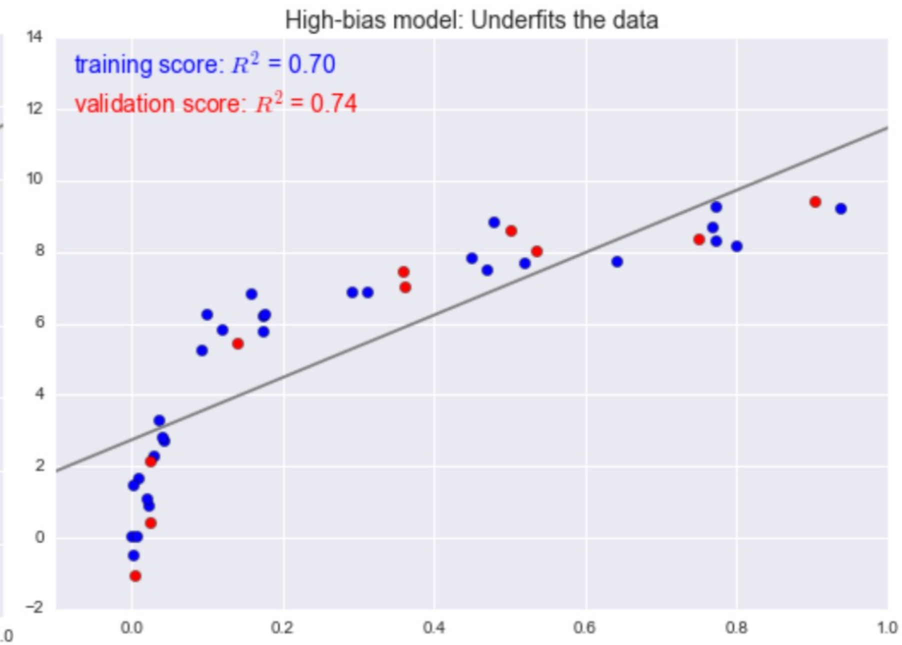
Bias and **variance** are two major sources of errors **that prevent supervised learning algorithms from generalizing beyond their training set**



Intuitive **High Bias** Example: low-order polynomial fitting (**underfit**)

$p = 1$





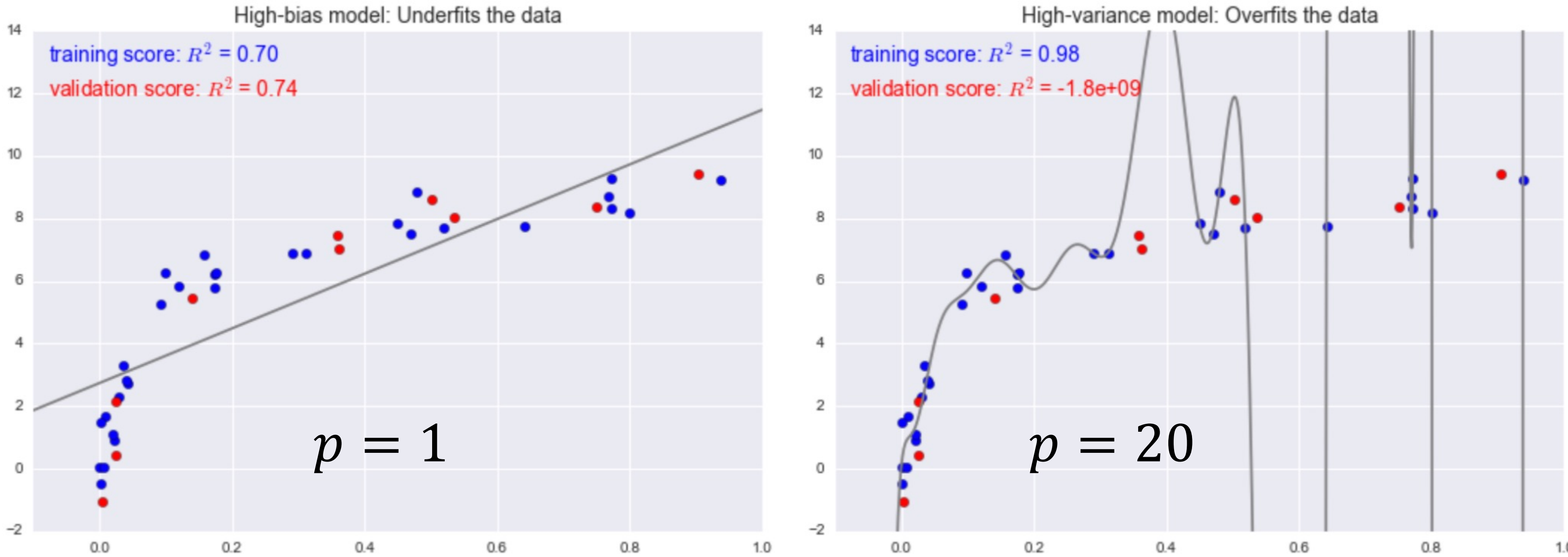
$R^2 = 1$ indicates a perfect match, $R^2 = 0$ indicates the model does no better than simply taking the mean of the data, and negative values mean even worse models.

Intuitive **High Variance** Example: high-order polynomial fitting (**overfit**)

$p = 20$

$R^2 = 1$ indicates a perfect match, $R^2 = 0$ indicates the model does no better than simply taking the mean of the data, and negative values mean even worse models.

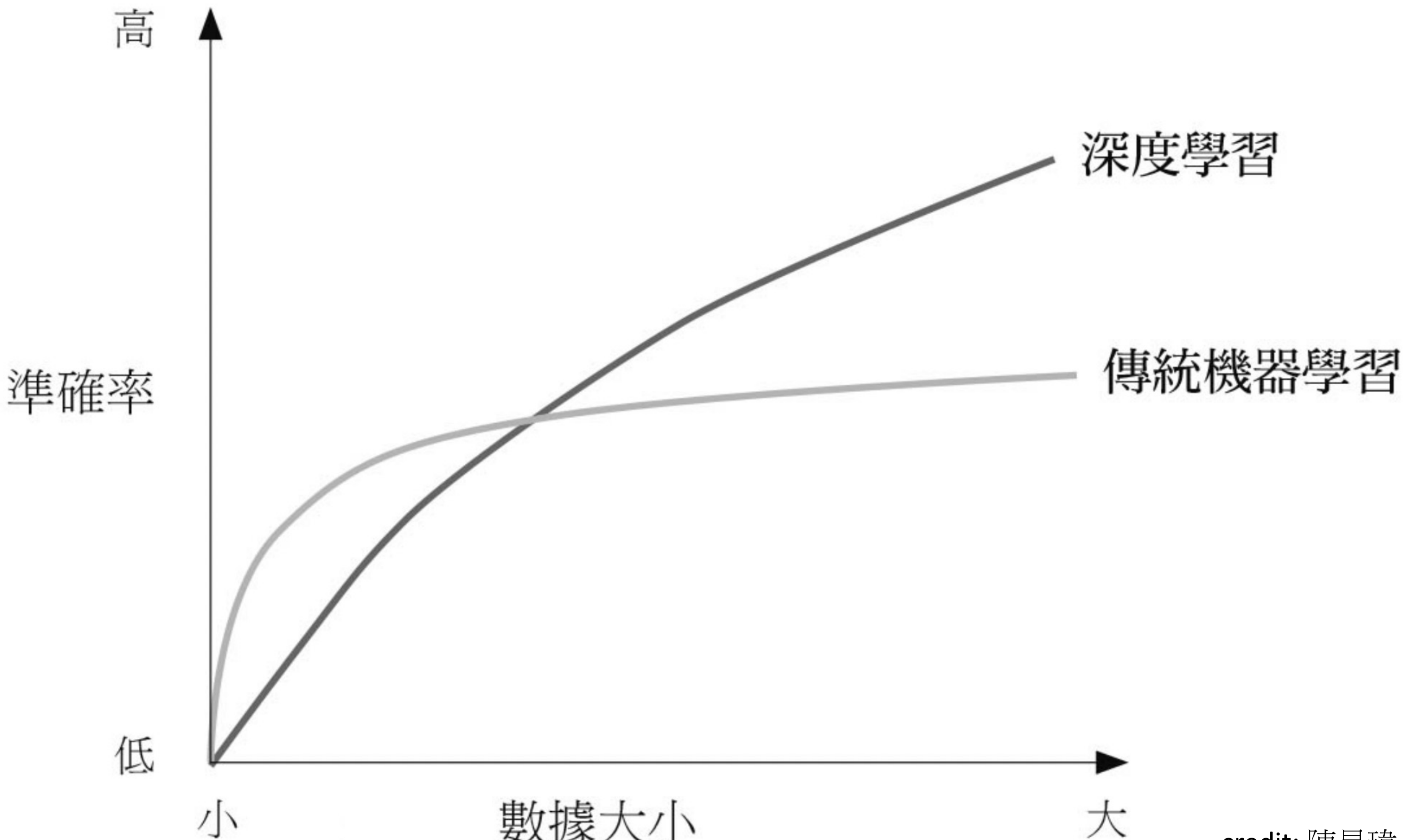
"Best" ML Model: find the optimal trade-off between **bias** and **variance**



- Given this training dataset, the optimal trade-off between bias and variance is a polynomial curve $p = 3$. (the best model)

(Fun Time) If **we increase the number of our training examples** from 40 to 200, the best ML model will likely be (1) $p = 3$ (2) $p > 3$ (3) $p < 3$

- A complex model (such as deep learning model) can easily overfit a small training dataset thus will lead to bad performance for new data.
- The deep learning model excels when the training data is large, and can often reach a much higher accuracy (sometimes unbelievably high accuracy, e.g., computer vision) comparing with the classical ML model.



credit: 陳昇瑋：人工智慧在台灣

Bias and variance: theoretical minimum and example

- The phrase "theoretical minimum" is taken from a very successful book series written by Leonard Susskind, a great physicist at Stanford University.
- "Theoretical minimum" means <u>just the minimum theories and equations you need to know in order to proceed to the next level.</u>
- See Bias_Variance.pdf

Summary

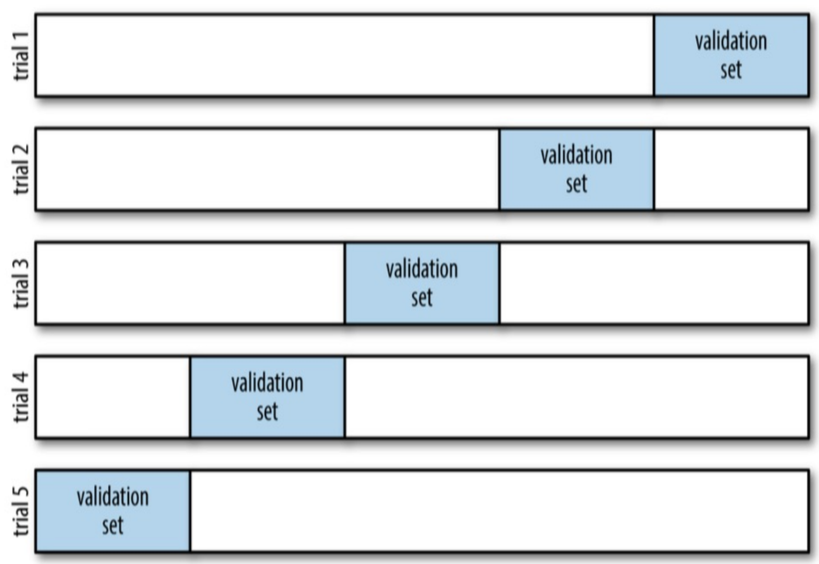
Learn the sources of errors when you apply your learning model for new data

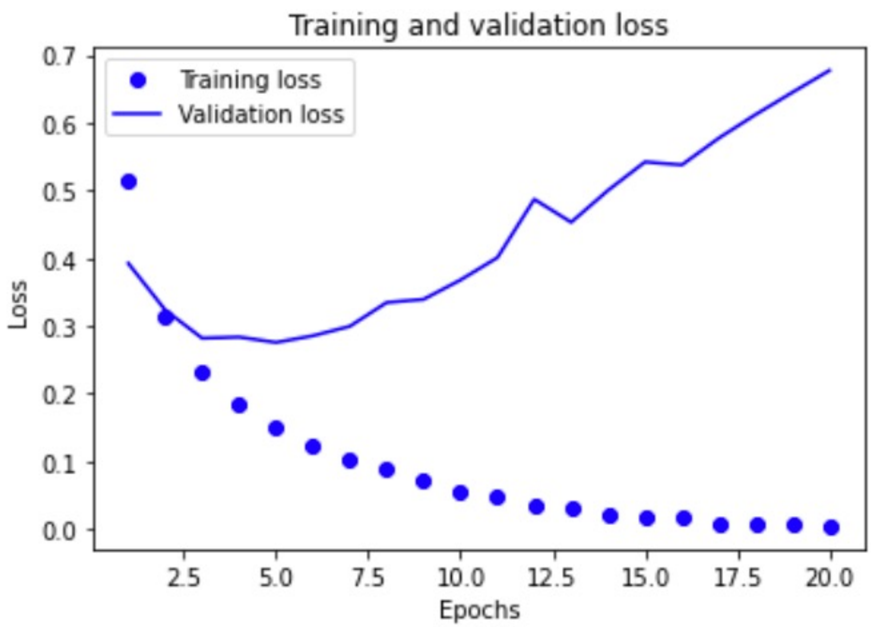
- **Bias** and **variance** are two major sources of errors that prevent supervised learning algorithms from generalizing beyond their training set.
- Fundamentally, the question of "the best ML model" is about **finding a sweet spot in the tradeoff between bias and variance**.
- The optimal model will generally depend **on the size** of training data.
- For general cases, it is not possible to explicitly compute bias and error; we rely on **the validation curve** and **the learning curve** to help us spot them.

Evaluate the performance of machine learning algorithms with cross validation

Machine Learning vs. Deep Learning: Performance Assessment

- K-fold cross validation is a gold standard in classical machine learning to evaluate performance but rarely used in deep learning (computational prohibited)
- Gold standard in deep learning is the validation curve.





Evaluate the performance of machine learning algorithms

Fun Time: Can we test our learning algorithm on the same training dataset to evaluate predictive performance of the algorithm for new data? (1) Yes (2) No (3) Maybe

Evaluate the performance of machine learning algorithms

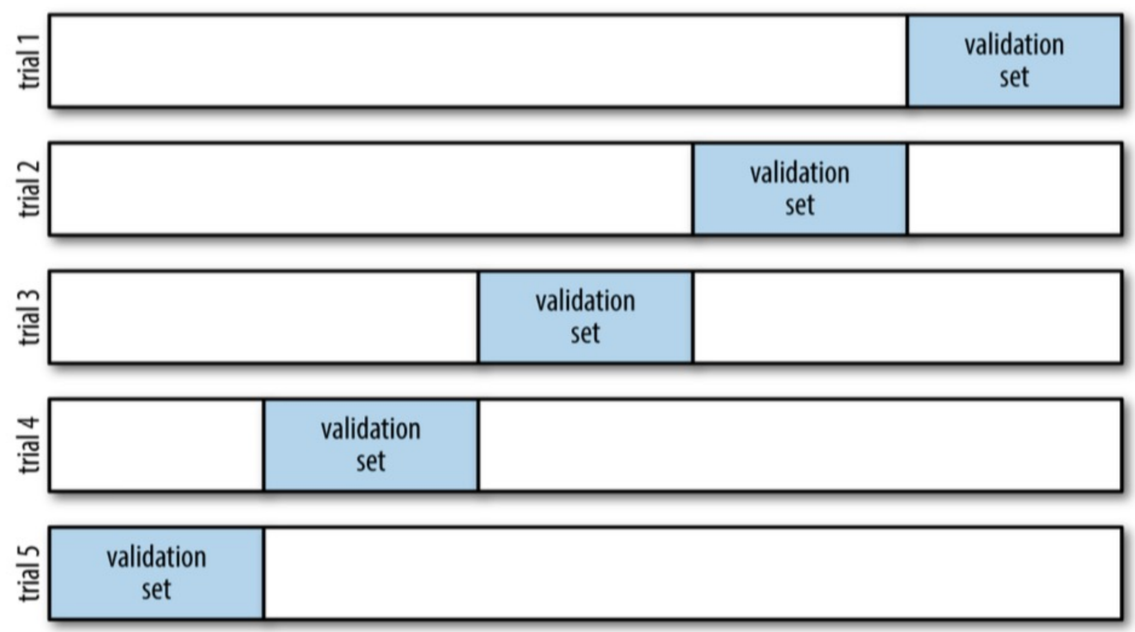
- **A simple approach: split your data into train and test data (say 50/50). Use train data to fit your model and use test data to evaluate performance of your algorithm.**

<u>Fun Time</u>: what is the potential problem of this approach?

Evaluate the performance of machine learning algorithms with cross validation

Cross_validation.ipynb

- A common approach is to use **k-fold** cross validation



Fundamentals and Landscape of Classical Machine Learning: Summary

- **Machine learning: use data to compute hypothesis *g* that approximate unknown target *f*.**
- **In practice, learning algorithm $\mathcal{A}$ takes training examples $\mathcal{D}$ and hypothesis set $\mathcal{H}$ to get final hypothesis *g*.**
- **Learning is only feasible in a *probabilistic* way and we can predict something useful outside the training set $\mathcal{D}$ using only $\mathcal{D}$.**
- **Scikit-Learn and Keras (now part of TensorFlow) are mostly widely used ML software frameworks by ML professionals.**
- **From 2016 to 2020, the entire machine learning industry has been dominated by deep learning and gradient boosted trees.**
- **Specifically, gradient boosted trees is used for problems where structured data is available, whereas deep learning is used for perceptual problems such as image classification.**

Fundamentals and Landscape of Classical Machine Learning: Summary

- **Bias and variance are two major sources of errors that prevent supervised learning algorithms from generalizing beyond their training set**
- **Fundamentally, the question of "the best ML model" is about finding a sweet spot in the tradeoff between bias and variance.**
- **For general cases, it is not possible to explicitly compute bias and error; we rely on the validation curve and the learning curve to help us spot them.**
- **K-fold cross validation (3, 5, 10) is a gold standard in classical machine learning to evaluate model performance but rarely used in deep learning (computational prohibited).**
- **Gold standard in deep learning is the validation curve.**