Fundamentals of 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/
#ML101

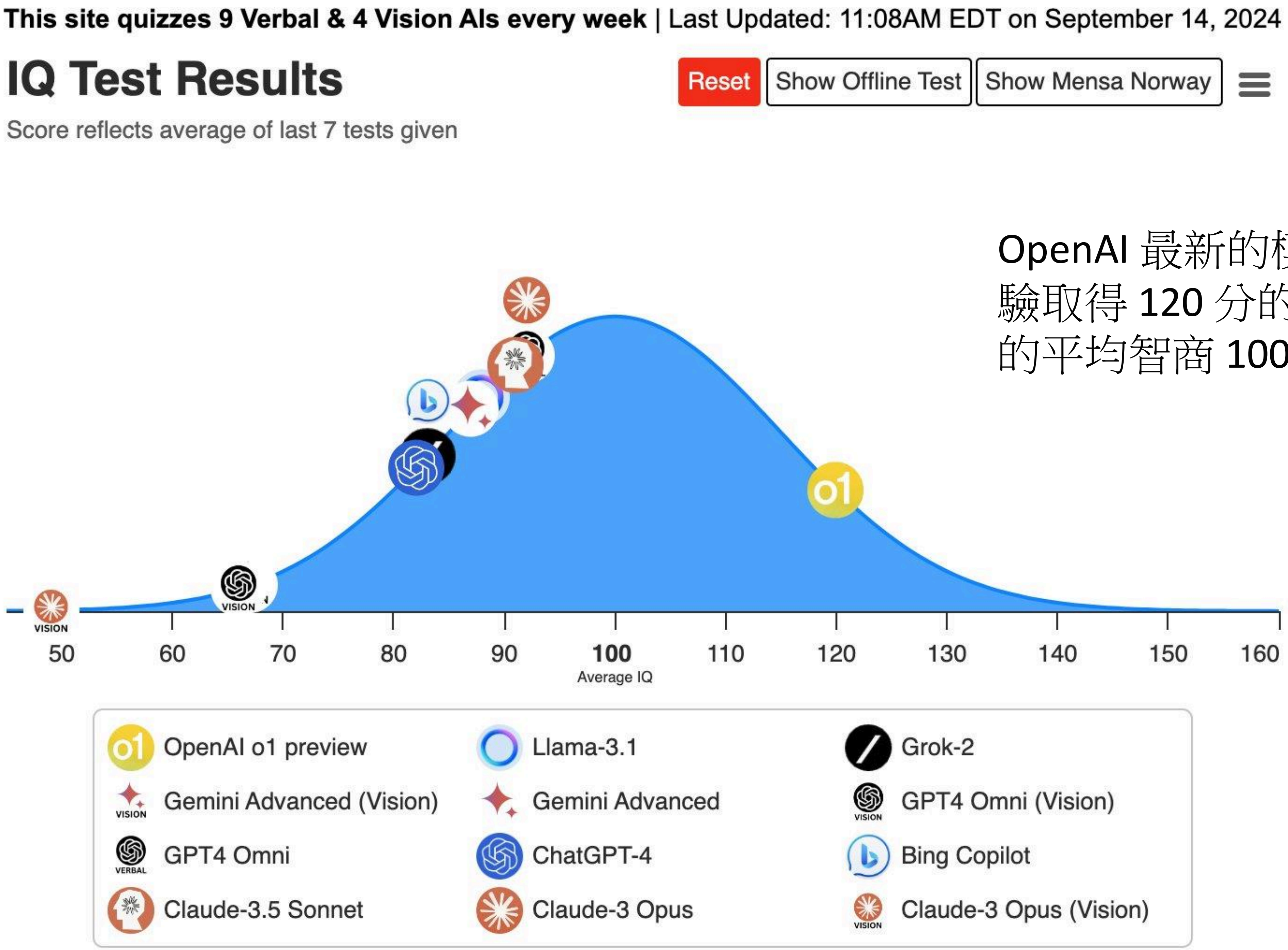
汝為Ai人，不可不知Ai事

OpenAI has officially released ‘o1’ (internally known as Project Strawberry/Q*), its first AI model with advanced 'reasoning' capabilities.

The details:

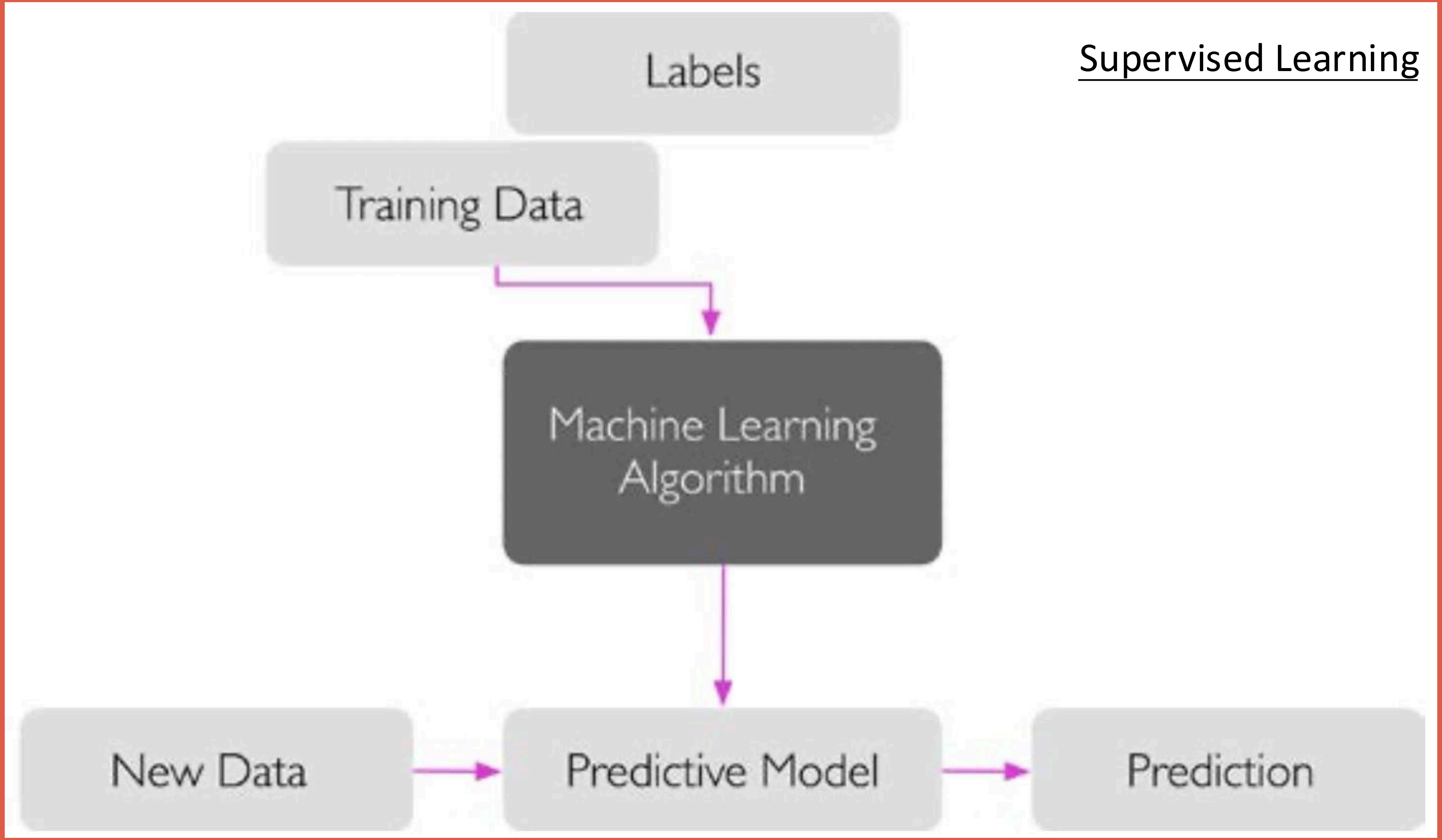
- o1 uses reinforcement learning and chain-of-thought processing to "think" before responding, mimicking human problem-solving.
- It outperforms expert humans on PhD-level science questions and ranks in the 89th percentile for competitive programming.
- The model also solved 83% of International Mathematics Olympiad qualifying exam problems, compared to GPT-4o's 13%.

Why it matters: OpenAI o1 is better than expert humans on PhD-level science questions. Its enhanced reasoning capabilities by "thinking" before responding will not only lead to more accurate AI responses, but opens up an entirely new world of real-world use cases for complex problems in science, coding, math, and more.

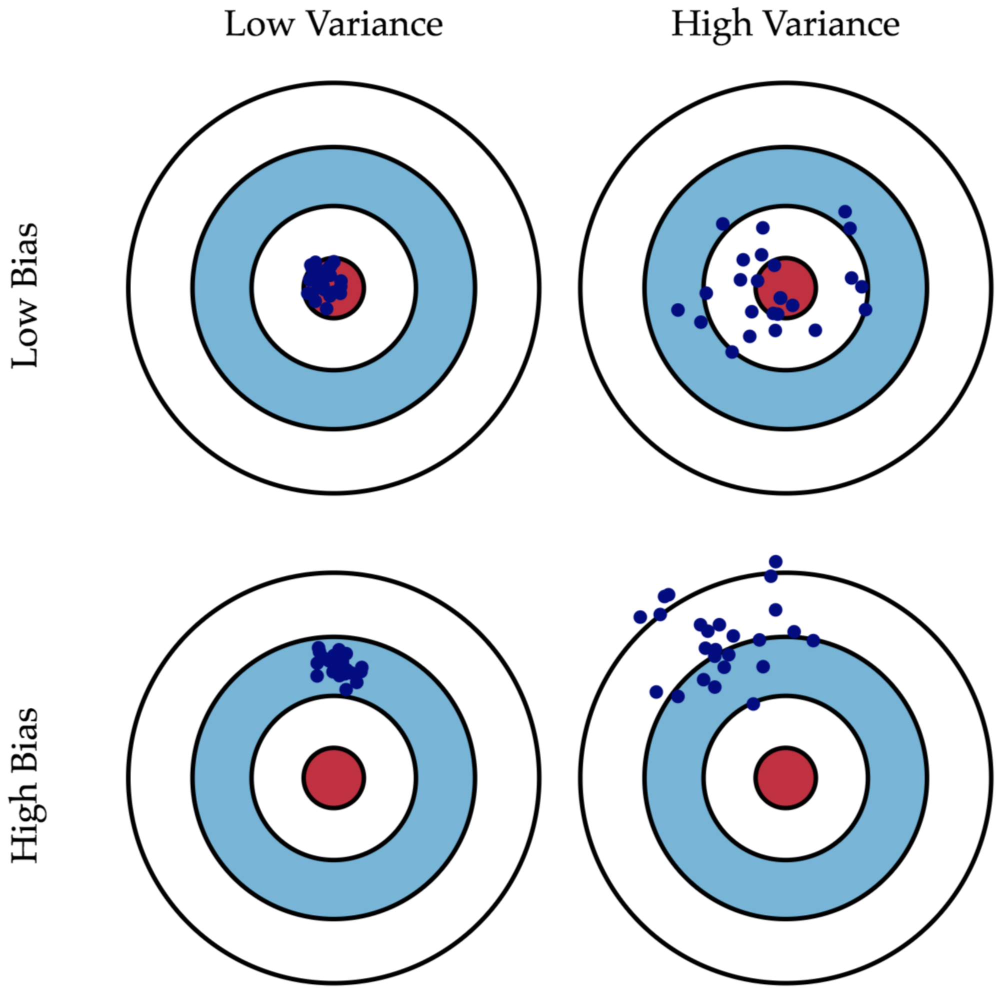


OpenAI 最新的模型 o1 在智力測驗取得 120 分的成績，超越人類的平均智商 100

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

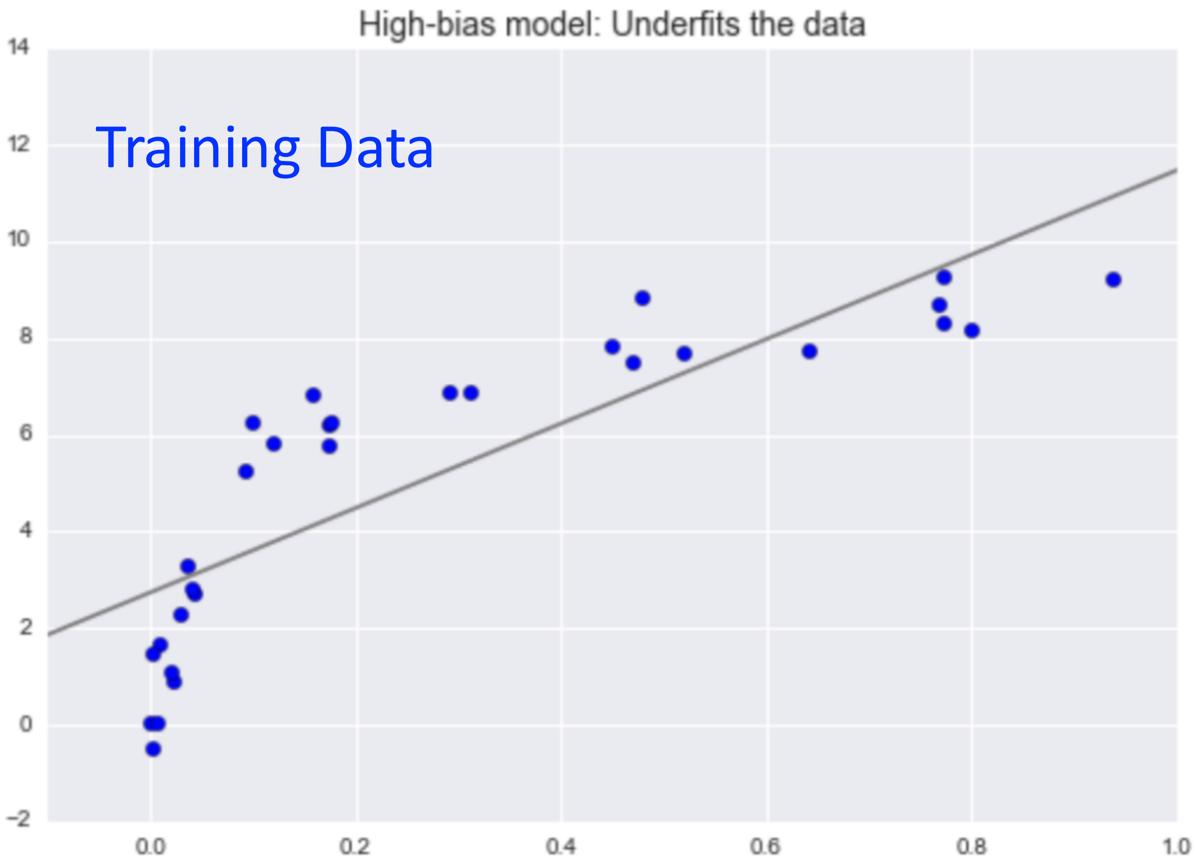


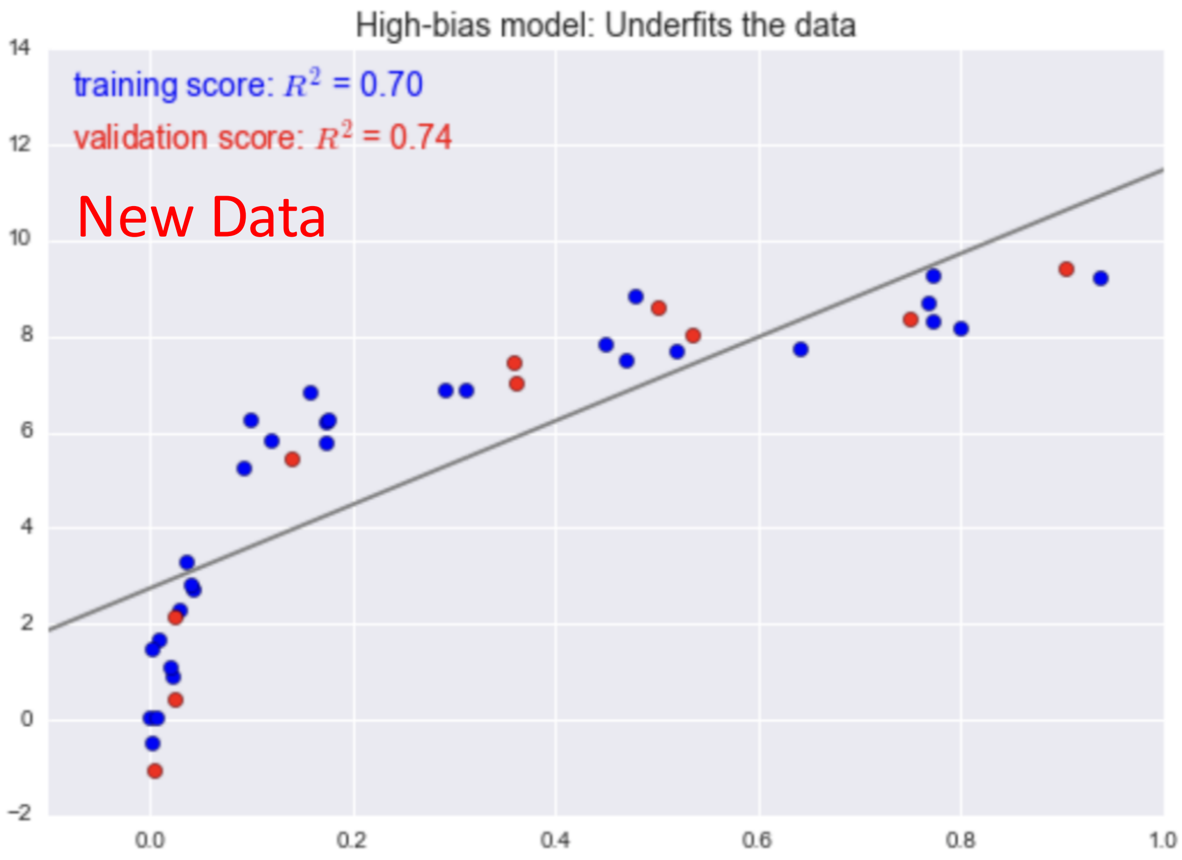
Bias and **variance** are two major sources of errors **that prevent supervised learning algorithms from generalizing beyond their training set**



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

$p = 1$

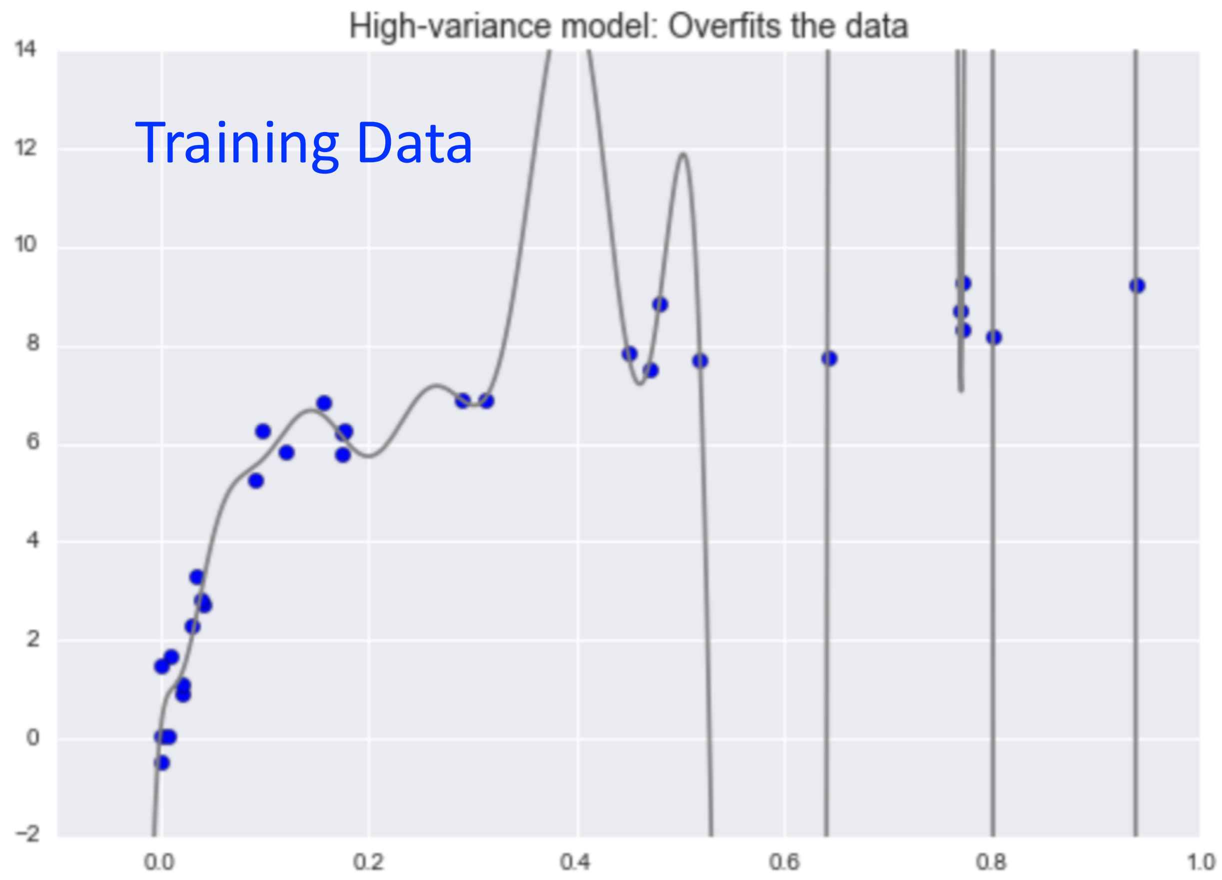


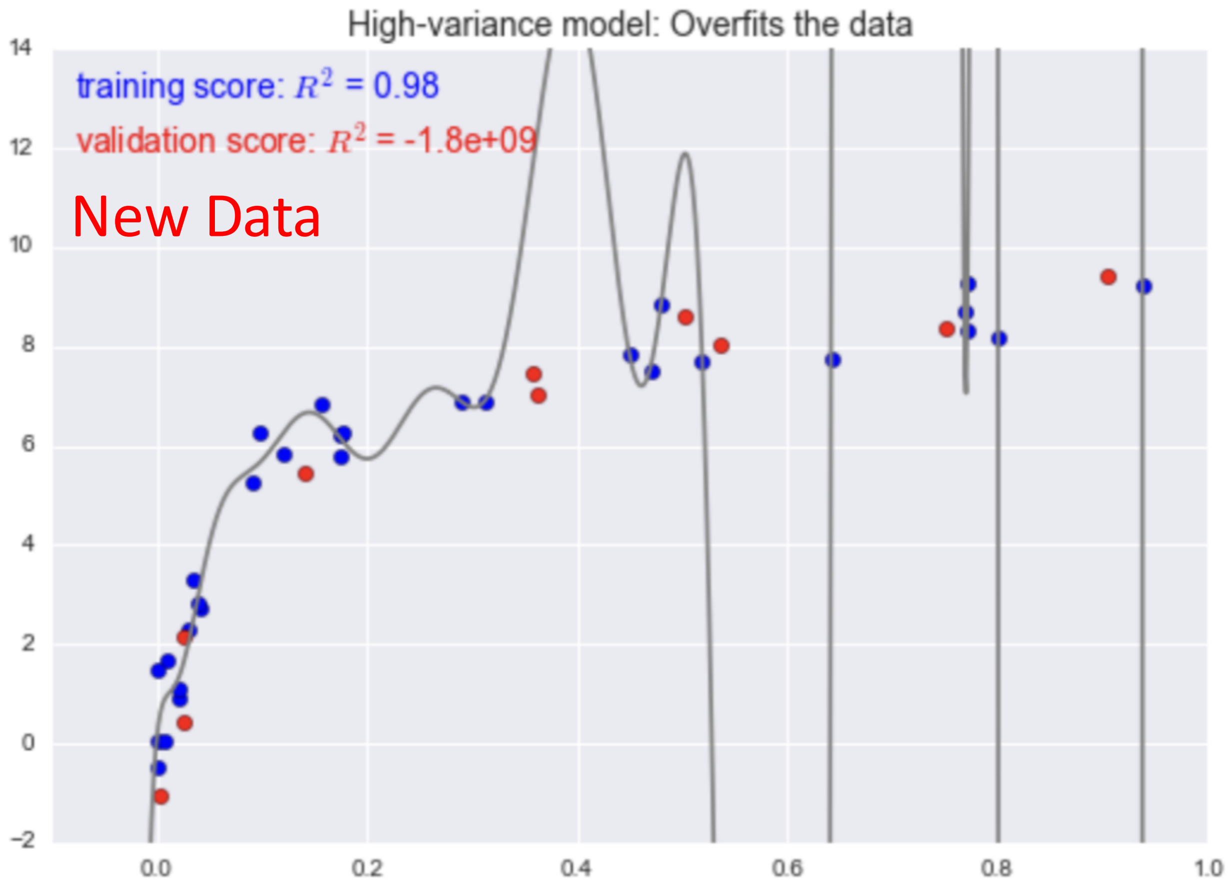


$R^2 = 1$ indicates a perfect match, $R^2 = 0$ indicates the model does no better than simply taking the mean of the data, and negative values mean even worse models.

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

$p = 20$



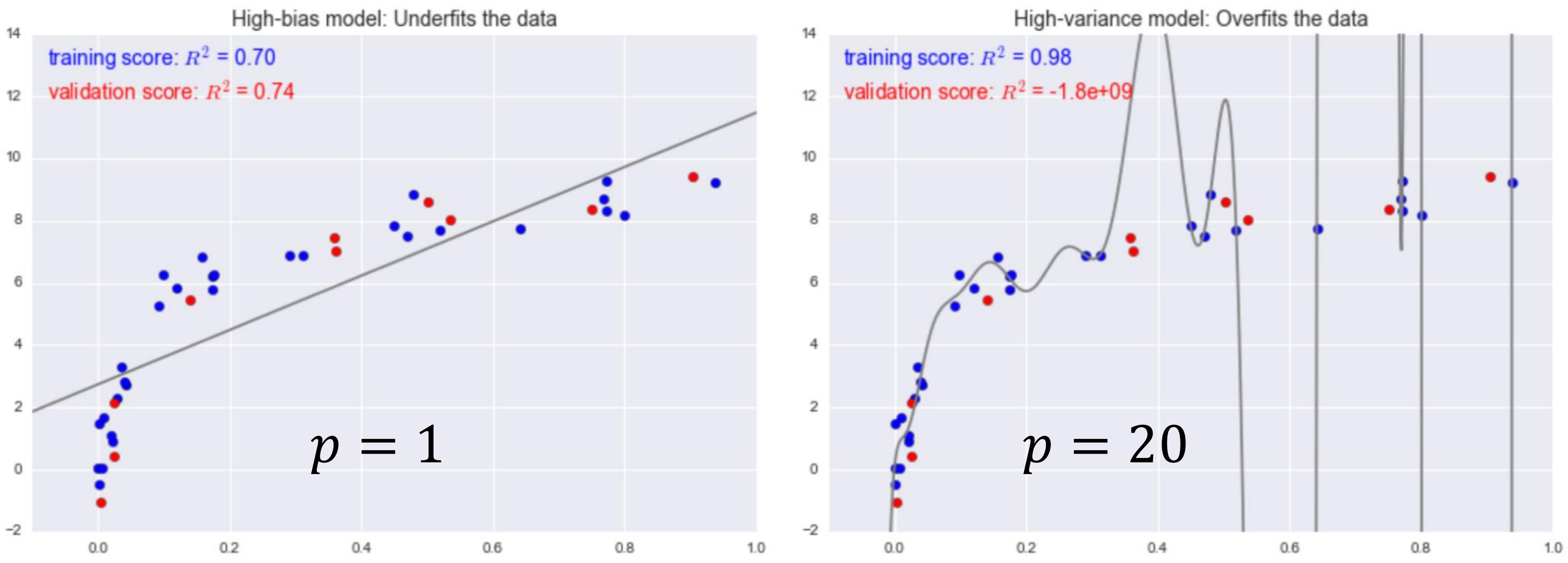


$R^2 = 1$ indicates a perfect match, $R^2 = 0$ indicates the model does no better than simply taking the mean of the data, and negative values mean even worse models.

Fun Time: 假設你正在開發一個機器學習模型來預測房價。你的模型在訓練資料上的預測誤差非常小，但在測試資料上的誤差卻非常大。這種現象最有可能是由哪種問題導致的？

(1) 高偏差 (High bias)
(2) 高變異 (High variance)
(3) 低偏差 (Low bias)
(4) 低變異 (Low variance)

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



- "Human" search: let us work through the example and find the best model (best polynomial curve)
- See Bias_var_intuitive_example.ipynb

Summary

Bias and variance tradeoff: an intuitive example

- **High bias** model (low order polynomial) tends underfit the data.
- **High variance** model (high order polynomial) tends overfit the data.
- ”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.

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

C-S David Chen, Department of Civil Engineering, National Taiwan University

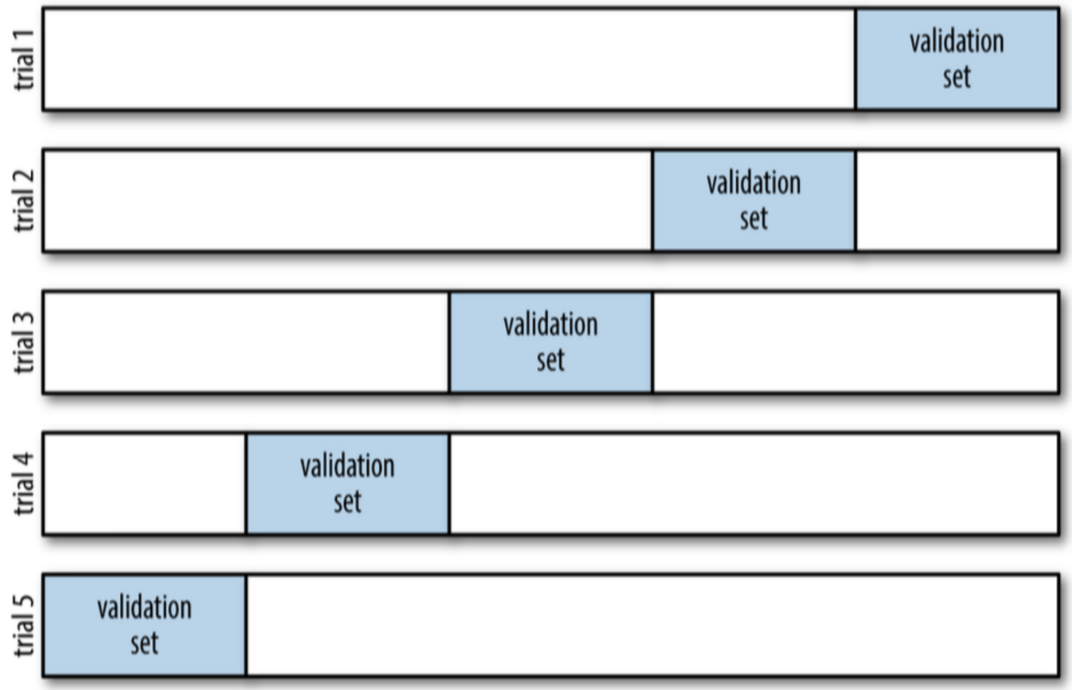
Fun Time: 在開發模型時你只會有一組千辛萬苦得到的資料，通常我們會把一部分資料用來訓練模型，一部分資料用來評估模型面對新資料的表現。

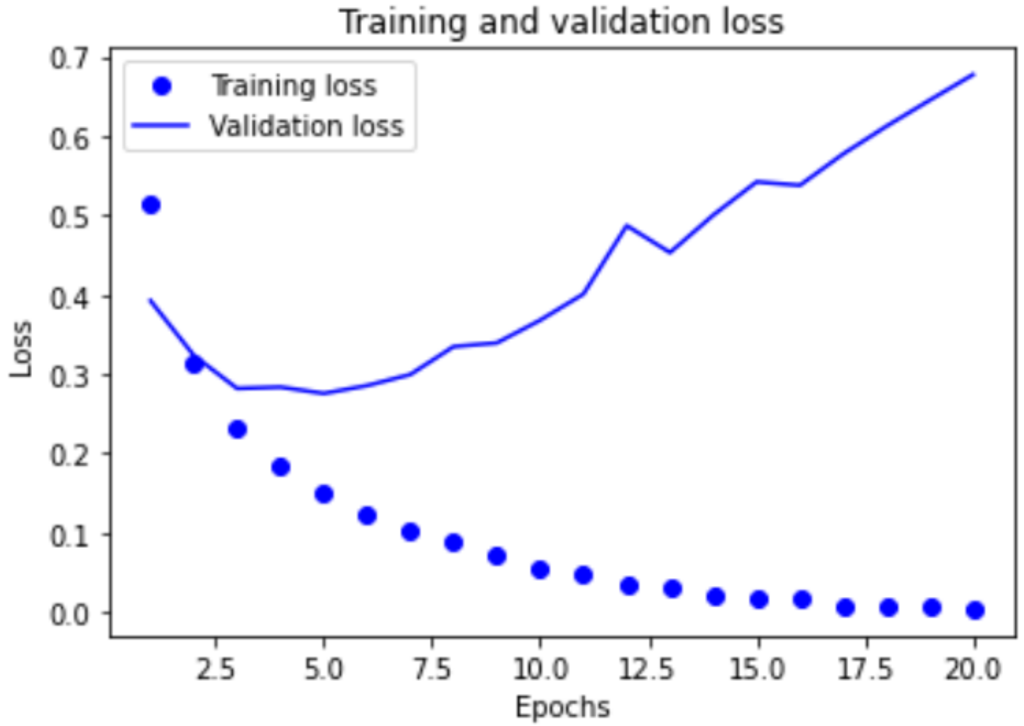
If we split data into train and test data (say 50/50). Use train data to fit the model and use test data to evaluate performance of your algorithm. What is the potential problem of this approach?

1. **waste of data for training**
2. **waste of data for testing**
3. **all of the above**
4. **none of the above.**

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.

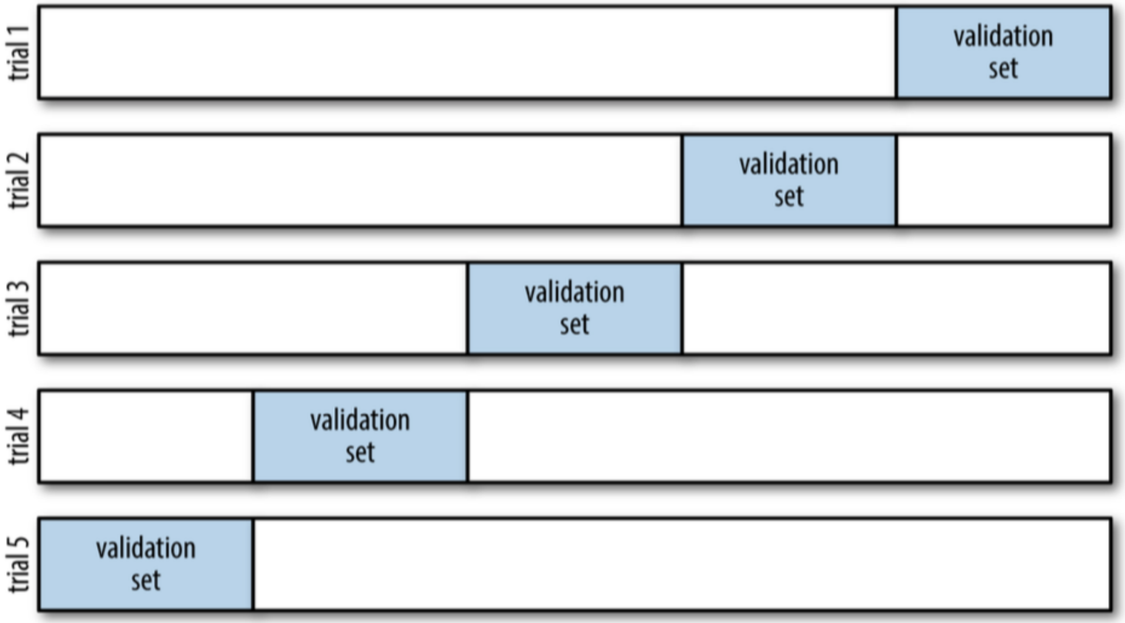




Evaluate the performance of machine learning algorithms with cross validation

Cross_validation.ipynb

- **K-fold cross validation is a gold standard in classical machine learning to evaluate performance**



Fundamentals of 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 of 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.**