深入淺出機器學習

Head-in machine learning

Agenda

- What is machine learning
- 機器學習三兄弟
- Types of learning algorithms

Machine learning =
let machine learn

Machine learning

- Find an appropriate decision function

$$f: \text{feature space} \rightarrow \text{decision space}$$

to predict future outcomes

- Mathematically, we solve the following problem

$$\min_{f(\mathbf{x})} E\left[L\left(y, f(\mathbf{x})\right)\right] \qquad (1)$$

Machine learning三兄弟

- Loss function $L\left(y, f(\mathbf{x})\right)$ measures the accuracy of a prediction

- Representation of the decision function $f(\mathbf{x})$

- Model complexity (regularization)

Loss function

- Numerical y: $L\left(y, f(\mathbf{x})\right) = \left(y - f(\mathbf{x})\right)^2$
- Categorical y: $L\left(y, f(\mathbf{x})\right) = I\left\{y \neq f(\mathbf{x})\right\}$
- Vector $\mathbf{y}$: $L\left(\mathbf{y}, f(\mathbf{x})\right) = \|\mathbf{y} - f(\mathbf{x})\|^2$
- Structural learning

Empirical risk

- When the data $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$ are randomly sampled, equation (1) can be approximated by its empirical version

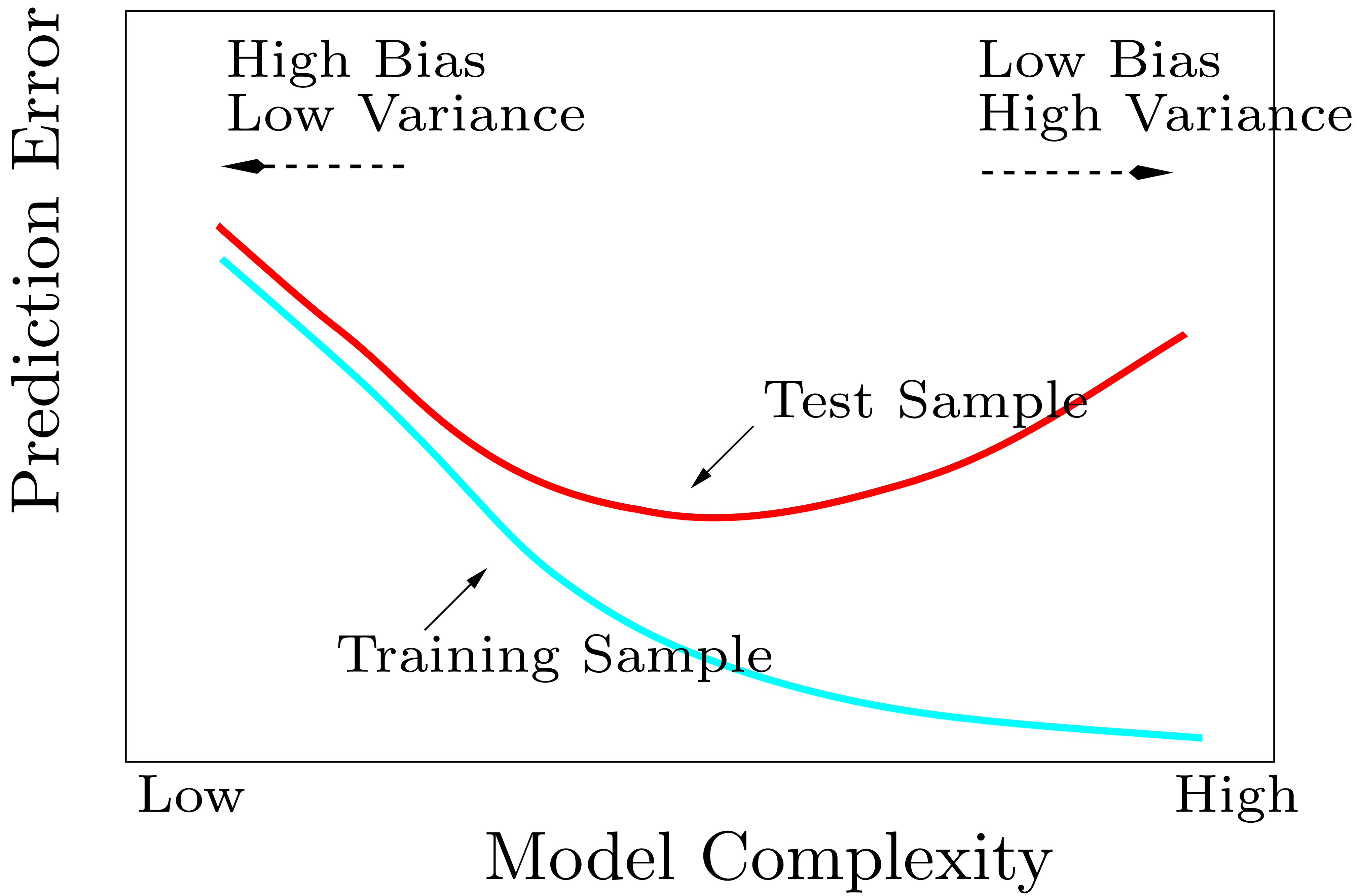
$$\min_{f(\mathbf{x})} \frac{1}{n} \sum_{i=1}^{n} \left[L\left(y_i, f(\mathbf{x}_i)\right) \right] \qquad (2)$$

- If $L\left(y, f(\mathbf{x})\right) = -\ell\left(y \,|\, f(\mathbf{x})\right)$, equation (2) is equivalent to MLE

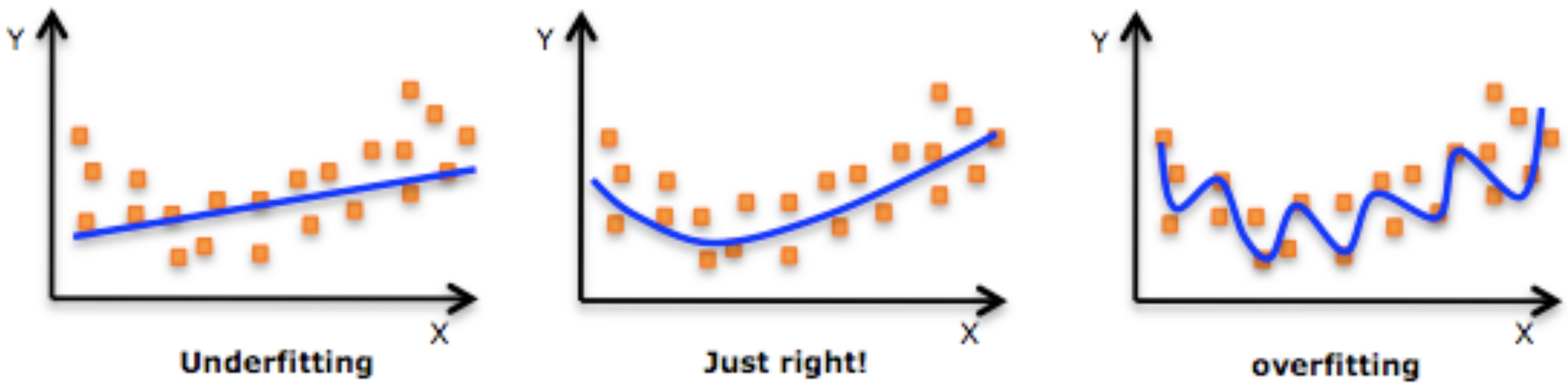
Representations

- Deep learning
- Kernel trick
- Gaussian processes
- Splines
- Wavelets
- Finite element methods

Overfitting



Overfitting



Structure risk (regularization)

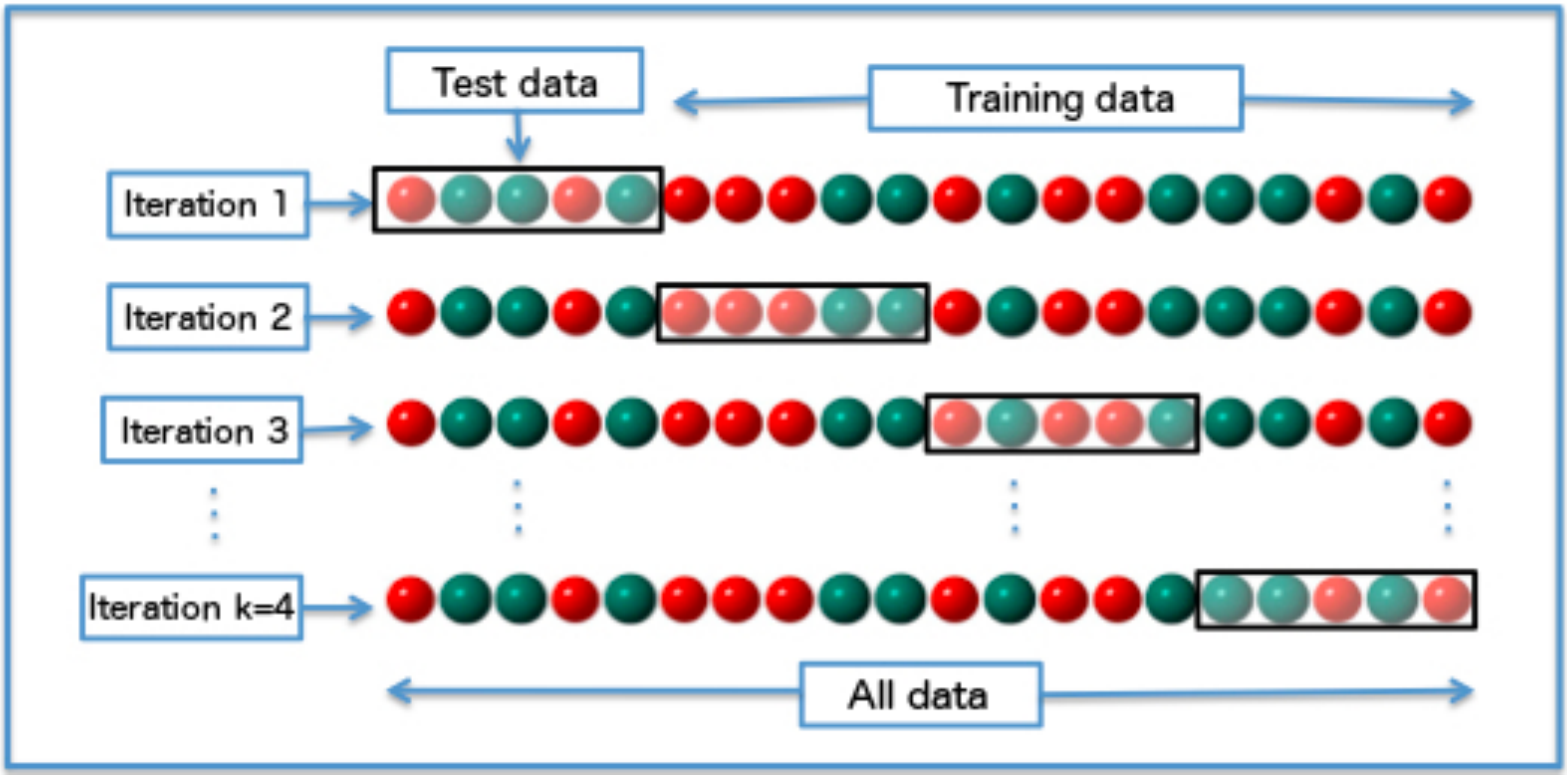
- Control the model complexity by introducing an additional regularization term:

$$\min_{f(\mathbf{x})} \frac{1}{n} \sum_{i=1}^{n} \left[L\left(y_i, f(\mathbf{x}_i)\right) \right] + \lambda ||f|| \quad (3)$$

- The tuning parameter λ needs to be chosen carefully (e.g. by cross-validation)

- $||f||$ can also be interpreted as the prior of f

Cross validation



Types of learning algorithms

- Supervised learning: when y_i's are known
- Semisupervised learning: when part of y_i's are known
- Unsupervised learning: when y_i's are unknown
- Autoencoder and GAN: $y_i = \mathbf{x}_i$
- Reinforcement learning

Homework: cross-validation

Find an appropriate tuning parameter C in homework 3 by:

- 5-fold CV
- 10-fold CV
- Shuffle split

Reference: sklearn.model_selection