CE6146
Introduction to Deep Learning
Deep Learning Basics

Chia-Ru Chung
Department of Computer Science and Information Engineering
National Central University

2023/9/21

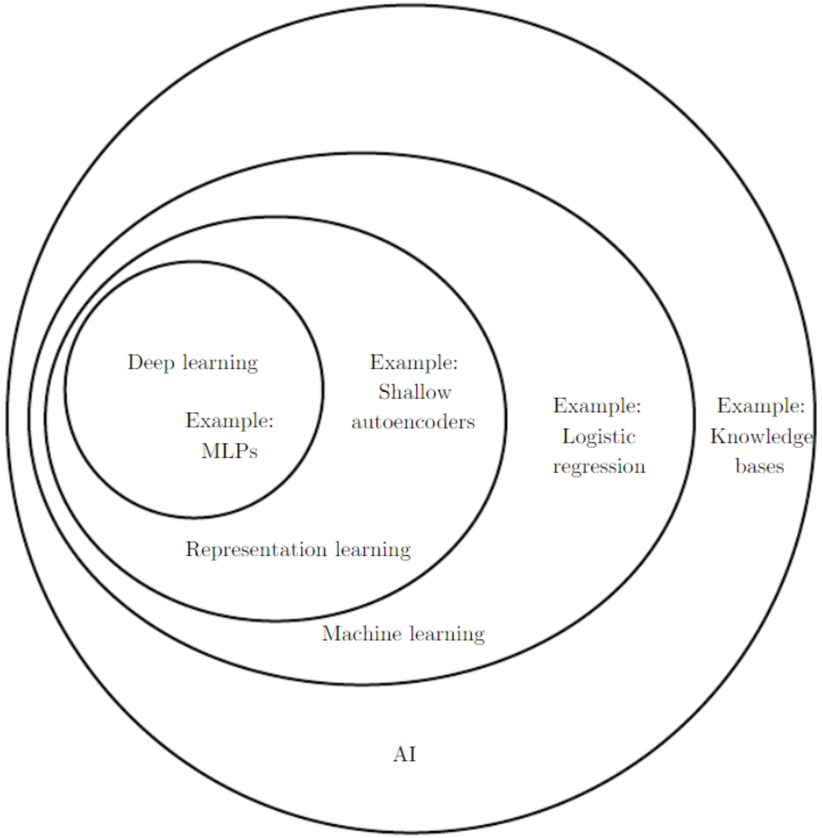
Outline

- Definition and Importance
- Essential Mathematics
- Programming and Data Structures
- Basic Concepts
- Deep Learning Libraries
- Summary

Definition and Importance

- What is deep learning
- Importance and applications

Deep Learning and AI



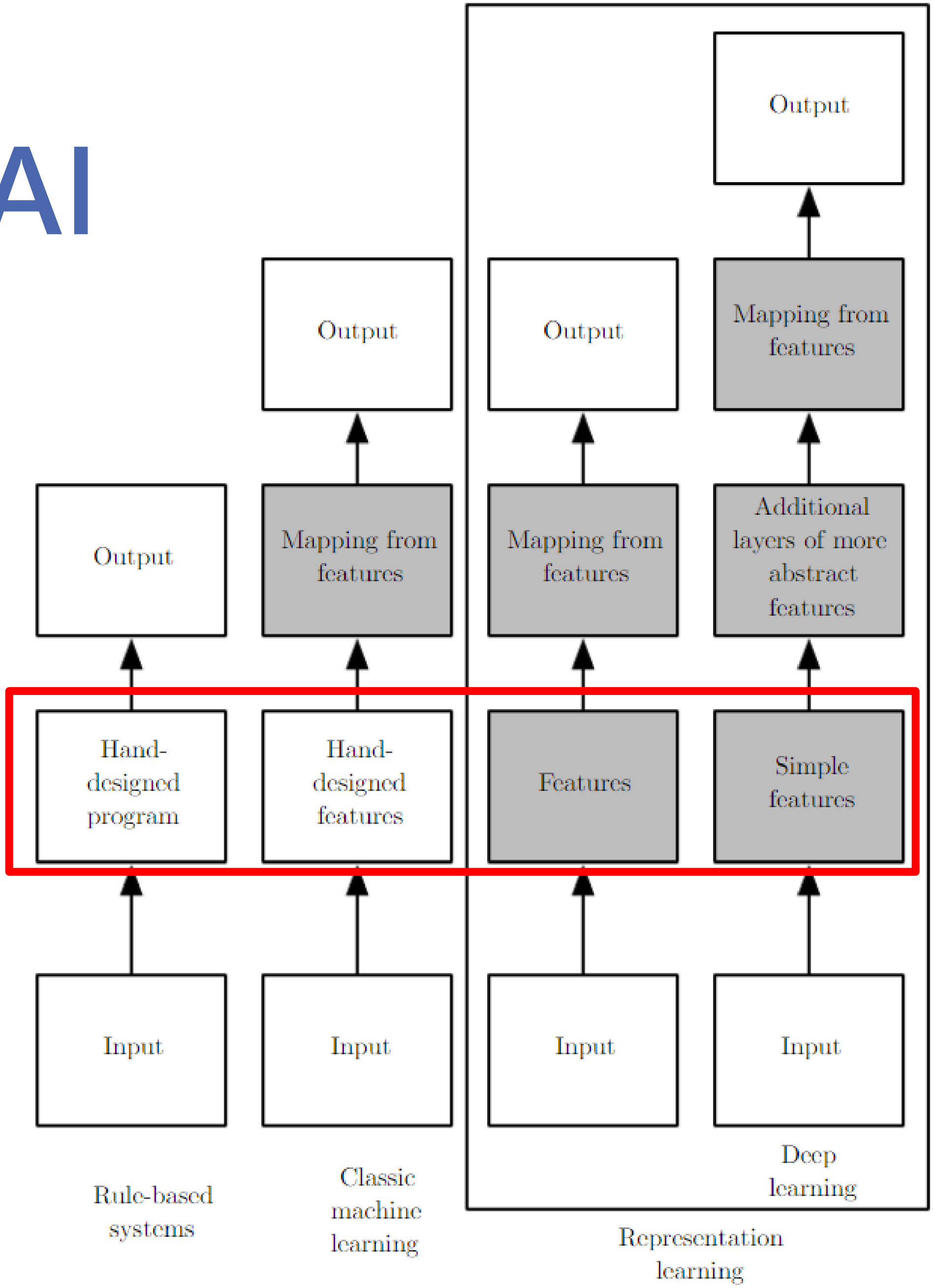


Figure 1.4 and 1.5 in Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

What is Deep Learning?

- Deep learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.
  - Focuses on learning representations from data.
  - Utilizes neural networks with many layers.

Different from Machine Learning

- Deep learning can automatically learn representations.
- No need for manual feature extraction.

Why Deep Learning?

- Deep Learning algorithms can automatically learn patterns.
- Effective in handling large and complex data.
- Revolutionized various domains including NLP, computer vision, and more.

Importance

- Revolutionized various domains including healthcare, autonomous vehicles, and more.
- Key enabler for tasks requiring human-like intelligence such as image and voice recognition.
- Widely adopted in industry and academia for complex problem-solving.

Essential Mathematics

- Linear Algebra
- Probability and Information Theory
- Numerical Computation

Linear Algebra

- Linear algebra is the branch of mathematics concerning linear equations, linear functions, and their representations through matrices and vector spaces.
- It forms the basis for understanding the structure of solutions to linear systems of equations, which appear across a variety of engineering disciplines.

Importance of Linear Algebra in Deep Learning (1/2)

- Data Representation: Data in machine learning is often represented in matrices.
- Efficiency: Linear algebra provides an efficient way of performing large-scale computations.
- Optimization: Techniques in linear algebra help in solving optimization problems in learning algorithms.

Importance of Linear Algebra in Deep Learning (2/2)

- Understanding Learning Algorithms: The inner workings of many machine learning algorithms are based on linear algebra.
- Feature Transformation: Principal Component Analysis (PCA), which is based on linear algebra, is used for feature transformation or dimensionality reduction.

Linear Algebra – Terminology

- Scalar: A single numerical value. Used for scaling vectors and matrices.
- Vector: An ordered array of numbers. Each element is identified by an index.
- Matrix: A 2D array of numbers, each identified by two indices.
- Tensor: An N-dimensional array, generalizing scalars, vectors, and matrices.

Linear Algebra – Norms

- Norms measure the length or size of vectors. Common norms include:
  - L1 Norm: Sum of absolute values.
  - L2 Norm: Square root of sum of squares.
- They are used in regularization, optimization, and defining distances.

Probability and Information Theory

- Probability theory is a mathematical framework for quantifying uncertainty, providing a set of rules for reasoning about uncertain events.
- Information theory provides measures to quantify information such as entropy, cross-entropy, and Kullback-Leibler divergence.

Importance of Probability and Information Theory in Deep Learning (1/2)

- Uncertainty Handling: Deep learning models often have to make decisions in uncertain situations. Probability theory helps in quantifying and handling this uncertainty.
- Probabilistic Models: Many deep learning models, especially classifiers, are probabilistic in nature. They predict the likelihood of different possible outcomes.

Importance of Probability and Information Theory in Deep Learning (2/2)

- Optimization: Concepts like Maximum Likelihood Estimation (MLE) are rooted in probability theory and are commonly used in training deep learning models.
- Regularization: Techniques like dropout can be better understood with a grounding in probability theory.
- Information Metrics: Measures like cross-entropy and KL-divergence are used in defining loss functions and are based on information theory.

Why Probability in Deep Learning

- Understanding probability is crucial for deep learning and machine learning for several reasons:
  - Data is inherently uncertain.
  - Models make probabilistic predictions.
  - Helps in understanding and modeling the 'randomness' in real-world data.

Probability Theory – Terminology

- Random variable: A random variable is a variable that can take different values randomly. Random variables can be discrete or continuous.
- Probability distribution: A probability distribution describes how the probabilities of a random variable are distributed. Common distributions include Gaussian, Bernoulli, and Poisson.

Importance of Numerical Computation

- Handling large-scale data: Deep learning often involves working with massive datasets, and efficient numerical computation is essential.
- Optimization: Understanding the numerical aspects is crucial for the optimization of deep learning algorithms.
- Stability and Accuracy: Ensuring the numerical stability and accuracy of algorithms is vital for their successful application.

Overflow and Underflow

- Overflow occurs when numbers are too large to be stored in the available memory.
- Underflow occurs when numbers are too close to zero.
- Techniques like log-sum-exp trick can help in mitigating these issues.

Conditioning and Stability

- Conditioning refers to how rapidly a function changes with respect to small changes in its inputs.
- Poorly conditioned problems are numerically unstable and can cause significant issues in deep learning models.

Gradient-Based Optimization (1/2)

- Basic Gradient Descent: The simplest form where gradients are calculated for the entire dataset.
- Stochastic Gradient Descent (SGD): A variant where a subset of data is used to estimate the gradient, making it computationally more efficient.
- Momentum: A technique to help accelerate gradients vectors in the right directions, thus leading to faster converging.
- Adaptive Learning Rates: Methods like Adagrad, Adadelta, RMSprop, and Adam adjust the learning rates during training.

Gradient-Based Optimization (2/2)

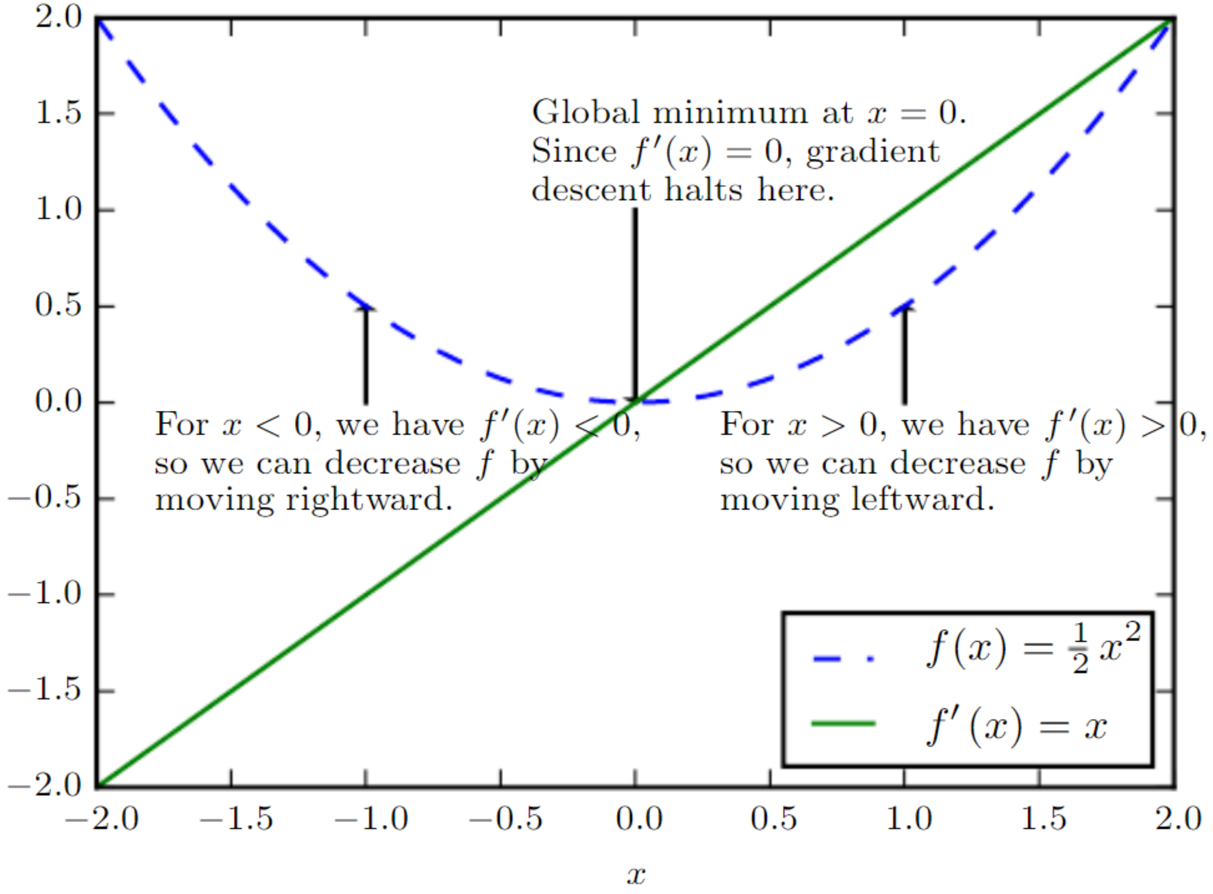
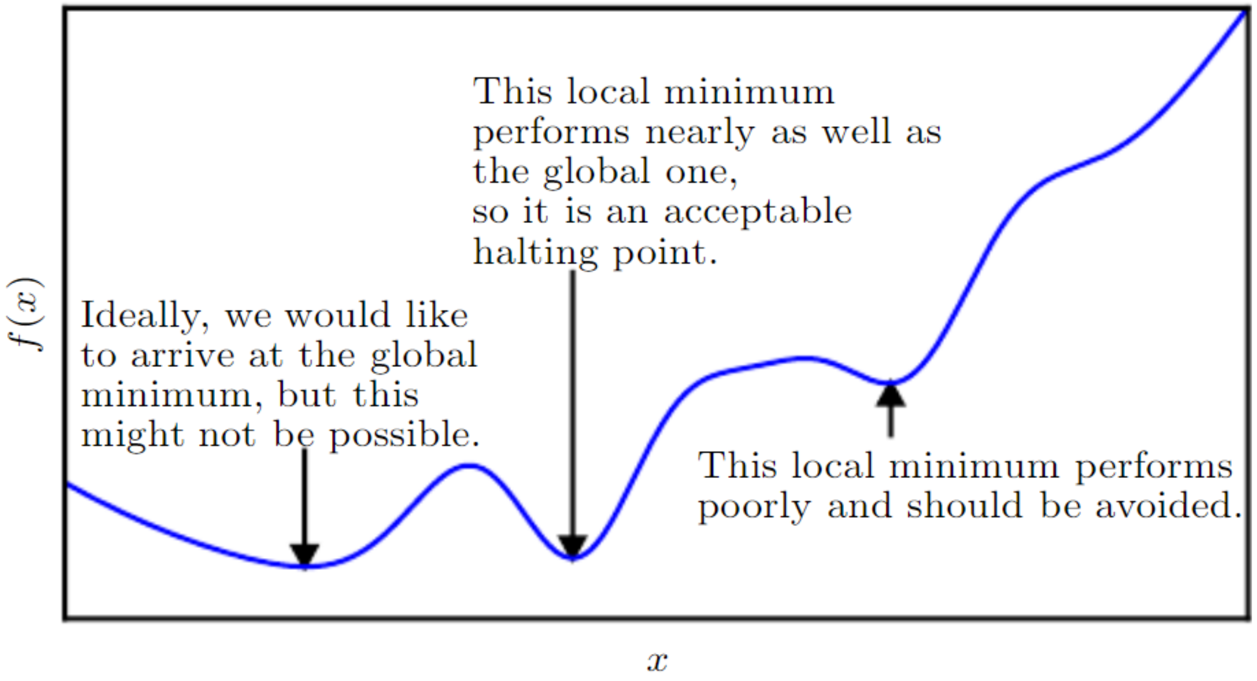


Figure 4.1 in Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

- An illustration of how the gradient descent algorithm uses the derivatives of a function to follow the function downhill to a minimum.

Approximate Optimization



- Optimization algorithms may fail to find a global minimum when there are multiple local minima or plateaus present. In the context of deep learning, we generally accept such solutions even though they are not truly minimal, so long as they correspond to significantly low values of the cost function.

Figure 4.3 in Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

Conditioning and Stability

- Numerical computation forms the backbone of deep learning algorithms. It's crucial for:
  - Efficiently performing matrix and tensor operations.
  - Solving optimization problems.
  - Handling large-scale data.

Conditioning and Stability

- Numerical computation forms the backbone of deep learning algorithms. It's crucial for:
  - Efficiently performing matrix and tensor operations.
  - Solving optimization problems.
  - Handling large-scale data.

Basic Concepts

- Overview of Machine Learning
- Learning Algorithms

Overview of Machine Learning

- A subset of AI that provides systems the ability to learn from data and improve.
- Types of ML: Supervised learning, Unsupervised learning, Reinforcement learning.
- Algorithms: Decision Trees, SVM, k-NN, Naive Bayes, etc.

Types of Machine Learning

- Supervised Learning: Learning a function that maps from input to output based on example input-output pairs.
- Unsupervised Learning: Learning patterns in the input when no specific output is required.
- Reinforcement Learning: Learning to act based on rewards.

Challenges in Machine Learning

- Insufficient Data
- Poor Data Quality
- Nonrepresentative Data
- Irrelevant Features

Learning Algorithms

- A machine learning algorithm is an algorithm that is able to learn from data.
- A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E (Mitchell, 1997).

Learning Algorithms – Tasks

- Classification
- Classification with missing inputs
- Regression
- Transcription
- Machine translation
- Structured output
- Anomaly detection
- Synthesis and sampling
- Imputation of missing values
- Denoising
- Density estimation or probability mass function estimation

Generalization

- Refers to how well the model performs on unseen data.
- An ideal model should generalize well from the training data to any data from the problem domain.
- This allows us to make predictions in situations we haven’ t encountered before.

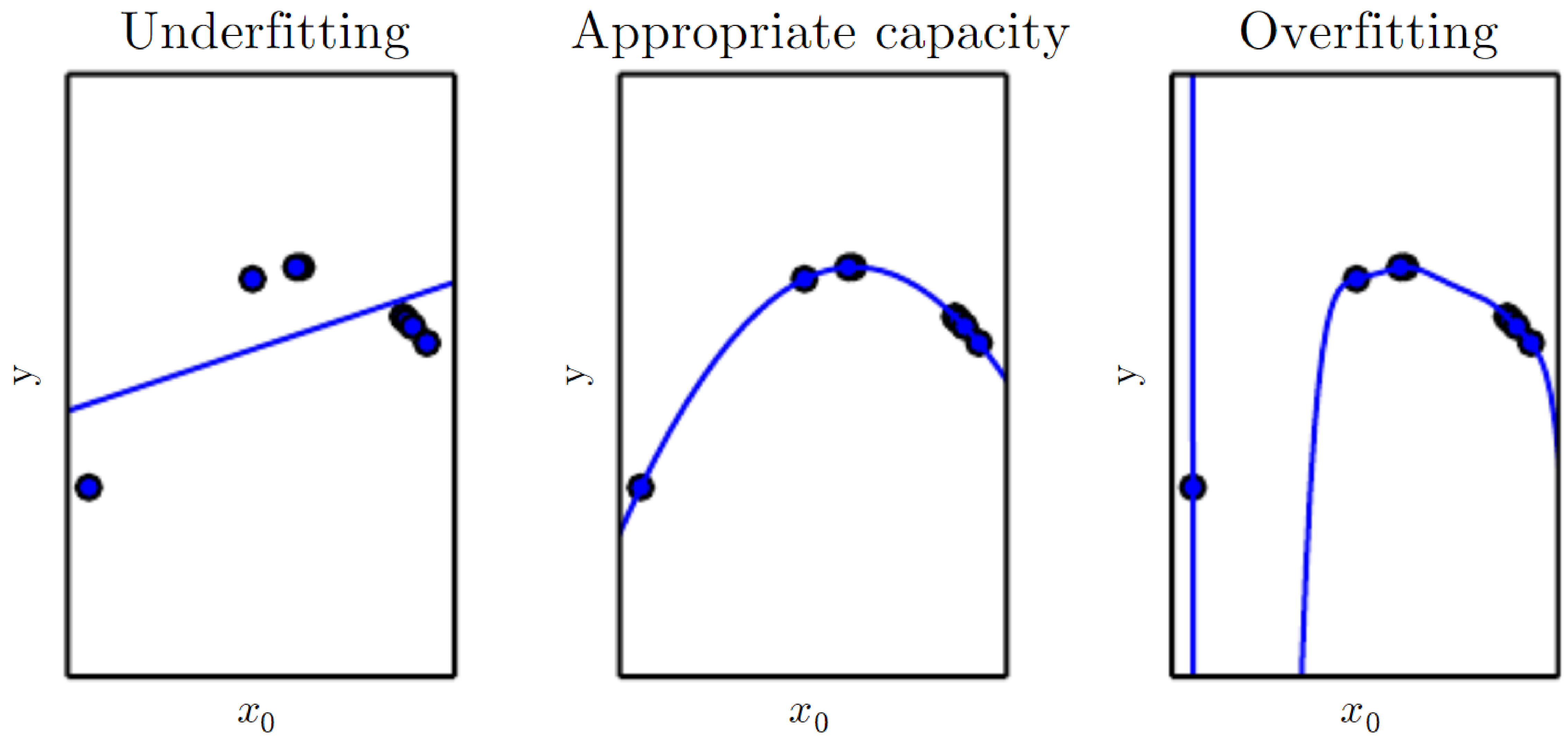


Figure 5.2 in Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

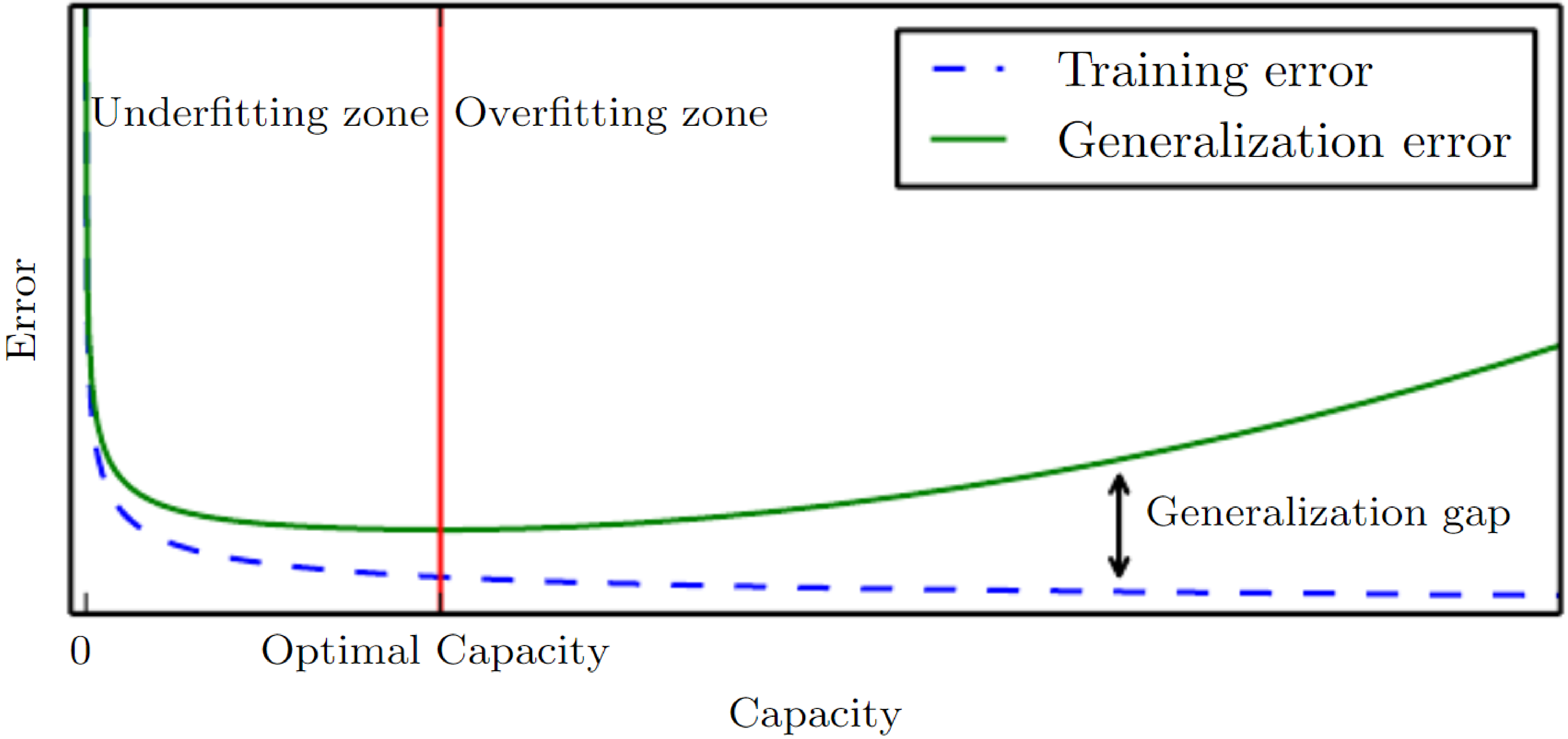
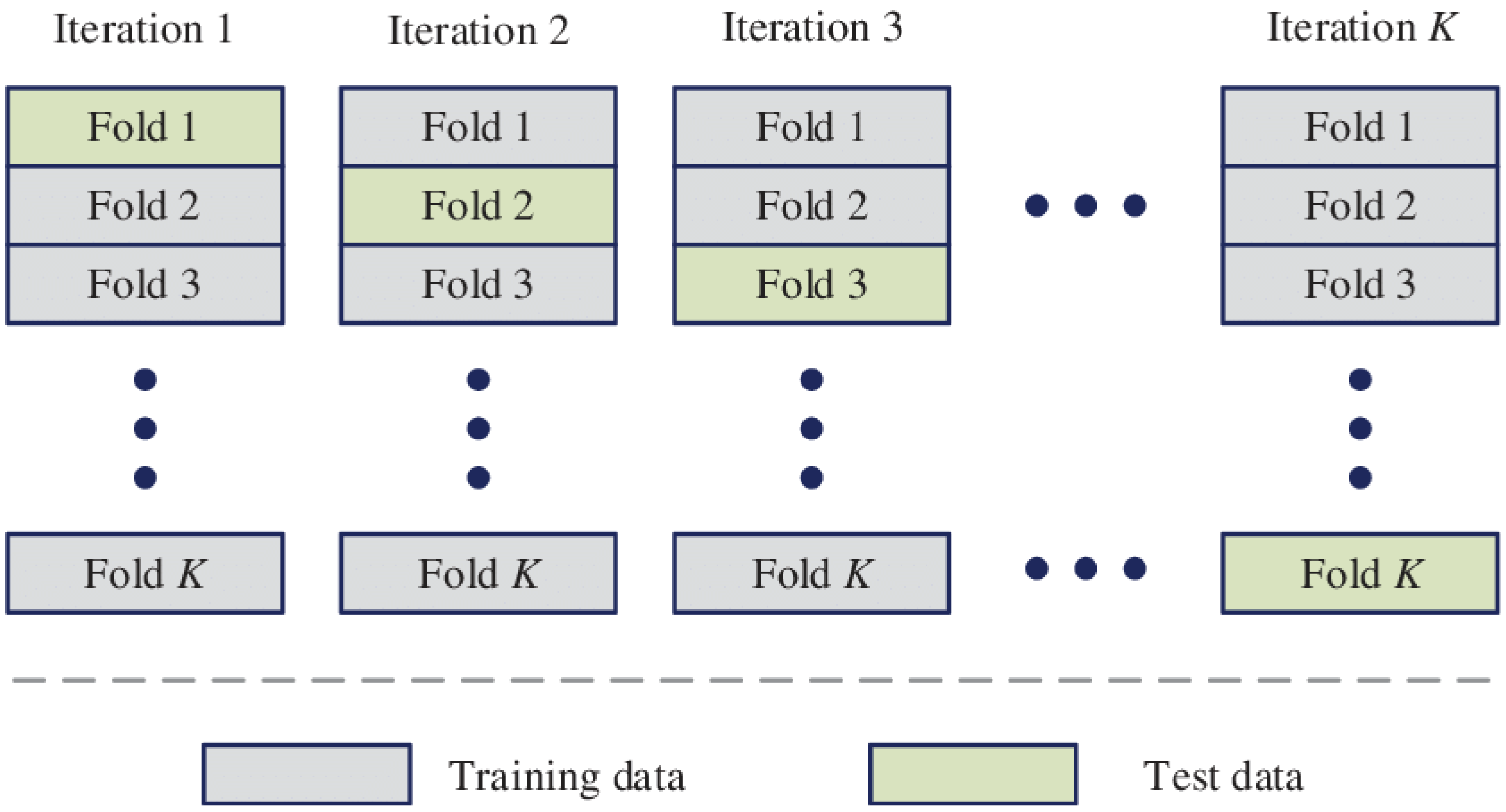


Figure 5.3 in Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

Cross-Validation (1/2)

- Dividing the dataset into a fixed training set and a fixed test set can be problematic if it results in the test set being small.
- A small test set implies statistical uncertainty around the estimated average test error, making it difficult to claim that algorithm A works better than algorithm B on the given task

Cross-Validation (2/2)



Source: https://doi.org/10.1080/20964471.2019.1572452

Hyperparameters and Validation Sets

- Hyperparameters are parameters not learned from the data but set prior to the training process.
- Examples include learning rate, batch size, and number of layers in a neural network.
- Validation sets are used for tuning these hyperparameters.

Estimators, Bias and Variance

- An estimator is a rule for calculating an estimate of a given quantity based on observed data.
- Bias is the error due to overly simple assumptions in the learning algorithm.
- Variance is the error due to too much complexity in the learning algorithm.
- The bias-variance tradeoff is crucial in machine learning.

Bias-Variance Tradeoff Explained

- Bias
- Vari
- Opt
  betv

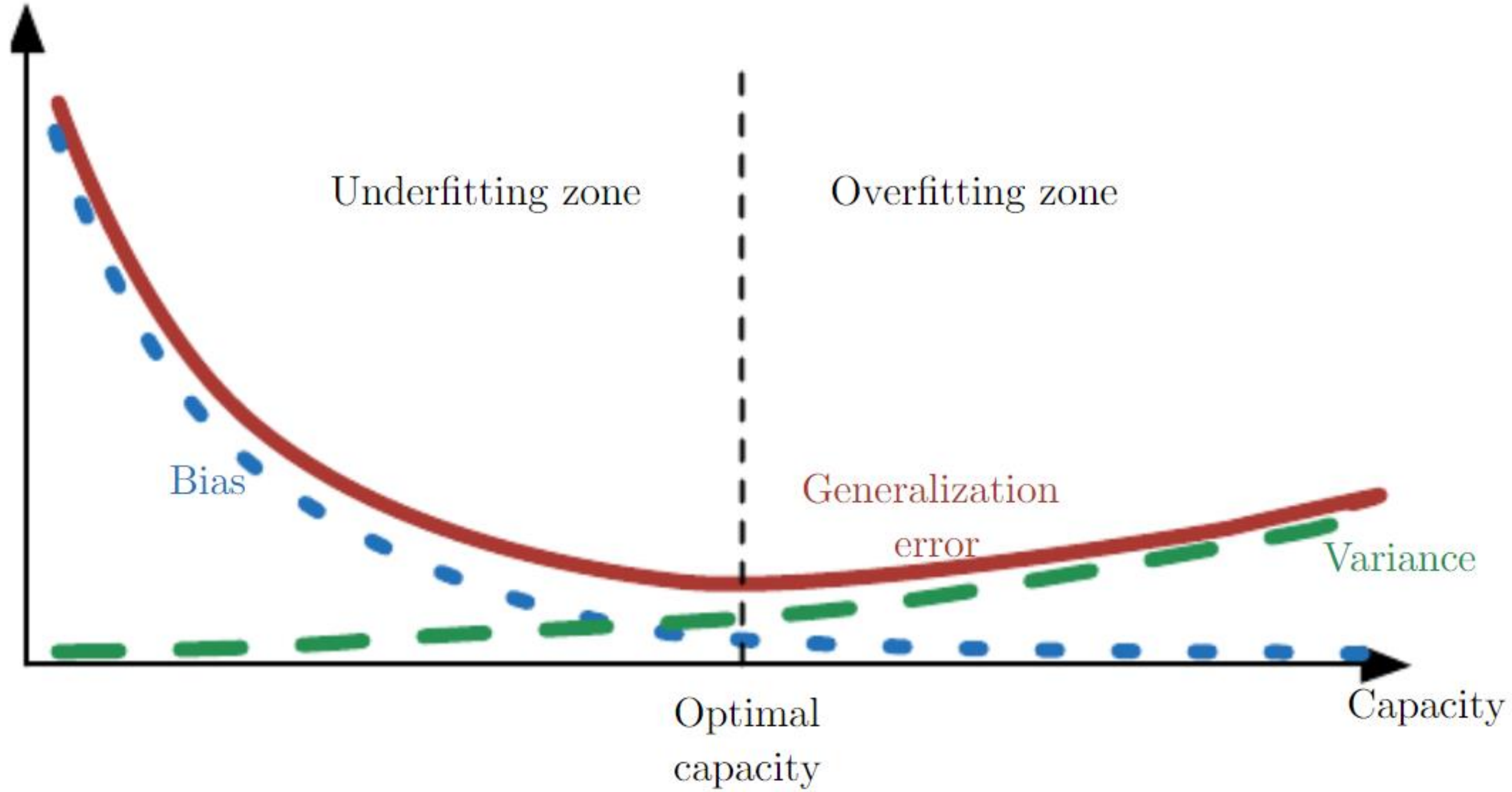


Figure 5.6 in Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

Maximum Likelihood Estimation

- MLE is a method of estimating the parameters of a statistical model.
- It maximizes the likelihood function to find the most probable parameters.
- Used in various machine learning algorithms including logistic regression and naive Bayes.

Curse of Dimensionality and Dimension Reduction

- The Curse of Dimensionality refers to the exponential increase in volume associated with adding extra dimensions to data.
- Dimensionality Reduction: Techniques like PCA can reduce the number of random variables under consideration.

Programming and Data Structures

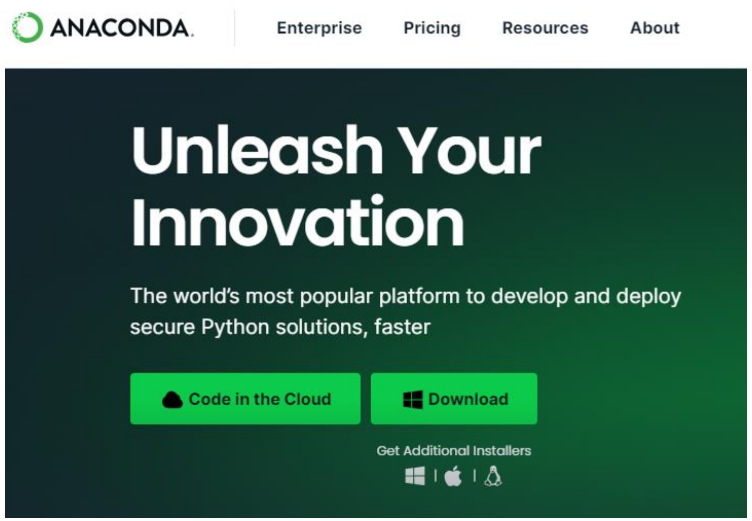
- Python Basics
- Data Structures

Python Basics

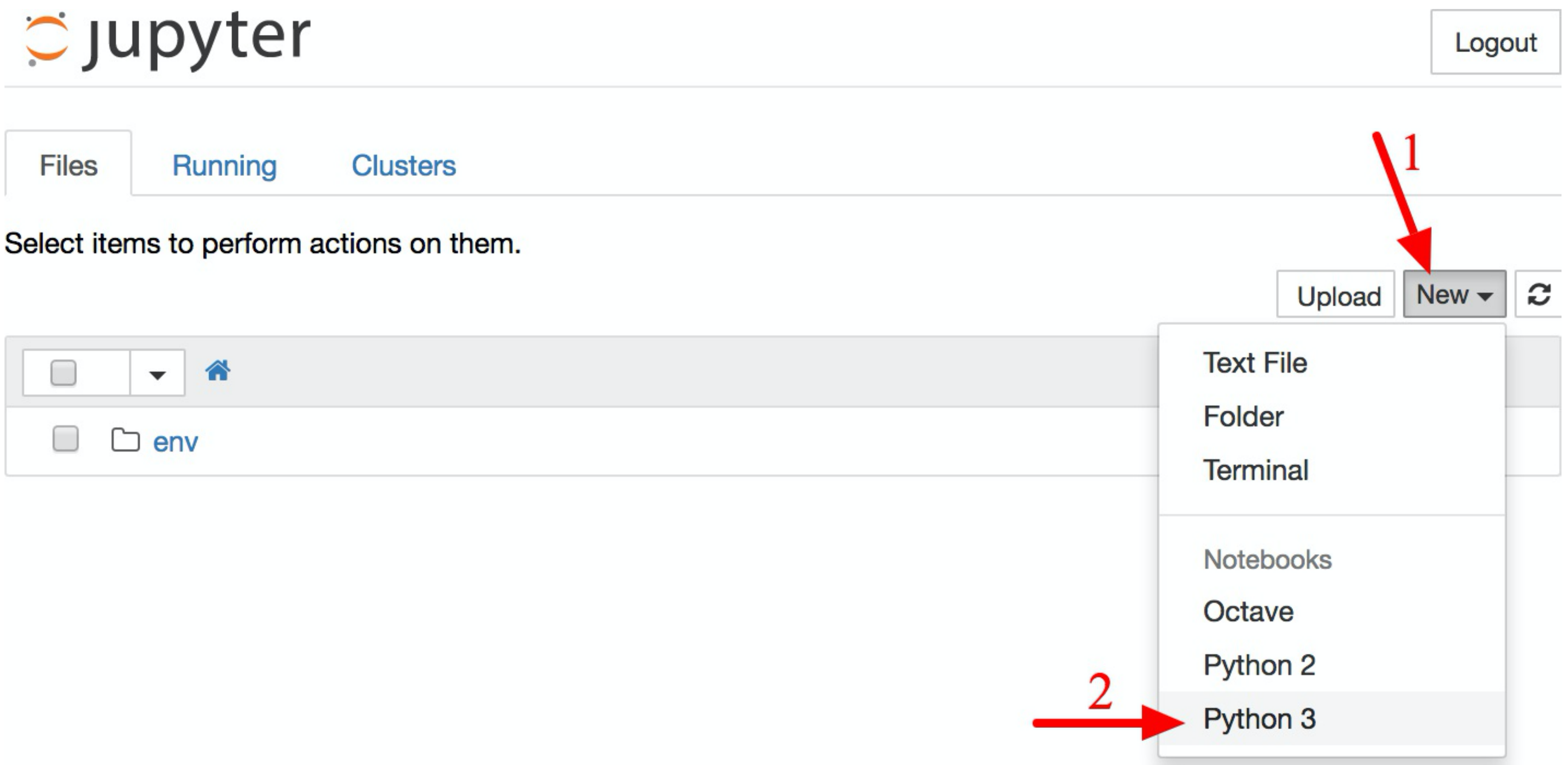
- Python's syntax is clean and its standard library is extensive.
- These features make Python an excellent choice for web development, scripting, data analysis, and, of course, artificial intelligence.

Choice of Python Distro - Anaconda

- There are many Python distributions, such as Cpython, ActivePython, PyPy, but we will use Anaconda.
- Why?
  - It support all different platforms, including Windows, MacOS, and Linux.
  - It comes with many IDE, including jupyter.
  - It is optimized for data analysis and machine learning.
- Download from https://www.anaconda.com/
- Make sure to choose the one w/ python 3.x



Workspace in Jupyter



Data Structures in Python

- Lists, dictionaries, and sets are some of the fundamental Python data structures useful in machine learning.
- Numpy arrays and Pandas dataframes are more specialized data structures that make data manipulation more straightforward.

Deep Learning Libraries

- Keras
- TensorFlow
- PyTorch

TensorFlow

- TensorFlow is an open-source machine learning framework developed by Google Brain.
- It's highly flexible and supports a wide array of machine learning algorithms.
- Strengths: Scalability, TensorBoard for visualization.

Keras

- Keras is a high-level neural networks API running on top of TensorFlow, CNTK, or Theano.
- Strengths: Simplicity, fast prototyping.

PyTorch

- PyTorch is an open-source machine learning framework developed by Facebook's AI Research lab.
- It is known for its dynamic computation graph.
- Strengths: Dynamic computation graph, debugging.

Thank you.