430.457

Introduction to Intelligent Systems

Prof. Songhwai Oh

ECE, SNU

Ch. 19 Learning from Examples

LEARNING

Learning

- **Learning**: to improve agent's performance on future tasks after making observations about the world.
- Components to be learned:
  1. A direct mapping from conditions on the current state to actions.
  2. A means to infer relevant properties of the world from the percept sequence.
  3. Information about the way the world evolves and about the results of possible actions the agent can take.
  4. Utility information indicating the desirability of world states.
  5. Action-value information indicating the desirability of actions.
  6. Goals that describe classes of states whose achievement maximizes the agents utility.
- Representation and prior knowledge: e.g., first-order logic, Bayesian networks; inductive learning vs. deductive learning.

Types of Learning

- Categorized by types of feedback available for learning.
- **Unsupervised learning**: learns patterns in the input even though no explicit feedback is supplied. E.g., clustering, density estimation, anomaly detection.
- **Reinforcement learning**: learns from a series of reinforcements – rewards or punishments.
- **Supervised learning**: observes some example input-output pairs and learns a function that maps from input to output.
- **Semi-supervised learning**: given a few labeled examples and must make what we can of a large collection of unlabeled examples.

Supervised Learning

- Categorized by types of feedback available for learning.
- **Supervised learning problem**: Given a **training set** of N example input-output pairs

$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N),$$

  where each y_j was generated by an unknown function $y = f(x)$, discover a function $h \in \mathcal{H}$ that approximates the true function f.
- Learning is a search through the space of possible hypotheses ($\mathcal{H}$, hypothesis space) for one that will perform well, even on new examples beyond the training set.
- A hypothesis **generalizes** well if it correctly predicts the value of y for new examples (**test set**).
- **Classification** (if $y \in \mathcal{Y}$ and $|\mathcal{Y}|$ is finite); **Regression** (if y is continuous).
- Supervised learning can be posed as a problem of finding h^* such that

$$h^* = \arg\max_{h \in \mathcal{H}} P(h|data) = \arg\max_{h \in \mathcal{H}} P(data|h)P(h)$$

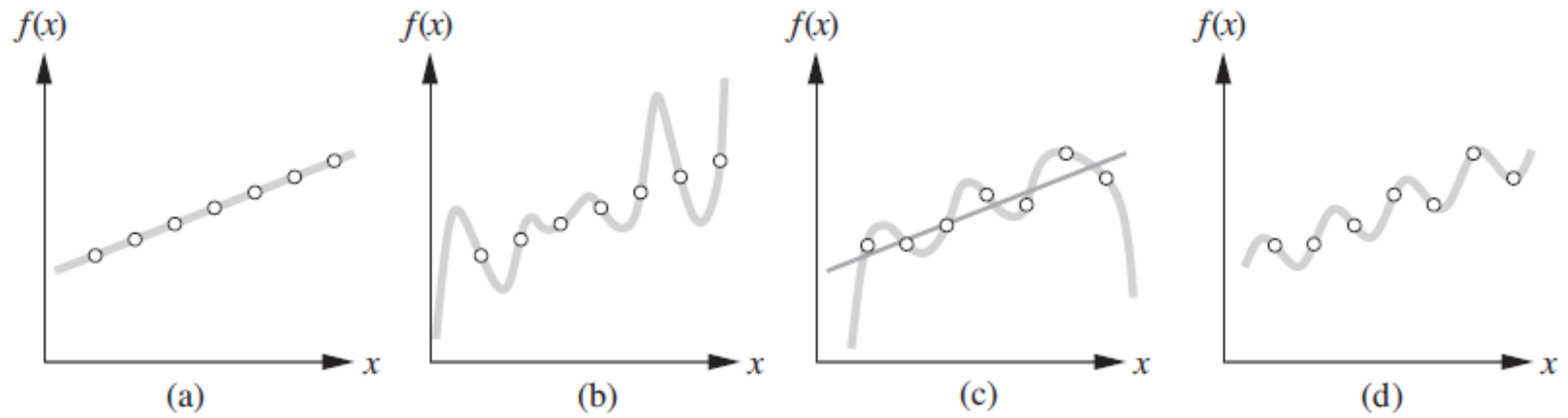

Figure 18.1 (a) Example $(x, f(x))$ pairs and a consistent, linear hypothesis. (b) A consistent, degree-7 polynomial hypothesis for the same data set. (c) A different data set, which admits an exact degree-6 polynomial fit or an approximate linear fit. (d) A simple, exact sinusoidal fit to the same data set.

- **Ockhams razor**: prefer the simplest hypothesis consistent with the data over more complex hypotheses.
- **Expressiveness-complexity tradeoff**: tradeoff between the expressiveness of a hypothesis space and the complexity of finding a good hypothesis within that space.