Artificial Intelligence

Learning Agent

Outline

- Learning agents
- Learning Types
- Inductive learning
- Supervised inductive learning
- Inductive bias

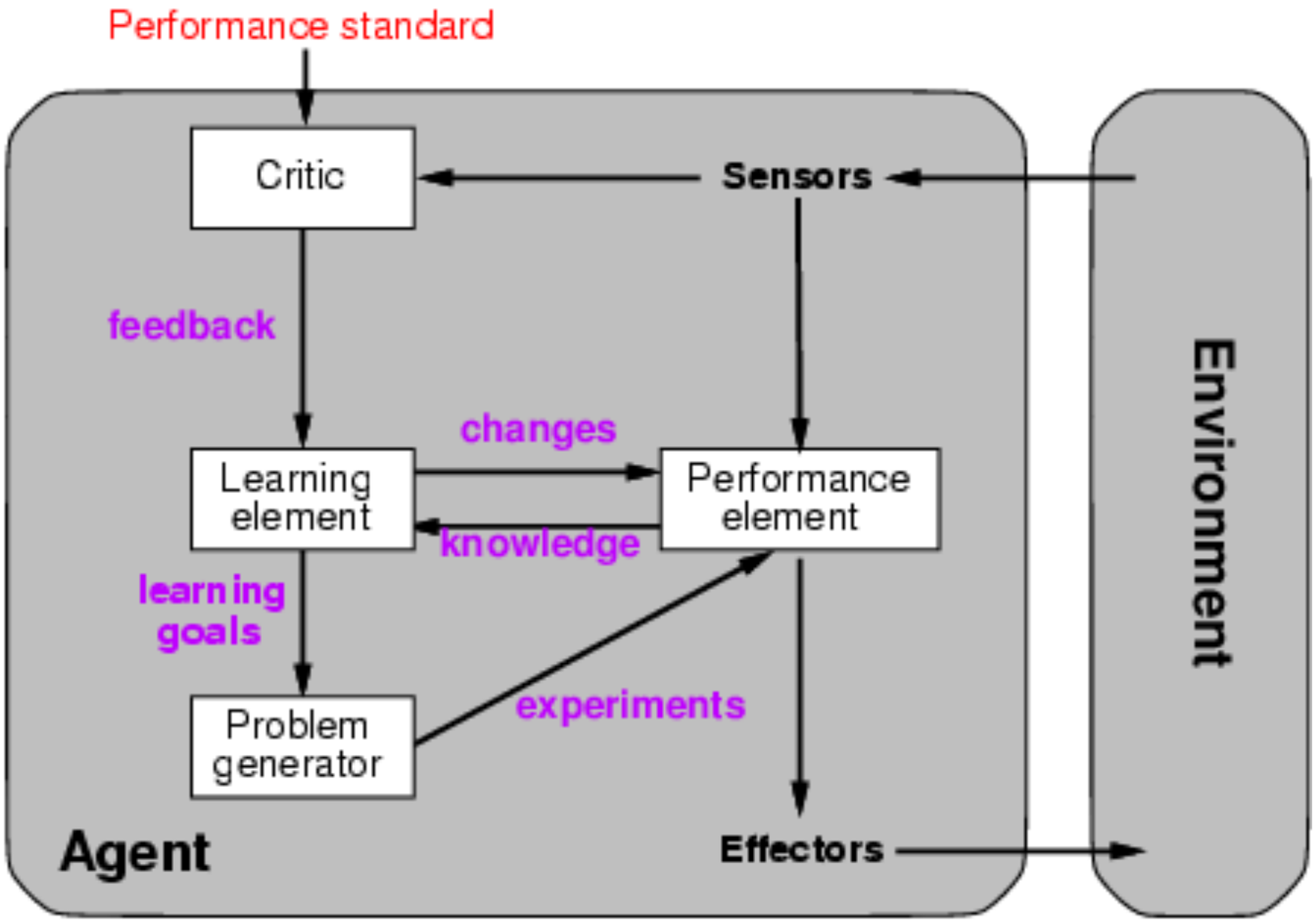
Learning

- Learning is essential for unknown environments especially when designer lacks omniscience.

- Learning is useful as a system construction method. A learning agent is exposed to reality to gain system functionality rather than having its system fixed at the beginning.

- Learning modifies the learning agent's decision mechanisms to improve its performance.

Learning agents



Learning element

- Design of a learning element is affected by
  - Which components of the performance element are to be learned
  - What feedback is available to learn these components
  - What representation is used for the components
- Goal of learning: establish a world model that can to used to predict future events

Types of Learning

- Deductive Learning: instructor-centred learning
    - Concepts, rules and generalization are imported into the learner from instructors/designers/higher beings;
    - Learner then needs to apply them.
- Inductive Learning: learn by observing examples
    - Supervised learning: correct answer for each example
    - Unsupervised learning: correct answers not given
    - Reinforced learning: occasional rewards

Supervised Learning

- Learning agent using supervised Learning techniques learns a model from a set of historical examples/observations/dataset.

- Typical steps:
  - Build a model (learning)
  - Calibrate the model (validation)
  - Use the model (deployment)

- Ultimate goal of learning: Construct a model that generalizes beyond the given dataset and that isn’t influenced by the noise in the dataset.

What does a model look like?

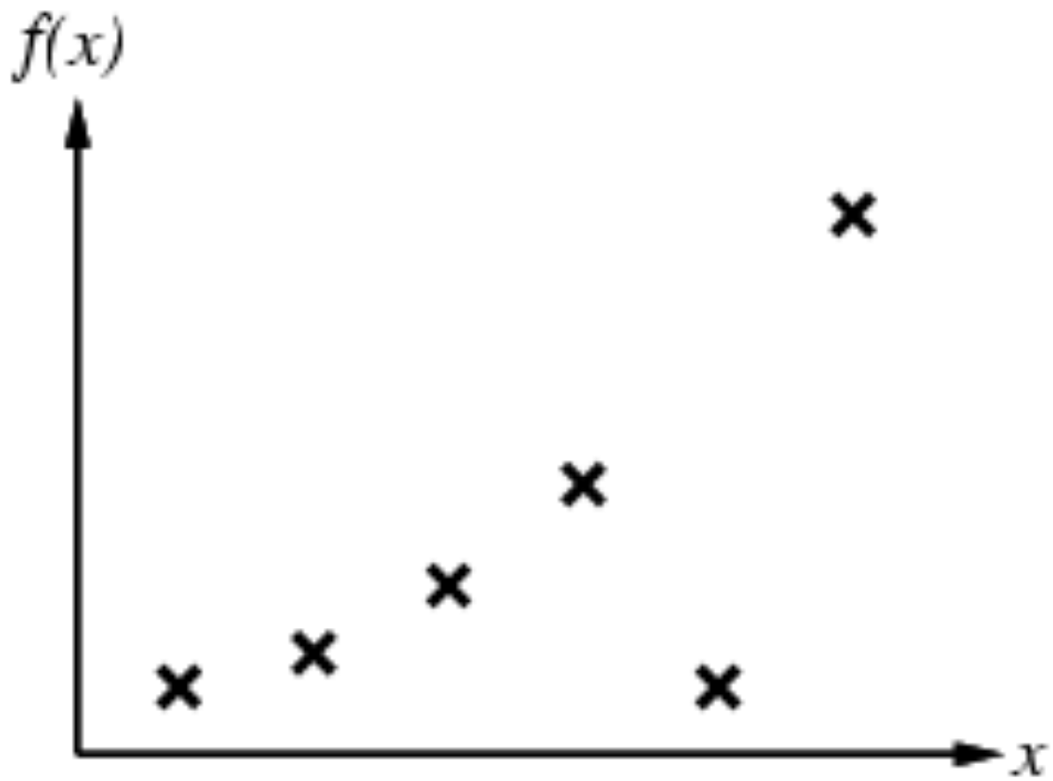
- A mathematical function
- A table
- A set of rules/logical expression
- A decision tree
- A diagram
- A program
- A black box
- …

Mathematical Representation of Inductive learning

- Simplest form: learn a function from examples
  - f is the target function
  - An example is a pair (x, f(x))
- Problem: find a hypothesis function h such that h ≈ f given a training set of examples
- This is a highly simplified model of real learning
  - Ignores prior knowledge
  - Assumes examples are given

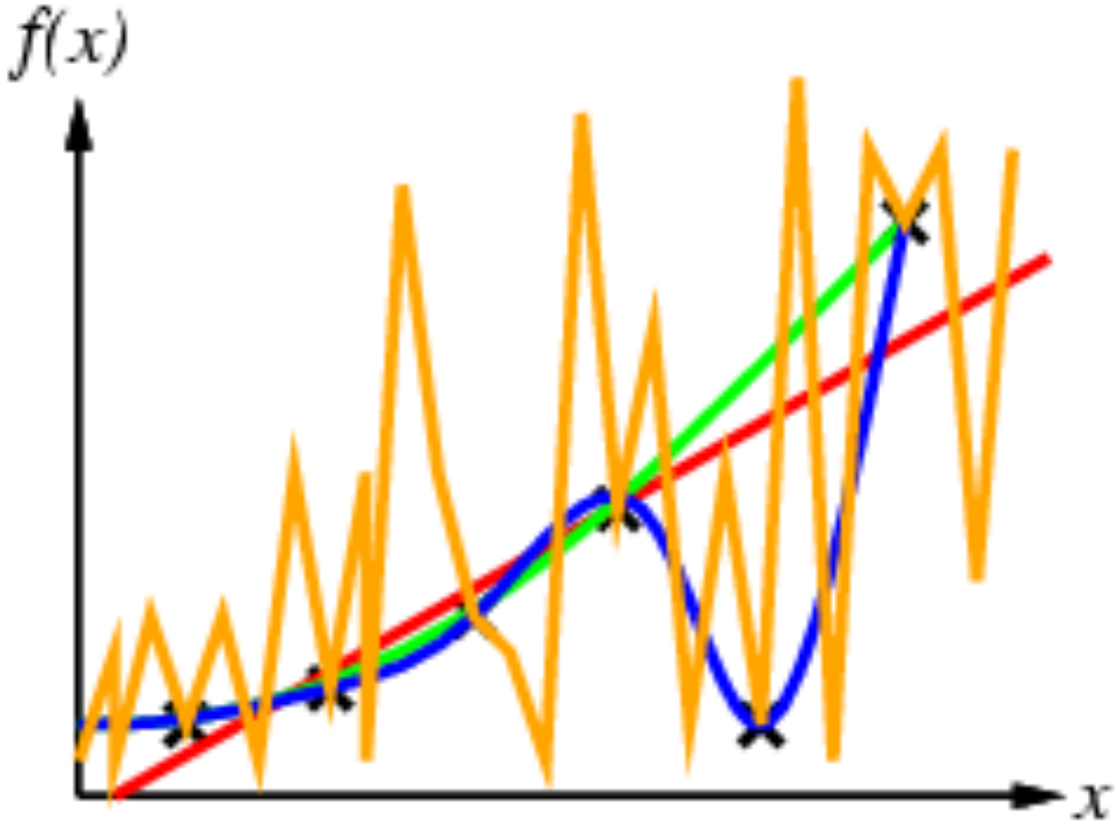
Inductive learning method

- Construct/adjust model h by searching through a set of possible models for the model that best captures the relationship between the input and the output



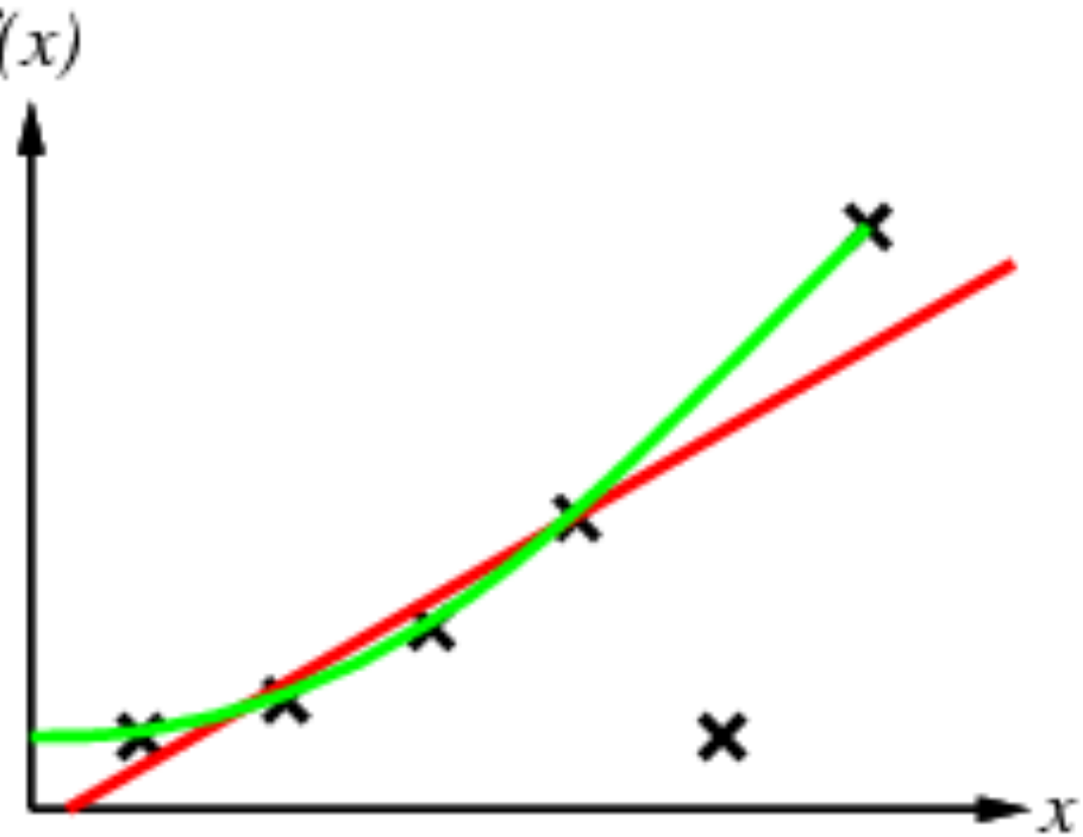
Inductive learning method

- h is consistent if it agrees with f on all examples



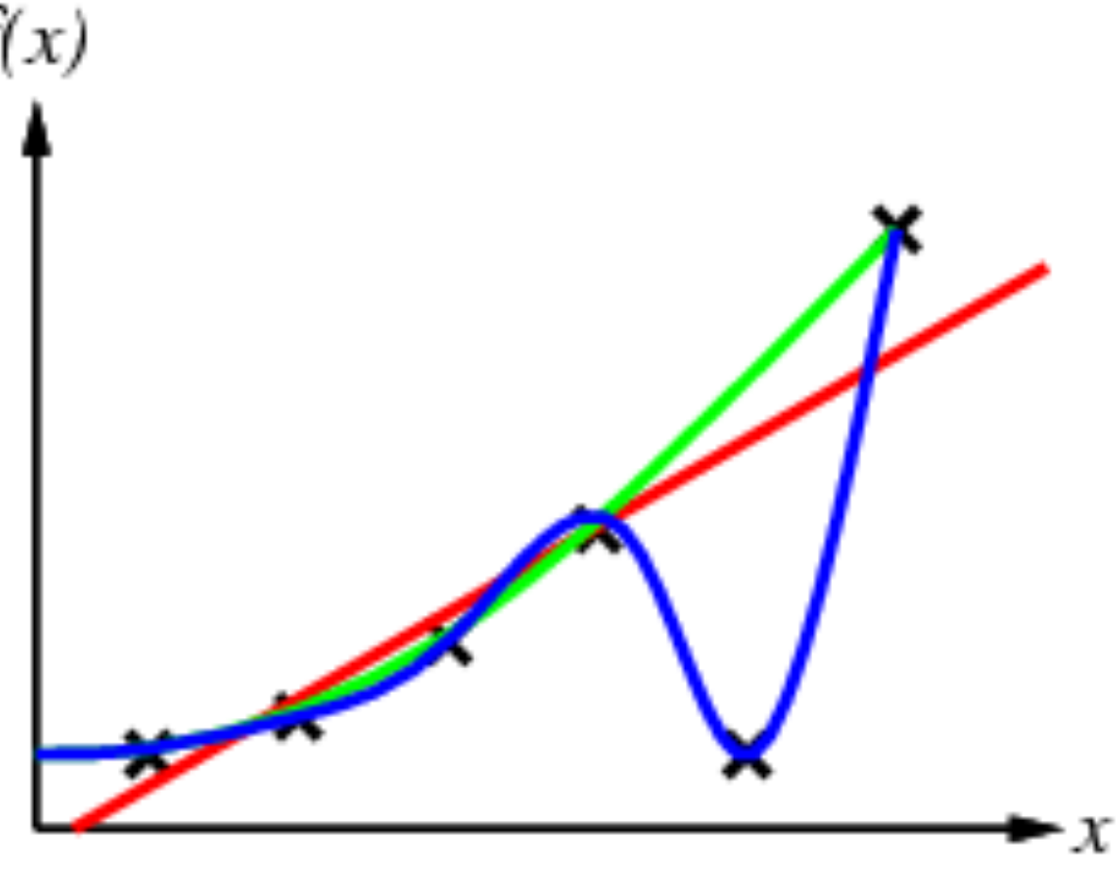
Inductive learning method

- Consistent model while ignoring noise in the dataset



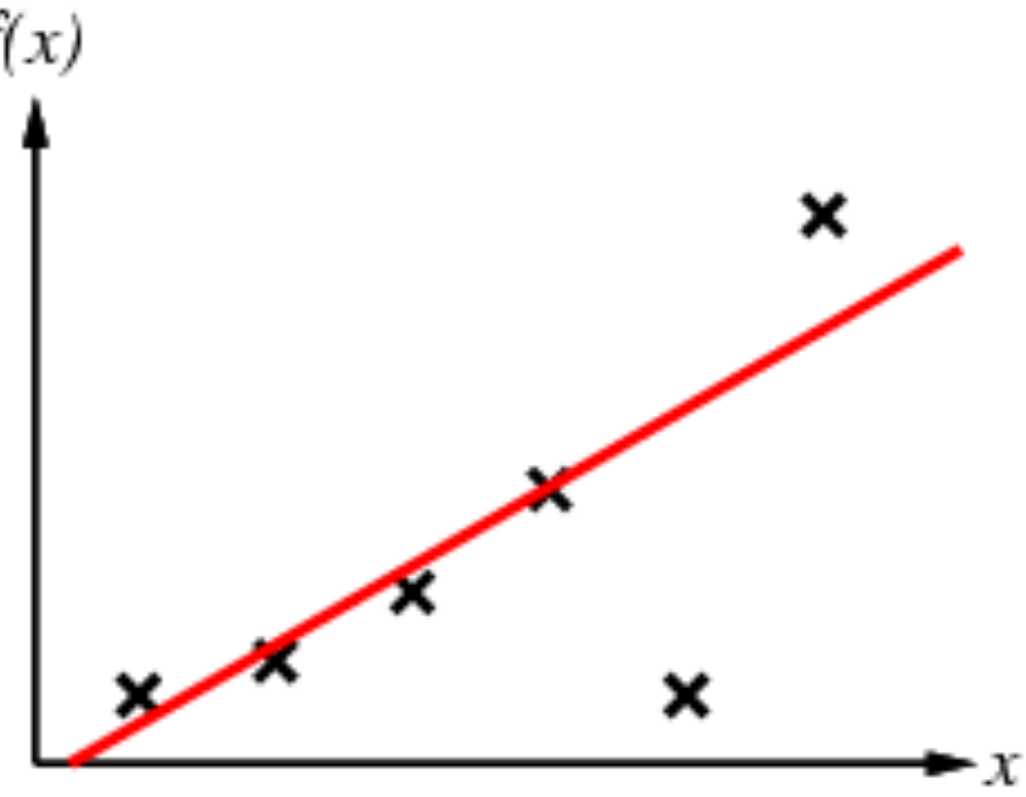
Inductive learning method

- Consistent model assuming there is no noise in the given dataset



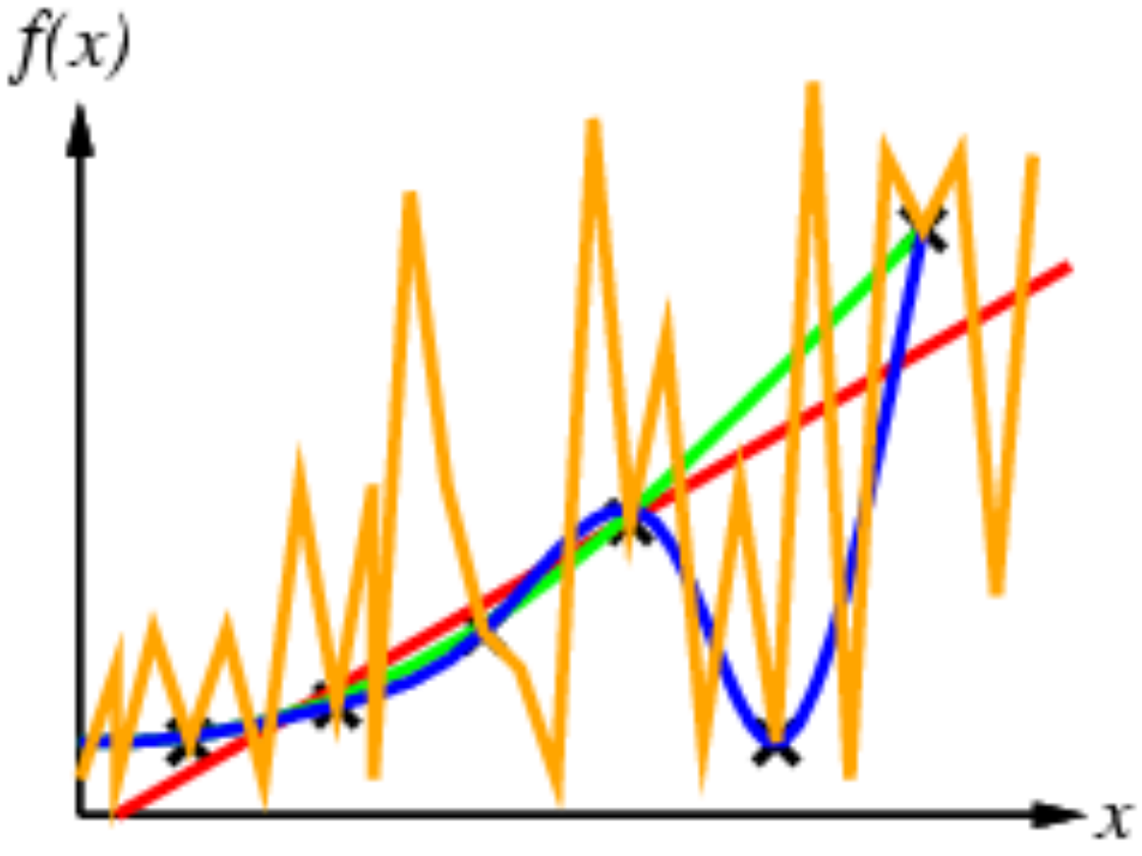
What can go wrong?

- Under-fitting



What can go wrong?

- Over-fitting



Inductive Bias

- The set of assumptions that define the model selection criteria of an inductive learning algorithm
    - The assumptions are necessary in learning since our agent can’t handle endlessly arbitrary situations.

- There are two types of bias:
    - Restriction bias: such as pre-determined model type restricts the set of hypothesis that will be considered.
    - Preference bias: some models are systematically preferred over others (with or without a reason)
        - Ockham’s razor: prefer the simplest hypothesis consistent with data

Summary

- Learning needed for unknown environments, lazy designers
- Learning agent = performance element + learning element
- For supervised learning, the aim is to find a simple hypothesis approximately consistent with training examples
- Inductive learning algorithms work by searching through sets of potential models.
- There are two sources of information that guide this search:
  - The training data
  - The inductive bias of the algorithm
- Striking the right balance between model complexity and simplicity (between under-fitting and over-fitting) is the hardest part of inductive learning.