Introduction to Learning

Alice Gao
Lecture 18-1

Based on work by K. Leyton-Brown, K. Larson, and P. van Beek

Outline

Learning Goals

Introduction to Learning

Supervised Learning

Overfitting

Revisiting the Learning goals

Learning Goals

By the end of the lecture, you should be able to

- Identify reasons for building an agent that can learn.
- Describe different types of learning.
- Define supervised learning, classification, and regression.
- For supervised learning, describe how we choose a hypothesis.
- Define and describe overfitting.

Agents that learn

We want to build agents that learn from their experience to improve their performance over time

Why would we want an agent to learn?

- Cannot anticipate all possible situations
- Cannot anticipate all changes over time
- No idea how to program a solution

One class of learning problems

from a collection of input–output pairs,
learn a function that predicts the output for new inputs.

Types of learning

Based on types of feedback available, there are different types of learning:

- Unsupervised learning: learn patterns in the input, no output available (or the output is the same for all data)
  clustering: detecting clusters of input examples
- Supervised learning: learn a function that maps from input to output
- Semi-supervised learning: learn from a few labeled examples and a large number of unlabeled examples, dealing with noise and lack of labels.

Supervised Learning

- Input: A collection of training examples

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

  which were generated by an unknown function $y = f(x)$.
  x_i are tuples of features, and y_is are some outcome.
  Hypothesis space H
- Output: A hypothesis $h \in H$ that approximates f.

Accuracy of a Hypothesis

We want a hypothesis h that generalizes well
— h correctly the value of y for unseen examples.

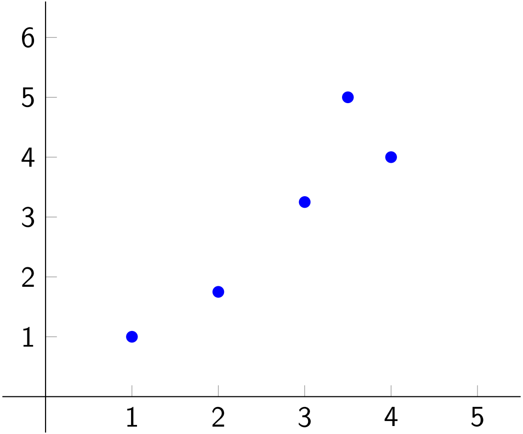
To achieve this, we measure the accuracy of the hypothesis on a test set of examples that are distinct from the training set.

Classification v.s. Regression

- Classification: the output variable has discrete values.
- Regression: the output variable has continuous values.

An example of regression

Training set: (1, 1), (2, 1.75), (3, 3.25), (3.5, 5), (4, 4)



Hypotheses

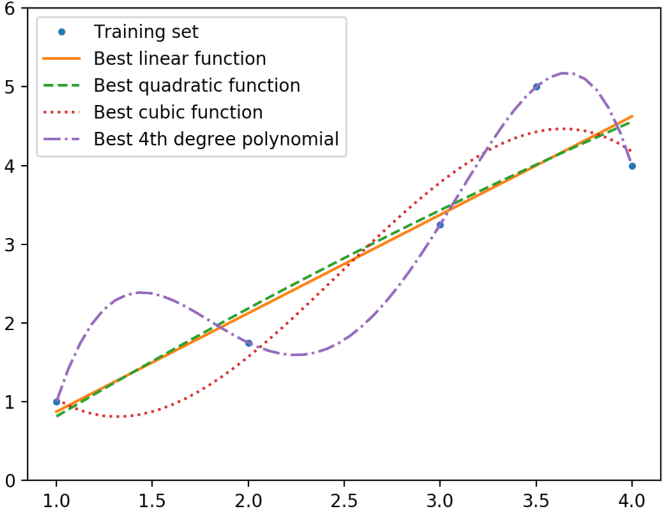
What are the possible hypothesis spaces H?

- straight line
- piece-wise linear function
- quadratic function
- cubic function

What is the best hypothesis in each space?

- consistent with all the training examples
- minimize error on training examples

Possible hypotheses



Choosing a hypothesis

We prefer the simplest hypothesis that agrees with all of the data points (or makes small errors)

- Generalizes well
- Easy to compute

Ockham's Razor (English philosophy William of Ockham)

> *The simplest solution tends to be the correct one.*

Over-fitting

As the hypothesis h becomes more complex

- The error on the training set decreases.
- The error on the test set decreases and and then increases.

Definition: Given a hypothesis space H,
a hypothesis $h \in H$ is said to over-fit the training data
if there exists some hypothesis $h' \in H$ such that

- h has smaller error than h' over the training set.
- h' has smaller error than h over all the data points.

Revisiting the Learning Goals

By the end of the lecture, you should be able to

- Identify reasons for building an agent that can learn.
- Describe different types of learning.
- Define supervised learning, classification, and regression.
- For supervised learning, describe how we choose a hypothesis.
- Define and describe overfitting.