Introduction to Machine Learning

Blake VanBerlo

Lecture 12

Readings: RN 19.1, 19.2. PM 7.1, 7.2.

Outline

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.
- Define bias, variance, and describe the trade-off between the two.
- Describe how prevent overfitting by performing cross validation.

Learning Goals

Introduction to Learning

Supervised Learning

Revisiting the Learning goals

Applications

- Medical diagnosis
- Spam filtering
- Facial recognition
- Speech understanding
- Handwriting recognition

Agents that learn

Learning is the ability of an agent to improve its performance on future tasks based on experience.

$\rightarrow$ The agent needs to remember its past in a way that is useful for its future.

We want an agent to

- Do more $\rightarrow$ The range of behaviours is expanded.
- Do things better $\rightarrow$ The accuracy on tasks is improved.
- Do things faster $\rightarrow$ The speed is improved.

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

The Learning Architecture

- Problem/Task

  → The behaviour or task that is being improved.

- Experiences/Data

  → The experiences that are being used to improve performance in the task.

- Background knowledge/Bias

- Measure of improvement

  → How can the improvement be measured? e.g. increasing accuracy in prediction, new skills that were not present initially, improved speed.

Types of learning problems

- **Supervised learning:**

  Given input features, target features, and training examples, predict the value of the target features for new examples given their values on the input features.

- **Unsupervised learning:**

  Learning classifications when the examples do not have targets defined.

  E.g. clustering, dimensionality reduction

- **Reinforcement Learning:**

  Learning what to do based on rewards and punishments.

Q: Supervised or Unsupervised Learning

Q #1: We are given information on a user's credit card transactions. We would like to detect whether some of the transactions are fraudulent by finding some transactions that are different from the other transactions. We have no information on whether any particular transaction is fraudulent or not.

Is this a supervised or unsupervised learning problem?

(A) Supervised learning

(B) Unsupervised learning

Q: Supervised or Unsupervised Learning

Q #1: We are given information on a user's credit card transactions. We would like to detect whether some of the transactions are fraudulent by finding some transactions that are different from the other transactions. We have no information on whether any particular transaction is fraudulent or not.

Is this a supervised or unsupervised learning problem?

(A) Supervised learning

(B) Unsupervised learning

→ (B) is correct. We do not have values of the target features.

Q: Supervised or Unsupervised Learning

Q #2: You are given historical data on the weather condition (sunny, cloudy, rain, or snow) on a particular day of the year. You want to predict the weather condition on this day next year.

Is this a supervised or unsupervised learning problem?

(A) Supervised learning

(B) Unsupervised learning

Q: Supervised or Unsupervised Learning

Q #2: You are given historical data on the weather condition (sunny, cloudy, rain, or snow) on a particular day of the year. You want to predict the weather condition on this day next year.

Is this a supervised or unsupervised learning problem?

(A) Supervised learning

(B) Unsupervised learning

→ (A) is correct. We have values of the target feature (i.e. weather condition).

Two types of supervised learning problems

- **Classification:** target features are discrete.

  $\rightarrow$ E.g. Predicting whether an image contains a dog or a cat

- **Regression:** target features are continuous.

  $\rightarrow$ E.g. Predicting tomorrow's temperature

Q: Classification or regression

Q #3: Is the following problem an instance of classification or regression?

You are given historical data on the weather condition (sunny, cloudy, rain, or snow) on a particular day of the year. You want to predict the weather condition of this day next year.

(A) Classification

(B) Regression

(C) This is not supervised learning.

Q: Classification or regression

Q #3: Is the following problem an instance of classification or regression?

You are given historical data on the weather condition (sunny, cloudy, rain, or snow) on a particular day of the year. You want to predict the weather condition of this day next year.

(A) Classification

(B) Regression

(C) This is not supervised learning.

→ A) is correct. We are predicting the value of a discrete variable.

Q: Classification or regression

Q #4: Is the following problem classification or regression?

You are given historical data on the price of a house at several points in time. You want to predict the price of this house next month.

(A) Classification

(B) Regression

(C) This is not supervised learning.

Q: Classification or regression

Q #4: Is the following problem classification or regression?

You are given historical data on the price of a house at several points in time. You want to predict the price of this house next month.

(A) Classification

(B) Regression

(C) This is not supervised learning.

$\rightarrow$ B) is correct. We are predicting the value of a continuous variable.

Supervised Learning

- Given training examples of the form $(x, f(x))$

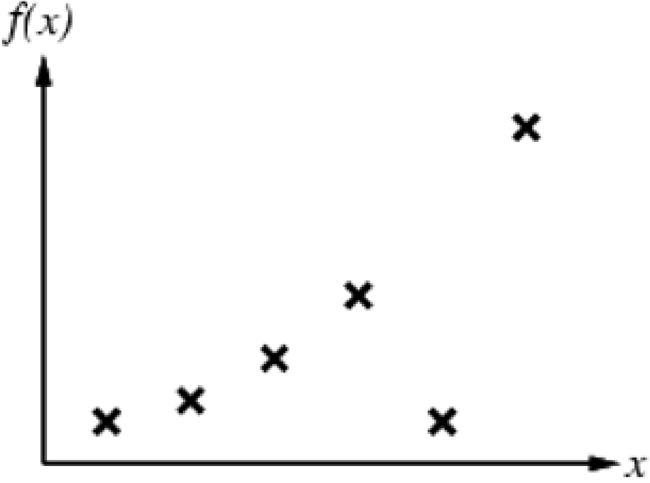
  $\rightarrow$ We assume that f exists. We don't have f. We never observe f.

- Return a function h (a.k.a a hypothesis) that approximates the true function f.

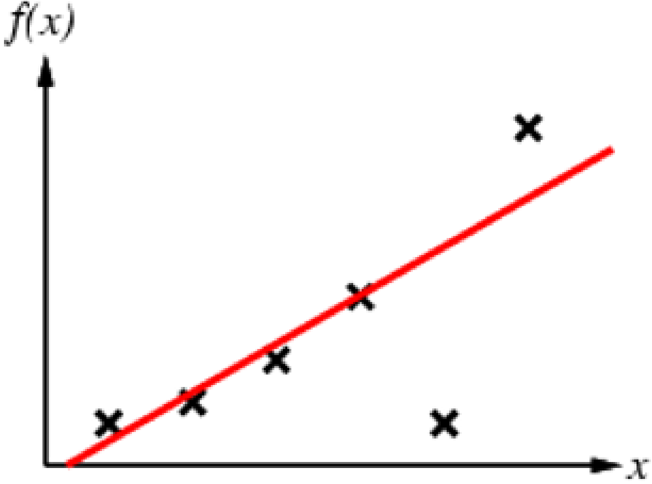
Learning as a search problem

- Given a hypothesis space, learning is a search problem.
- Search space is prohibitively large for systematic search.
- ML techniques are often some forms of local search.

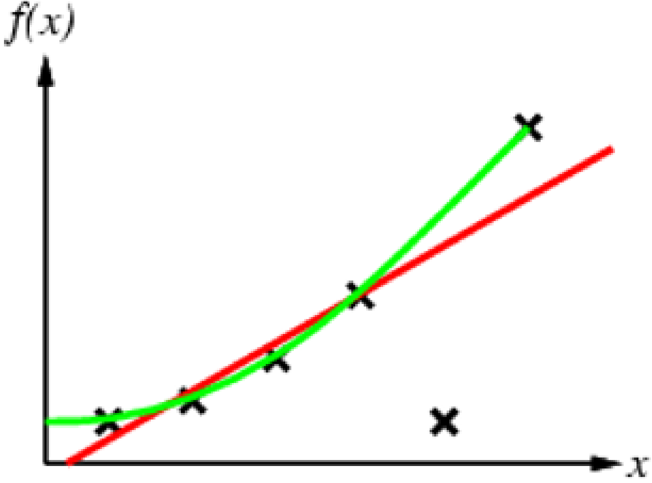
Example: A prediction task. Fit a polynomial.



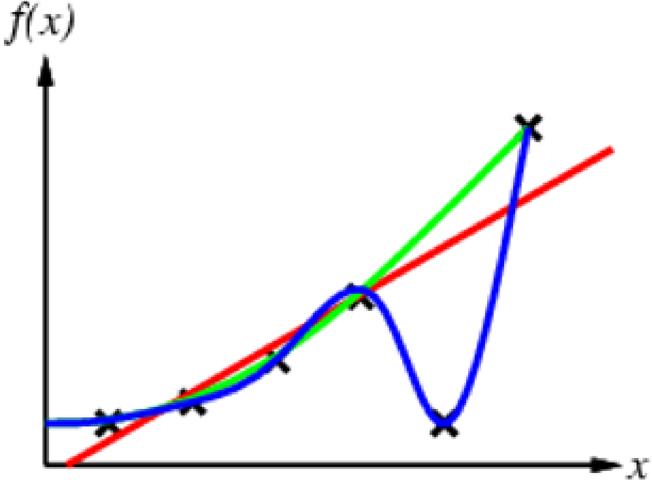
Example: A prediction task. Fit a polynomial.



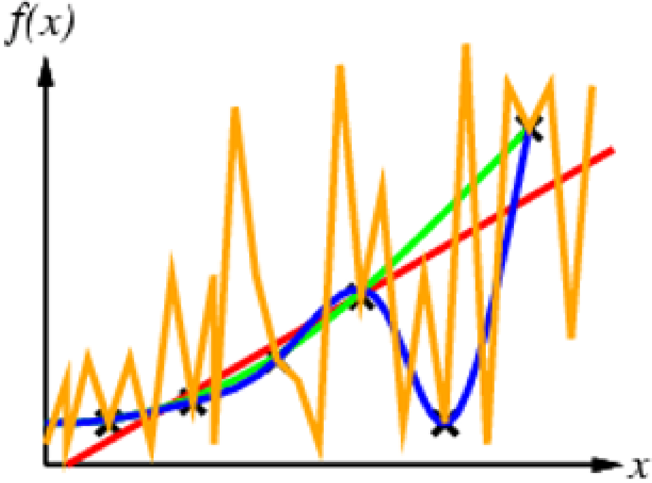
Example: A prediction task. Fit a polynomial.



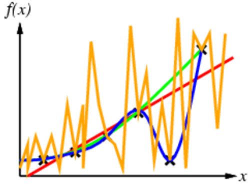
Example: A prediction task. Fit a polynomial.



Example: A prediction task. Fit a polynomial.



Example: A prediction task. Fit a polynomial.



- Which function is the correct one?
- If some data points are outliers, or the data is noisy, then the simpler curves are better.
- If we are predicting stock market prices, perhaps the more complex curve is better.
- There is no perfect answer. All curves can be justified as the correct one from some perspective.
- *No free lunch theorem:* In order to learn something useful, we have to make some assumptions — have an inductive bias.
- Assumptions can include: Do I have any outliers? Does the curve follow a particular parametric form?

Generalization

- Goal of ML is to find a hypothesis that can predict unseen examples correctly.

  → Goal is not to predict the data we already have correctly. This makes ML difficult but exciting.

- How do we choose a hypothesis that generalizes well?
  - *Ockham's razor*

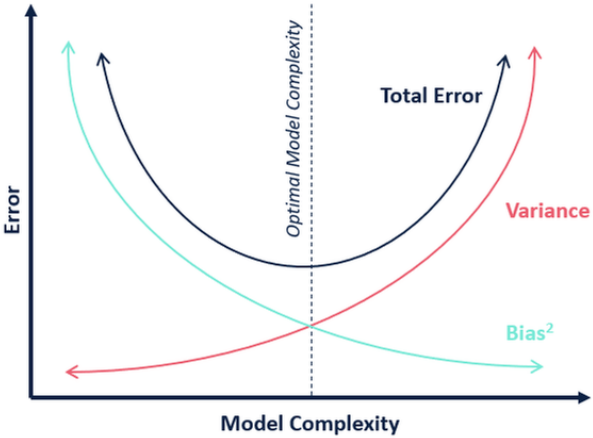
    → prefer the simplest hypothesis consistent with the data.
  - *Cross-validation*

    → a more principled approach to choose a hypothesis.

- A trade-off between
  - complex hypotheses that fit the training data well
  - simpler hypotheses that may generalize better

Bias-Variance Trade-off

How well does the hypothesis fit the data as the hypothesis becomes more complex?



Bias-Variance Trade-off

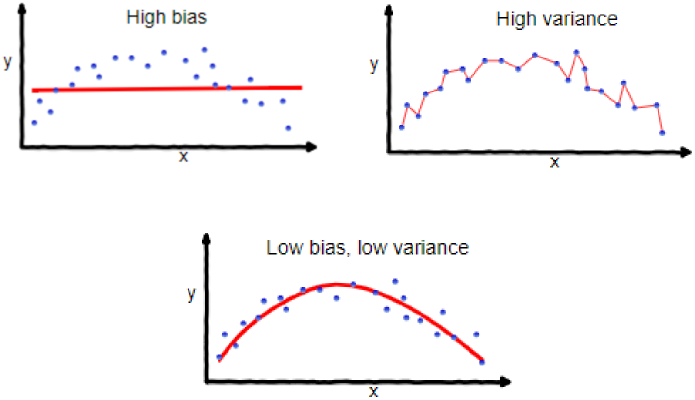
- **Bias:** If I have infinite data, how well can I fit the data with my learned hypothesis?

  A hypothesis with high bias makes strong assumptions, too simplistic, has few degrees of freedom, does not fit the training data well.

- **Variance:** How much does the learned hypothesis vary given different training data?

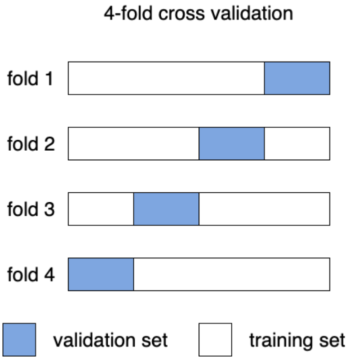
  A hypothesis with high variance has a lot of degrees of freedom, is very flexible, and fits the training data well. Whenever the training data changes, the hypothesis changes a lot.

Bias-Variance Trade-off



Cross-validation

How do we find a hypothesis that has low bias and low variance?
Use cross validation.



→ Use part of the training data as a surrogate for test data (called validation data).

Use validation data to choose the hypothesis.

K-fold Cross Validation

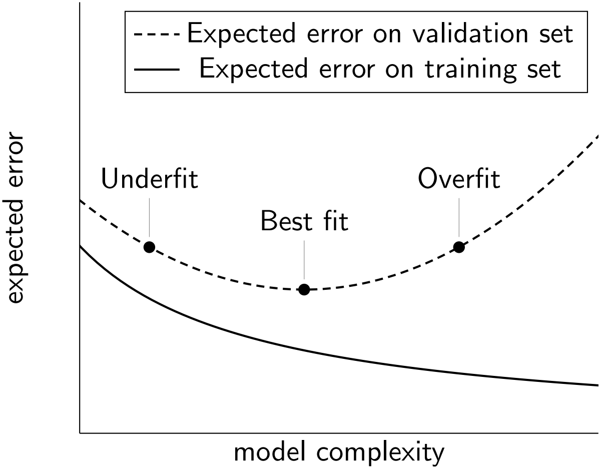
1. Break training data into K equally sized partitions.
2. Train a learning algorithm on $K - 1$ partitions (training set).
3. Test on the remaining 1 partition (validation set).
4. Do this K times, each time testing on a different partition.
5. Calculate the average error on the K validation sets.

After cross validation

After running cross validation, you can:

- Select one of the K trained hypotheses as your final hypothesis.
- Train a new hypothesis on all of the data, using parameters selected by cross validation.

Overfitting



Q: Which hypothesis is prone to overfitting?

Q #5: Suppose that we are considering a simple hypothesis (a straight line) and a complex hypothesis (a 4$^{\text{th}}$ degree polynomial). Which of the two is more likely to overfit the training data?

(A) The simple hypothesis

(B) The complex hypothesis

(C) I don't know

Q: Which hypothesis is prone to overfitting?

Q #5: Suppose that we are considering a simple hypothesis (a straight line) and a complex hypothesis (a 4^{th} degree polynomial). Which of the two is more likely to overfit the training data?

(A) The simple hypothesis

(B) The complex hypothesis

(C) I don't know

$\rightarrow$ (B) is the correct answer.

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.
- Define bias, variance, and describe the trade-off between the two.
- Describe how prevent overfitting by performing cross validation.