Outline of the Course

1. The Learning Problem (April 3)
2. Is Learning Feasible? (April 5)
3. The Linear Model I (April 10)
4. Error and Noise (April 12)
5. Training versus Testing (April 17)
6. Theory of Generalization (April 19)
7. The VC Dimension (April 24)
8. Bias–Variance Tradeoff (April 26)
9. The Linear Model II (May 1)
10. Neural Networks (May 3)

11. Overfitting (May 8)
12. Regularization (May 10)
13. Validation (May 15)
14. Support Vector Machines (May 17)
15. Kernel Methods (May 22)
16. Radial Basis Functions (May 24)
17. Three Learning Principles (May 29)
18. Epilogue (May 31)

- theory; mathematical
- technique; practical
- analysis; conceptual
Learning From Data

Yaser S. Abu-Mostafa
California Institute of Technology

Lecture 1: The Learning Problem
The learning problem – Outline

- Example of machine learning
- Components of Learning
- A simple model
- Types of learning
- Puzzle
Example: Predicting how a viewer will rate a movie

10% improvement = 1 million dollar prize

The essence of machine learning:

- A pattern exists.
- We cannot pin it down mathematically.
- We have data on it.
**Movie rating - a solution**

**Viewer**
- Does the viewer like comedies?
- Does the viewer like action?
- Does the viewer prefer blockbusters?
- Does the viewer like Tom Cruise?

**Movie**
- Blockbuster content?
- Comedy content?
- Action content?

**Process**
1. Match movie and viewer factors.
2. Add contributions from each factor.
3. Predicted rating.

© Creator: Yaser Abu-Mostafa - LFD Lecture 1
The learning approach

viewer

movie

rating

LEARNING
**Components of learning**

**Metaphor:** Credit approval

**Applicant information:**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>23 years</td>
</tr>
<tr>
<td>gender</td>
<td>male</td>
</tr>
<tr>
<td>annual salary</td>
<td>$30,000</td>
</tr>
<tr>
<td>years in residence</td>
<td>1 year</td>
</tr>
<tr>
<td>years in job</td>
<td>1 year</td>
</tr>
<tr>
<td>current debt</td>
<td>$15,000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Approve credit?
Components of learning

Formalization:

- **Input**: \( x \)  
  *(customer application)*

- **Output**: \( y \)  
  *(good/bad customer?)*

- **Target function**: \( f : \mathcal{X} \rightarrow \mathcal{Y} \)  
  *(ideal credit approval formula)*

- **Data**: \((x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\)  
  *(historical records)*

- **Hypothesis**: \( g : \mathcal{X} \rightarrow \mathcal{Y} \)  
  *(formula to be used)*
HYPOTHESIS SET

\( \mathcal{H} \)

(set of candidate formulas)

LEARNING ALGORITHM

\( \mathcal{A} \)

FINAL HYPOTHESIS

\( g \approx f \)

(final credit approval formula)

TRAINING EXAMPLES

\((x_1, y_1), \ldots, (x_N, y_N)\)

(historical records of credit customers)

UNKNOWN TARGET FUNCTION

\( f: X \to Y \)

(ideal credit approval function)
The 2 solution components of the learning problem:

- The Hypothesis Set
  \[ \mathcal{H} = \{ h \} \quad g \in \mathcal{H} \]

- The Learning Algorithm

Together, they are referred to as the learning model.
A simple hypothesis set - the ‘perceptron’

For input $\mathbf{x} = (x_1, \cdots, x_d)$ ‘attributes of a customer’

Approve credit if $\sum_{i=1}^{d} w_i x_i > \text{threshold}$,

Deny credit if $\sum_{i=1}^{d} w_i x_i < \text{threshold}$.

This linear formula $h \in \mathcal{H}$ can be written as

$$h(\mathbf{x}) = \text{sign} \left( \left( \sum_{i=1}^{d} w_i x_i \right) - \text{threshold} \right)$$
\[ h(\mathbf{x}) = \text{sign} \left( \left( \sum_{i=1}^{d} w_i x_i \right) + w_0 \right) \]

Introduce an artificial coordinate \( x_0 = 1 \):

\[ h(\mathbf{x}) = \text{sign} \left( \sum_{i=0}^{d} w_i x_i \right) \]

In vector form, the perceptron implements

\[ h(\mathbf{x}) = \text{sign}(\mathbf{w}^\top \mathbf{x}) \]
A simple learning algorithm – PLA

The perceptron implements
\[ h(\mathbf{x}) = \text{sign}(\mathbf{w}^\top \mathbf{x}) \]

Given the training set:
\[ (\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N) \]

pick a misclassified point:
\[ \text{sign}(\mathbf{w}^\top \mathbf{x}_n) \neq y_n \]

and update the weight vector:
\[ \mathbf{w} \leftarrow \mathbf{w} + y_n \mathbf{x}_n \]
Iterations of PLA

- One iteration of the PLA:
  \[ \mathbf{w} \leftarrow \mathbf{w} + y \mathbf{x} \]
  where \((\mathbf{x}, y)\) is a misclassified training point.

- At iteration \(t = 1, 2, 3, \ldots\), pick a misclassified point from
  \((\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\)
  and run a PLA iteration on it.

- That's it!
The learning problem - Outline

- Example of machine learning
- Components of learning
- A simple model
- Types of learning
- Puzzle
Basic premise of learning

“using a set of observations to uncover an underlying process”

broad premise $\implies$ many variations

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
Supervised learning

Example from vending machines – coin recognition
Unsupervised learning

Instead of \((\text{input, correct output}), \) we get \((\text{input, ?})\)
Reinforcement learning

Instead of \((\text{input, correct output})\), we get \((\text{input, some output, grade for this output})\).

The world champion was a neural network!
A Learning puzzle

\[ f = -1 \]
\[ f = +1 \]
\[ f = ? \]