Machine Learning as Function Approximation
Q: Should I go outside today?

A: Absolutely, yes! Unless it’s this Thursday morning…
Q: In Lecture 1, why did we use the term *experience* instead of just *data*?

A: Because our concern isn’t just the data itself, but also where the data comes from (e.g. an agent interacting with the world vs. knowledge from a book).

As well, the word *experience* better aligns with the notion of what humans require in order to learn.
**Q&A**

**Q:** Did your definition of error rate include a typo?

**A:** Oops, yes! My mistake.

*Def:* **error rate** is the proportion of **test** examples on which we predicted the wrong label.

With the correct definition, we can now talk about:

1. *Def:* **training error rate** is the error rate on the training data.
2. *Def:* **test error rate** is the error rate on the test data.
Q: What does the technical term “point” refer to?

A: **Def:** a **point** is a collection of **features** (aka. **attributes**)

**Def:** an **example** contains a **label** (aka. **class**) and a point
Q: What is “test time”?  

A: Good question!
Q: Can we have the handwritten notes from lectures?

A: Okay fine...

https://1drv.ms/u/s!Aqk9RupCw3gqixxHH34qLcj5uJTQ?e=E9OYu7

...but just be warned that lots of education research suggests that taking your own notes is the best way to learn!
Reminders

• **Homework 1: Background**
  – Out: Wed, Jan 19 (1st lecture)
  – Due: Wed, Jan 26 at 11:59pm
  – Two parts:
    1. written part to Gradescope
    2. programming part to Gradescope
  – unique policy for this assignment:
    1. **two submissions** for written (see writeup for details)
    2. **unlimited submissions** for programming (i.e. keep submitting until you get 100%)
  – unique policy for this assignment: we will grant (essentially) any and all extension requests

• Please set your name in Gather.Town to be identical to your name in OHQueue.
Big Ideas

1. How to formalize a learning problem
2. How to learn an expert system (i.e. Decision Tree)
3. Importance of inductive bias for generalization
4. Overfitting
FUNCTION APPROXIMATION
Function Approximation

Quiz: Implement a simple function which returns $-\sin(x)$.

A few constraints are imposed:

1. You can’t call any other trigonometric functions
2. You can call an existing implementation of $\sin(x)$ a few times (e.g. 100) to test your solution
3. You only need to evaluate it for $x$ in $[0, 2\pi]$
SUPERVISED MACHINE LEARNING
Medical Diagnosis

• Setting:
  – Doctor must decide whether or not patient is sick
  – Looks at attributes of a patient to make a medical diagnosis
  – (Prescribes treatment if diagnosis is positive)

• Key problem area for Machine Learning

• Potential to reshape health care
Interview Transcript
Date: Jan. 15, 2022
Parties: Matt Gormley and Doctor S.
Topic: Medical decision making

Matt: Welcome. Thanks for interviewing with me today.
Doctor S: Interviewing…?
Matt: Yes. For the record, what type of doctor are you?
Doctor S: Who said I'm a doctor?
Matt: I thought when we set up this interview you said—
Doctor S: I'm a preschooler.
Matt: Good enough. Today, I'd like to learn how you would determine whether or not your little brother is allergic to cats given his symptoms.
Doctor S: He's not allergic.
Matt: We haven't started yet. Now, suppose he is sneezing. Does he have allergies to cats?
Doctor S: No, that's just the sniffles.
Matt: What if he is itchy; Does he have allergies?
Doctor S: No, that's just a mosquito.
[Editor's note: preschoolers unilaterally agree that itchiness is always caused by mosquitos, regardless of whether mosquitos were/are present.]
Matt: What if he's both sneezing and itchy?
Doctor S: Then he's allergic.
Matt: Got it. What if your little brother is sneezing and itchy, plus he's a doctor.
Doctor S: Then he's not allergic.
Matt: How do you know?
Doctor S: Doctors don't get allergies.
Matt: What if he is not sneezing, but is itchy, and he is a fox—
— and the fox is in the bottle where the tweetle beetles battle with their paddles in a puddle on a noodle—
— eating poodle.
Doctor S: Then he is must be a tweetle beetle noodle poodle bottled paddled muddled dudled fuddled wuddled fox in socks, sir. That means he's definitely allergic.
Matt: Got it. Can I use this conversation in my lecture?
Doctor S: Yes.
Interview Transcript

Date: Jan. 15, 2022
Parties: Matt Gormley and Doctor S.
Topic: Medical decision making

Matt: Welcome. Thanks for interviewing with me today.
Dr. S: Interviewing...?
Matt: Yes. For the record, what type of doctor are you?
Dr. S: Who said I'm a doctor?
Matt: I thought when we set up this interview you said—
Dr. S: I'm a preschooler.
Matt: Good enough. Today, I'd like to learn how you would determine whether or not your little brother is allergic to cats given his symptoms.
Dr. S: He's not allergic.
Matt: We haven't started yet. Now, suppose he is sneezing. Does he have allergies to cats?
Dr. S: Well, we don't even have a cat, so that doesn't make any sense.
Matt: What if he is itchy; Does he have allergies?
Dr. S: No, that's just a mosquito.
[Editor’s note: preschoolers unilaterally agree that itchiness is always caused by mosquitos, regardless of whether mosquitos were/are present.]
Matt: What if he's both sneezing and itchy?
Dr. S: Then he’s allergic.
Matt: Got it. What if your little brother is sneezing and itchy, plus he’s a doctor.
Dr. S: Then, thumbs down, he's not allergic.
Matt: How do you know?
Dr. S: Doctors don’t get allergies.
Matt: What if he is not sneezing, but is itchy, and he is a fox....
Matt: ...and the fox is in the bottle where the tweetle beetles battle with their paddles in a puddle on a noodle-eating poodle.
Dr. S: Then he is must be a tweetle beetle noodle poodle bottled paddled muddled dudded fuddled wuddled fox in socks, sir. That means he's definitely allergic.
Matt: Got it. Can I use this conversation in my lecture?
Dr. S: Yes
Doctor diagnoses the patient as sick or not $y \in \{+, -\}$ based on attributes of the patient $x_1, x_1, \ldots, x_M$

<table>
<thead>
<tr>
<th>i</th>
<th>$y$</th>
<th>$x_1$ (allergic?)</th>
<th>$x_2$ (hives?)</th>
<th>$x_3$ (sneezing?)</th>
<th>$x_4$ (red eye?)</th>
<th>has cat?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Medical Diagnosis Dataset

Doctor diagnoses the patient as sick or not $y \in \{+, -\}$ based on attributes of the patient $x_1, x_2, \ldots, x_M$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$y$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>allergic?</td>
<td>hives?</td>
<td>sneezing?</td>
<td>red eye?</td>
<td>has cat?</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>
Medical Diagnosis Dataset

Doctor diagnoses the patient as sick or not $y \in \{+, -\}$ based on attributes of the patient $x_1, x_1, \ldots, x_M$

<table>
<thead>
<tr>
<th>i</th>
<th>y</th>
<th>x_1</th>
<th>x_2</th>
<th>x_3</th>
<th>x_4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>allergic?</td>
<td>hives?</td>
<td>sneezing?</td>
<td>red eye?</td>
<td>has cat?</td>
</tr>
<tr>
<td>1</td>
<td>$y^{(1)}$ -</td>
<td>$x_1^{(1)}$ Y</td>
<td>$x_2^{(1)}$ N</td>
<td>$x_3^{(1)}$ N</td>
<td>$x_4^{(1)}$ N</td>
</tr>
<tr>
<td>2</td>
<td>$y^{(2)}$ -</td>
<td>$x_1^{(2)}$ N</td>
<td>$x_2^{(2)}$ Y</td>
<td>$x_3^{(2)}$ N</td>
<td>$x_4^{(2)}$ N</td>
</tr>
<tr>
<td>3</td>
<td>$y^{(3)}$ +</td>
<td>$x_1^{(3)}$ Y</td>
<td>$x_2^{(3)}$ Y</td>
<td>$x_3^{(3)}$ N</td>
<td>$x_4^{(3)}$ N</td>
</tr>
<tr>
<td>4</td>
<td>$y^{(4)}$ -</td>
<td>$x_1^{(3)}$ Y</td>
<td>$x_2^{(3)}$ N</td>
<td>$x_3^{(3)}$ Y</td>
<td>$x_4^{(3)}$ Y</td>
</tr>
<tr>
<td>5</td>
<td>$y^{(5)}$ +</td>
<td>$x_1^{(4)}$ N</td>
<td>$x_2^{(4)}$ Y</td>
<td>$x_3^{(4)}$ Y</td>
<td>$x_4^{(4)}$ N</td>
</tr>
</tbody>
</table>
# Medical Diagnosis Dataset

Doctor diagnoses the patient as sick or not \( y \in \{+, -\} \) based on attributes of the patient \( x_1, x_1, \ldots, x_M \)

<table>
<thead>
<tr>
<th>( i )</th>
<th>( y )</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</th>
<th>( x_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( y^{(1)} )</td>
<td>( y^{(1)} )</td>
<td>( x_1^{(1)} ) Y</td>
<td>( x_2^{(1)} ) N</td>
<td>( x_3^{(1)} ) N</td>
</tr>
<tr>
<td>2</td>
<td>( y^{(2)} )</td>
<td>( y^{(2)} )</td>
<td>( x_1^{(2)} ) N</td>
<td>( x_2^{(2)} ) Y</td>
<td>( x_3^{(2)} ) N</td>
</tr>
<tr>
<td>3</td>
<td>( y^{(3)} )</td>
<td>( y^{(3)} )</td>
<td>( x_1^{(3)} ) Y</td>
<td>( x_2^{(3)} ) Y</td>
<td>( x_3^{(3)} ) N</td>
</tr>
<tr>
<td>4</td>
<td>( y^{(4)} )</td>
<td>( y^{(4)} )</td>
<td>( x_1^{(4)} ) Y</td>
<td>( x_2^{(4)} ) N</td>
<td>( x_3^{(4)} ) Y</td>
</tr>
<tr>
<td>5</td>
<td>( y^{(5)} )</td>
<td>( y^{(5)} )</td>
<td>( x_1^{(5)} ) N</td>
<td>( x_2^{(5)} ) Y</td>
<td>( x_3^{(5)} ) Y</td>
</tr>
</tbody>
</table>

\( N = 5 \) training examples

\( M = 4 \) attributes
ML as Function Approximation

Chalkboard

– ML as Function Approximation
  • Problem setting
  • Input space
  • Output space
  • Unknown target function
  • Hypothesis space
  • Training examples
  • Goal of Learning
Supervised Machine Learning

$D_{\text{train}}$

$c^*(x)$

Learning Algorithm

$h(x)$
Medical Diagnosis Dataset

Doctor diagnoses the patient as sick or not $y \in \{+, -\}$ based on attributes of the patient $x_1, x_2, \ldots, x_M$

<table>
<thead>
<tr>
<th>i</th>
<th>allergic?</th>
<th>hives?</th>
<th>sneezing?</th>
<th>red eye?</th>
<th>has cat?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$y^{(1)}$</td>
<td>$x_1^{(1)}$</td>
<td>$x_2^{(1)}$</td>
<td>$x_3^{(1)}$</td>
<td>$x_4^{(1)}$</td>
</tr>
<tr>
<td>2</td>
<td>$y^{(2)}$</td>
<td>$x_1^{(2)}$</td>
<td>$x_2^{(2)}$</td>
<td>$x_3^{(2)}$</td>
<td>$x_4^{(2)}$</td>
</tr>
<tr>
<td>3</td>
<td>$y^{(3)}$</td>
<td>$x_1^{(3)}$</td>
<td>$x_2^{(3)}$</td>
<td>$x_3^{(3)}$</td>
<td>$x_4^{(3)}$</td>
</tr>
<tr>
<td>4</td>
<td>$y^{(4)}$</td>
<td>$x_1^{(3)}$</td>
<td>$x_2^{(3)}$</td>
<td>$x_3^{(3)}$</td>
<td>$x_4^{(3)}$</td>
</tr>
<tr>
<td>5</td>
<td>$y^{(5)}$</td>
<td>$x_1^{(4)}$</td>
<td>$x_2^{(4)}$</td>
<td>$x_3^{(4)}$</td>
<td>$x_4^{(4)}$</td>
</tr>
</tbody>
</table>

$N = 5$ training examples  
$M = 4$ attributes

Example hypothesis function:  
$h(x) = \begin{cases} + & \text{if sneezing} = Y \\ - & \text{otherwise} \end{cases}$
Supervised Machine Learning

• **Problem Setting**
  – Set of possible inputs, $x \in \mathcal{X}$ (all possible patients)
  – Set of possible outputs, $y \in \mathcal{Y}$ (all possible diagnoses)
  – Exists an unknown target function, $c^* : \mathcal{X} \rightarrow \mathcal{Y}$ (the doctor’s brain)
  – Set, $\mathcal{H}$, of candidate hypothesis functions, $h : \mathcal{X} \rightarrow \mathcal{Y}$ (all possible decision trees)

• **Learner is given** $N$ training examples $D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(N)}, y^{(N)})\}$ where $y^{(i)} = c^*(x^{(i)})$ (history of patients and their diagnoses)

• **Learner produces** a hypothesis function, $\hat{y} = h(x)$, that best approximates unknown target function $y = c^*(x)$ on the training data
Supervised Machine Learning

• **Problem Setting**
  – Set of possible inputs, \( x \in \mathcal{X} \) (all possible patients)
  – Set of possible outputs, \( y \in \mathcal{Y} \) (all possible diagnoses)
  – Exists an unknown target function, \( c^* : \mathcal{X} \rightarrow \mathcal{Y} \) (the doctor’s brain)
  – Set, \( \mathcal{H} \), of candidate hypothesis functions, \( h : \mathcal{X} \rightarrow \mathcal{Y} \) (all possible decision trees)

• **Learner is given** \( N \) training examples \( D = \{ (x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(N)}, y^{(N)}) \} \)
  where \( y^{(i)} = c^*(x^{(i)}) \) (history of patients and their diagnoses)

• **Learner produces** a hypothesis function, \( \hat{y} = h(x) \), that best approximates unknown target function \( y = c^*(x) \) on the training data

Two important settings we’ll consider:

1. **Classification**: the possible outputs are **discrete**
2. **Regression**: the possible outputs are **real-valued**
Quiz: Implement a simple function which returns \(-\sin(x)\).

A few constraints are imposed:
1. You can’t call any other trigonometric functions
2. You can call an existing implementation of \(\sin(x)\) a few times (e.g. 100) to test your solution
3. You only need to evaluate it for \(x\) in \([0, 2\pi]\)
Supervised Machine Learning

- **Problem Setting**
  - Set of possible inputs, \( x \in \mathcal{X} \) (all values in \([0, 2\pi]\))
  - Set of possible outputs, \( y \in \mathcal{Y} \) (all values in \([-1,1]\))
  - Exists an unknown target function, \( c^* : \mathcal{X} \rightarrow \mathcal{Y} \) \((c^*(x) = \sin(x))\)
  - Set, \( \mathcal{H} \), of candidate hypothesis functions, \( h : \mathcal{X} \rightarrow \mathcal{Y} \) (all possible piecewise linear functions)

- **Learner is given** N training examples \( D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(N)}, y^{(N)})\} \)
  where \( y^{(i)} = c^*(x^{(i)}) \)
  (true values of \( \sin(x) \) for a few random \( x \)'s)

- **Learner produces** a hypothesis function, \( \hat{y} = h(x) \), that best approximates unknown target function \( y = c^*(x) \) on the training data
EVALUATION OF MACHINE LEARNING ALGORITHM
Supervised Machine Learning

$D_{\text{train}}$

$c^*(x)$

$D_{\text{test}}$

$h(x)$

Predictions

Learning Algorithm
Quiz: Implement a simple function which returns $-\sin(x)$.

A few constraints are imposed:

1. You can’t call any other trigonometric functions
2. You can call an existing implementation of $\sin(x)$ a few times (e.g. 100) to test your solution
3. You only need to evaluate it for $x$ in $[0, 2\pi]$
Evaluation of ML Algorithms

Chalkboard

– How to evaluate an ML algorithm?
– Definition: Loss function
  • Example for regression
  • Example for classification
– Definition: Error Rate
– Test dataset
Supervised Machine Learning

$D_{\text{train}}$

$c^*(x)$

$D_{\text{test}}$

$h(x)$

Learning Algorithm

Predictions

Test Error Rate
Error Rate

• Consider a hypothesis $h$ its…
  ... error rate over all training data: $\text{error}(h, D_{\text{train}})$
  ... error rate over all test data: $\text{error}(h, D_{\text{test}})$
  ... true error over all data: $\text{error}_{\text{true}}(h)$

In practice, $\text{error}_{\text{true}}(h)$ is unknown
Majority Vote Classifier Example

Dataset:
Output Y, Attributes A and B

<table>
<thead>
<tr>
<th>Y</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>-</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>+</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>+</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>+</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>+</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>+</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>+</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

In-Class Exercise

What is the **training error** (i.e. error rate on the training data) of the majority vote classifier on this dataset?

Choose one of: 
{0/8, 1/8, 2/8, ..., 8/8}
LEARNING ALGORITHMS FOR SUPERVISED CLASSIFICATION
ML as Function Approximation

**Chalkboard**

– Algorithm 0: Memorizer
– Aside: Does memorization = learning?
– Algorithm 1: Majority Vote
ML as Function Approximation

Chalkboard

– Algorithm 2: Decision Stump
– Algorithm 3 (preview): Decision Tree
Tree to Predict C-Section Risk

Learned from medical records of 1000 women  (Sims et al., 2000)

Negative examples are C-sections

\[ [833+,167-] .83+ .17- \]

Fetal_Presentation = 1: [822+,116-] .88+ .12-
| Previous_Csection = 0: [767+,81-] .90+ .10-
| | Primiparous = 0: [399+,13-] .97+ .03-
| | Primiparous = 1: [368+,68-] .84+ .16-
| | | Fetal_Distress = 0: [334+,47-] .88+ .12-
| | | | Birth_Weight < 3349: [201+,10.6-] .95+ .05-
| | | | Birth_Weight >= 3349: [133+,36.4-] .78+
| | | Fetal_Distress = 1: [34+,21-] .62+ .38-
| | Previous_Csection = 1: [55+,35-] .61+ .39-
Fetal_Presentation = 2: [3+,29-] .11+ .89-
Fetal_Presentation = 3: [8+,22-] .27+ .73-

Figure from Tom Mitchell