Today

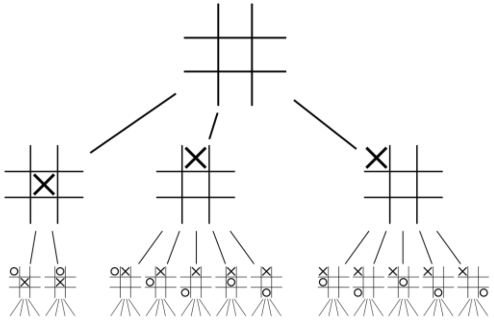
- Recap
- Classification

Announcement: PS1 is now posted,
due by midnight Feb 12

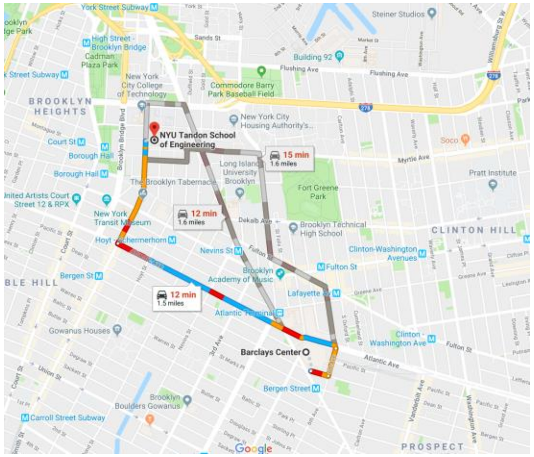
Recap

Artificial Intelligence (AI)

- **Artificial Intelligence**: A branch of computer science that allows computers to make predictions and decisions



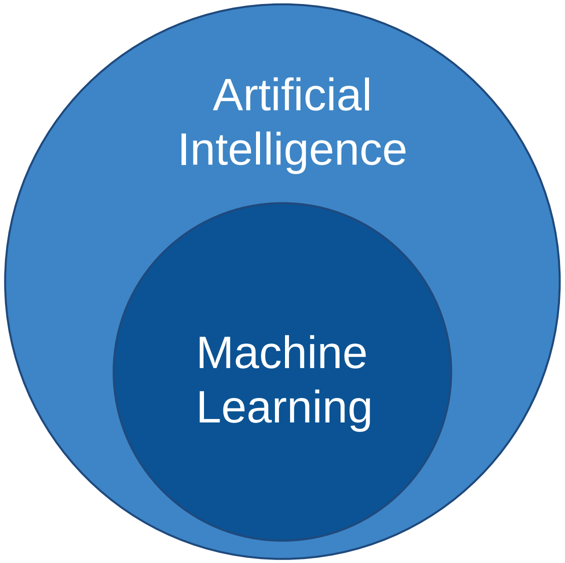
- **Dijkstra’s Shortest Path Algorithm**: A search algorithm that finds the shortest path between two nodes on a weighted graph



Recap

How is Machine Learning (ML) different from AI?

- Machine learning is a type of artificial intelligence
- Machine learning makes decisions based on data it has seen
- Not all AI algorithms need to do this
- Many of the latest AI systems all make use of ML
- For this reason, many people use the terms AI and ML interchangeably



Recap

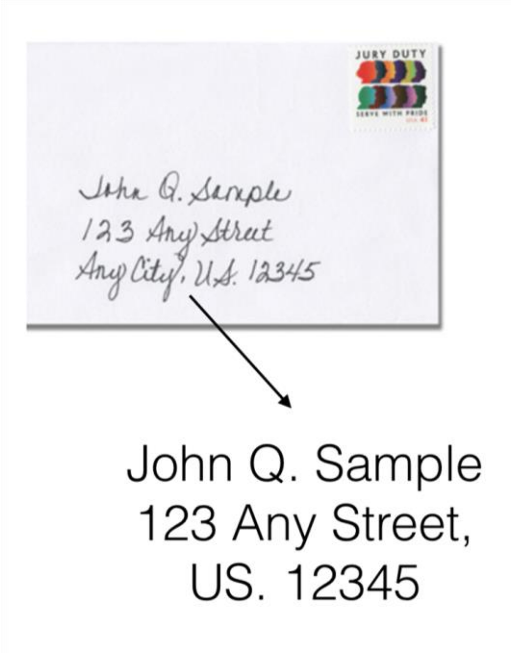
When do we use ML?

- Human expertise doesn’t exist (mars exploration)
- Can’t explain our expertise (speech recognition)
- Large datasets (census)
- Customized models (personalized medicine)

Recap

Why do we use ML?

- Some tasks are too complex for us to fully describe on our own
- Imagine a program to look at handwritten text and figure out what it says
- There are too many different handwriting styles for us to describe *all* variations
- ML gives us a way to tell the computer *how* to learn without needing to describe *what* to learn



Recap

Elements of a ML Algorithm

- **Representation:** how we are gathering and processing our data
  - Linear regression
  - Neural Networks
- **Optimization:** how to best find the patterns that fit our data
  - Backpropagation
  - Rule Learning
- **Evaluation:** how accurate those patterns are
  - Accuracy
  - Squared Error

Recap

So how can we teach machines to learn?

Recap

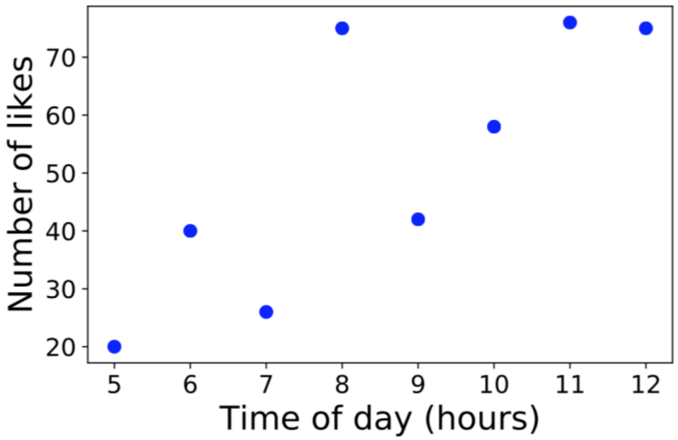
Types of Machine Learning

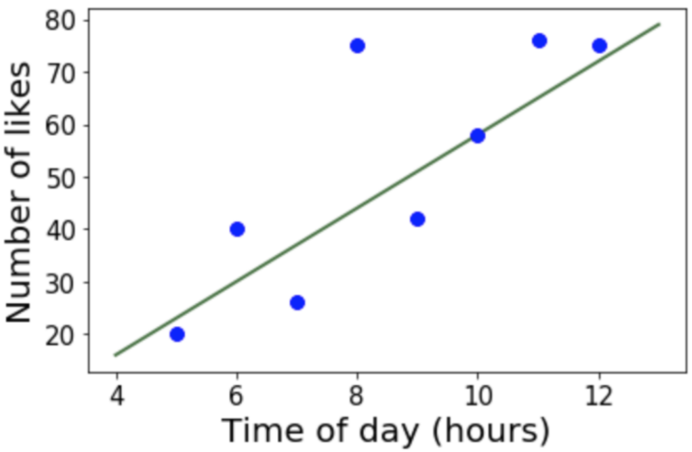
- **Supervised Learning**
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning

Recap

Example of Supervised Learning: Linear Regression

- When the label is a real number
- Training a **model** to find a relationship between input and output values
- Learning a **line of best fit**





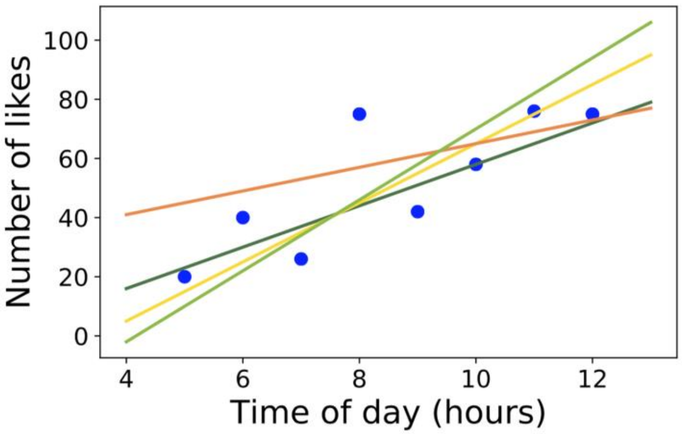
Recap

Linear Regression: Model Parameters

- Learning a best fit line means learning parameters (or weights) for our model.

$$h_\theta(x) = \theta_0 + \theta_1 x$$

θ_i's: Parameters



Recap

Linear Regression: Cost Function

- How do we know that the model parameters result in a “best fit” line? If parameter values minimize our cost function.

Cost Function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

SSD = sum of squared differences, also
SSE = sum of squared errors

Recap

Multivariate Linear Regression

Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

For convenience of notation, define $x_0 = 1$.

θ_i 's: Parameters

Cost Function:

$$J(\theta_0, \theta_1, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Goal: $\underset{\theta_0, \theta_1, \ldots, \theta_n}{\text{minimize}} \; J(\theta_0, \theta_1, \ldots, \theta_n)$ **How??**

Recap

Two potential solutions

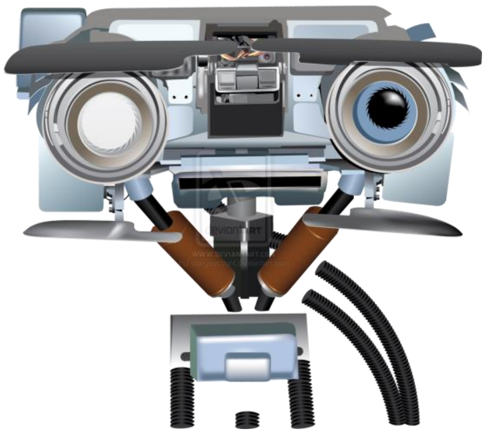
$$\min_{\theta} J(\theta; x^{(1)}, y^{(1)}, \dots, x^{(m)}, y^{(m)})$$

Gradient descent (or other iterative algorithm)

- Start with a guess for θ
- Change θ to decrease $J(\theta)$
- Until reach minimum

Direct minimization

- Take derivative, set to zero
- Sufficient condition for minima
- Not possible for most “interesting” cost functions



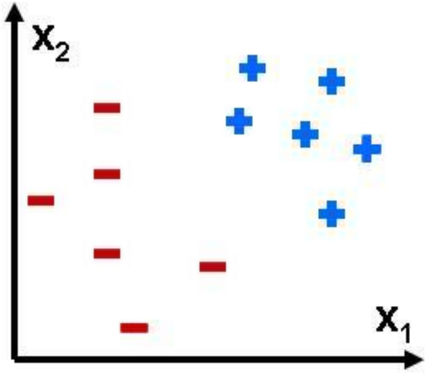
Supervised Learning: Classification

Classification

$y \in \{0,1\}$

0: “Negative Class” (e.g., benign tumor)
1: “Positive Class” (e.g., malignant tumor)

Tumor: Malignant / Benign?
Email: Spam / Not Spam?
Video: Viral / Not Viral?



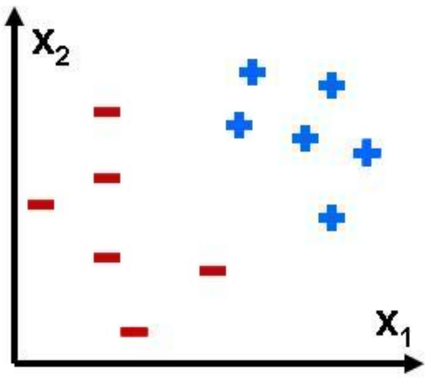
Classification

$y \in \{0,1\}$

0: "Negative Class" (e.g., benign tumor)
1: "Positive Class" (e.g., malignant tumor)

Why not use least squares regression?

$$\underset{\theta}{\operatorname{argmin}} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$



Classification

$y \in \{0,1\}$

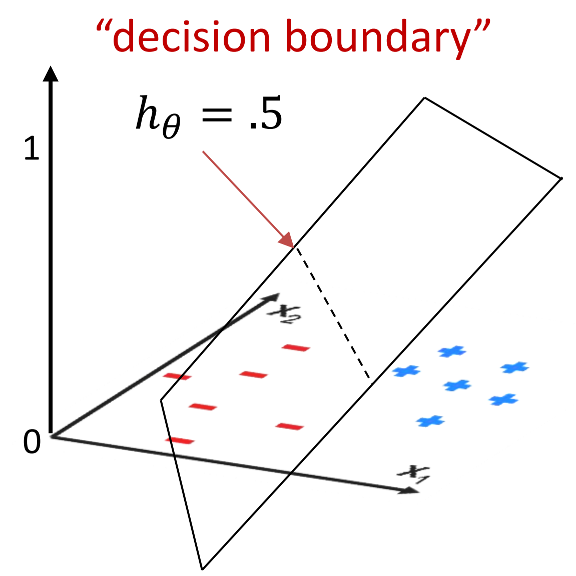
0: “Negative Class” (e.g., benign tumor)

1: “Positive Class” (e.g., malignant tumor)

Why not use least squares regression?

$$\underset{\theta}{\operatorname{argmin}} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- Indeed, this is possible!
  - Predict 1 if $h_\theta(x) > .5$, 0 otherwise
- However, outliers lead to problems…
- Instead, use logistic regression



Least Squares vs. Logistic Regression for Classification

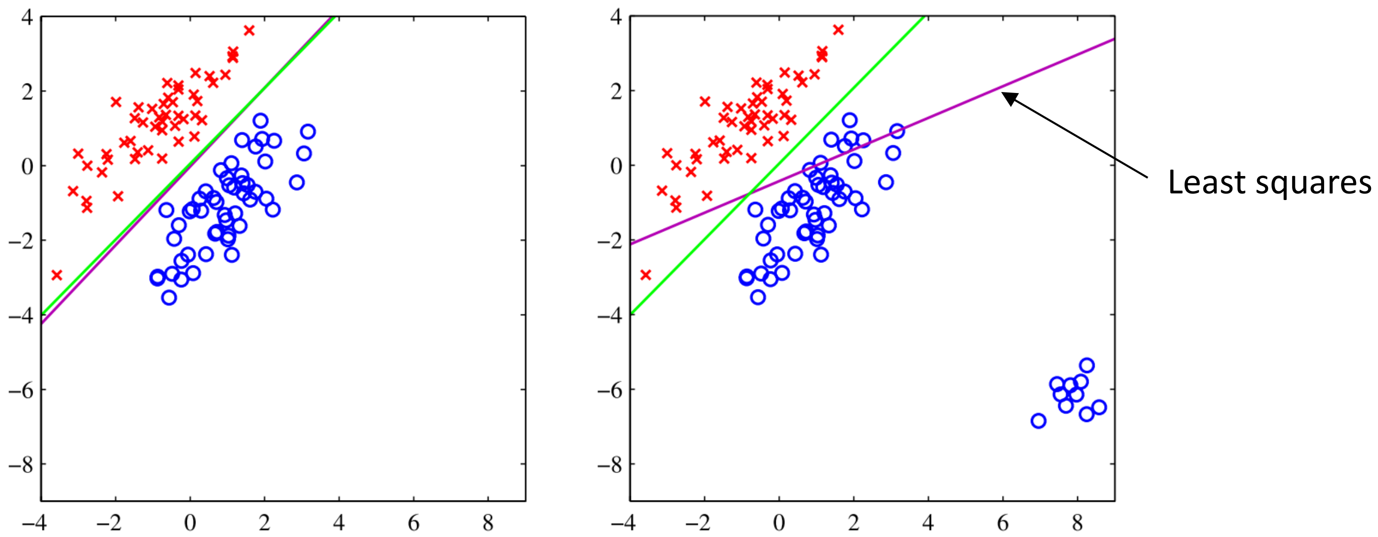


Figure 4.4 from Bishop. The left plot shows data from two classes, denoted by red crosses and blue circles, together with the decision boundary found by least squares (magenta) and also by the logistic regression model (green). The right-hand plot shows the corresponding results obtained when extra data points are added at the bottom left of the diagram, showing that least squares is highly sensitive to outliers, unlike logistic regression.

(see Bishop 4.1.3 for more details)

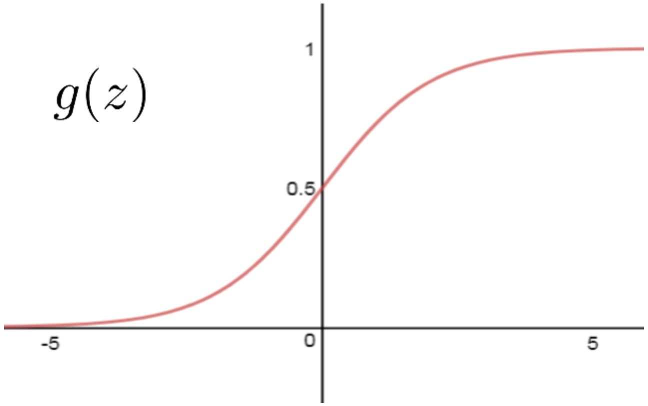
Logistic Regression

$$0 \leq h_\theta(x) \leq 1$$

map to (0, 1) with "sigmoid" function

$$g(z) = \frac{1}{1 + e^{-z}}$$

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$



$$h_\theta(x) = p(y = 1|x)$$ "probability of class 1 given input"

Logistic Regression

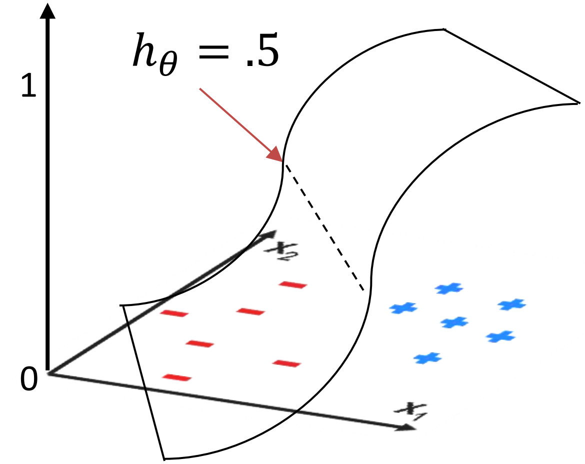
Hypothesis:

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

Predict "$y = 1$" if $h_\theta(x) \geq 0.5$

Predict "$y = 0$" if $h_\theta(x) < 0.5$

"decision boundary"



Logistic Regression Cost Function

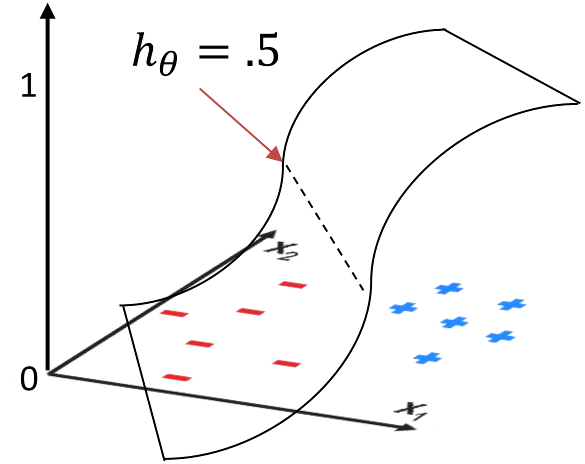
Hypothesis:

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

θ: parameters

$D = \left(x^{(i)}, y^{(i)}\right)$: data

“decision boundary”



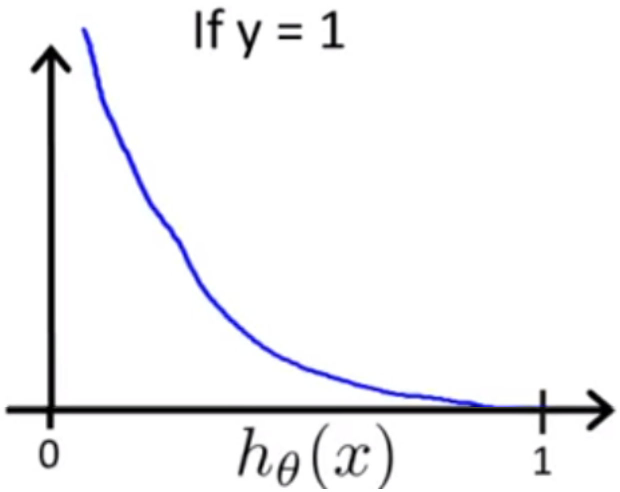
Cost Function: cross entropy

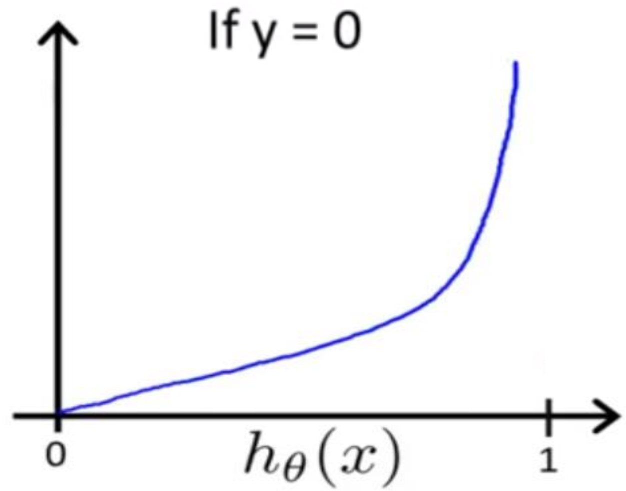
$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

Goal: minimize cost $\quad \min_\theta J(\theta)$

Logistic Regression Cost Function

$$\text{Cost}\left(h_\theta(x^{(i)}), y^{(i)}\right) = \begin{cases} -\log\left(h_\theta\left(x^{(i)}\right)\right) & if\ y = 1 \\ -log\left(1 - h_\theta\left(x^{(i)}\right)\right) & if\ y = 0 \end{cases}$$





$\text{Cost} = 0\ if\ y^{(i)} = 1,\ h_\theta\left(x^{(i)}\right) = 1$

Similarly desirable behavior

But as $h_\theta\left(x^{(i)}\right) \to 0$

$\text{Cost} \to \infty$

[Andrew Ng's Illustration]

Cross Entropy Cost

$$\text{Cost}\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(h_\theta\left(x^{(i)}\right)\right) & if\ y = 1 \\ -log\left(1 - h_\theta\left(x^{(i)}\right)\right) & if\ y = 0 \end{cases}$$

Can be written more compactly as:

$$\text{Cost}\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = -y^{(i)} \log h_\theta(x^{(i)}) - (1 - y^{(i)}) \log\left(1 - h_\theta(x^{(i)})\right)$$

Cross Entropy Cost

- Cross entropy compares distribution *q* to reference *p*

$$H(p,q) = -\sum_x p(x)\,\log q(x).$$

- Here q is predicted probability of $y = 1$ given x, reference distribution is $p = y^{(i)}$, which is either *1 or 0*

$$-\frac{1}{m}[\sum_{i=1}^{m} y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$$

Gradient of Cross Entropy Cost

- Cross entropy cost

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$= -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))]$$

- No direct closed-form solution

- its gradient w.r.t θ is:

$$(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

Gradient descent for Logistic Regression

Cost

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log (1-h_\theta(x^{(i)}))]$$

Want $\min_\theta J(\theta)$:

Repeat $\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

(simultaneously update all θ_j)

$\}$

Gradient descent for Logistic Regression

Cost

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))]$$

Want $\min_\theta J(\theta)$:

Repeat $\{$

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

(simultaneously update all θ_j)

$\}$

Boston University - Spring 2020

CS 542: Machine Learning

+ Add Syllabus

Course Information | Staff | Resources

Edit Resource Sections

Problem Sets

Manually sort using   Sort on: - -

Problem Sets	Date	Actions
Problem Set 1	Feb 12, 2020	Edit / Post a note / Update File / Delete

Add Links   Add Files

Submission Form

CS.542 Spring 2020 Problem Set 1

Important: This homework tests basic mathematical skills required for the course. Note that the order/points of questions do not always imply difficulty.

Submission: Please complete the homework and submit a PDF file (the only accepted file format). The *submission link* can be found here. Typed or scanned handwritten solutions are fine if they are easily readable. ***The deadline for this problem set is: Feb 12, 2020.***

Late Assignment Policy: Late problem sets will be levied a late penalty of 0.5% per hour (up to 72 hours). After 72 hours, no credit will be given.

Grading: The homework assignments in this course are self-graded. Once the solutions are available, please use them to assign points to your own **submitted** solution and submit the points. The *grading link* can be found here. A subset of the assignments will be chosen at random and double checked by the course staff. It is in your best interest to complete the assignments as they will prepare you for the exams. ***The deadline for submitting a grade for this problem set is: Feb 19, 2020.***

Total: 50 points.

CS 542 Problem Set 1: Submission Form

Please complete the homework and submit a PDF file (the only accepted file format). Typed or scanned handwritten solutions are fine if they are easily readable. It is in your best interest to complete the assignments individually as they will prepare you for the exams. The deadline for this problem set is: Feb 12, 2020.

NOTICE: You will not be able to access your PDF after submission, but you can upload a new file. Only the last uploaded file and its timestamp will be counted.

The name, username and photo associated with your Google account will be recorded when you upload files and submit this form. Not **sbargal@bu.edu**? Switch account

* Required

Submit Your Answers Here *

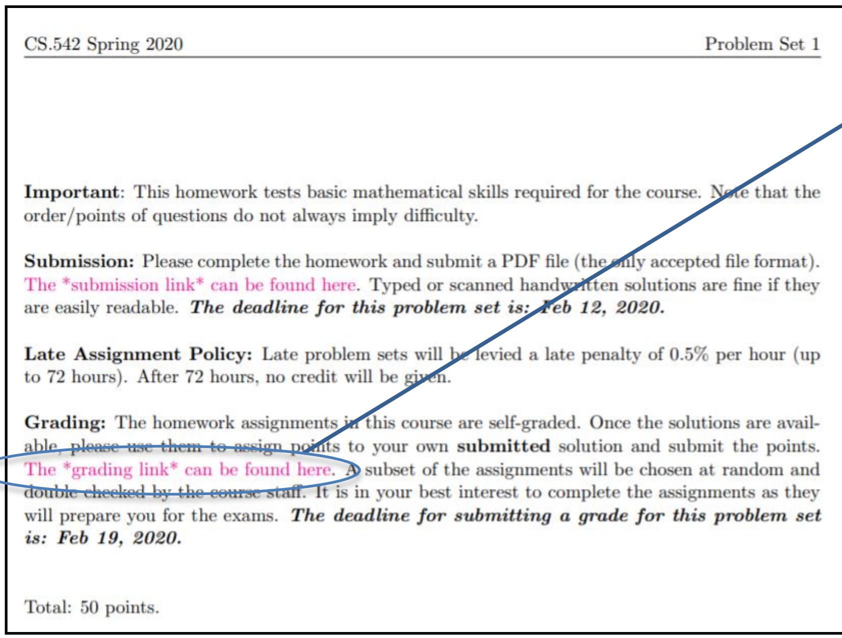
Submit a PDF file **named as** [BUusername]_[first]_[last]_ps1.pdf. Your filename should look like sbargal_sarah_bargal_ps1.pdf

Add file

A copy of your response will be emailed to sbargal@bu.edu.

Submit

Grading Form



CS.542 Spring 2020 Problem Set 1

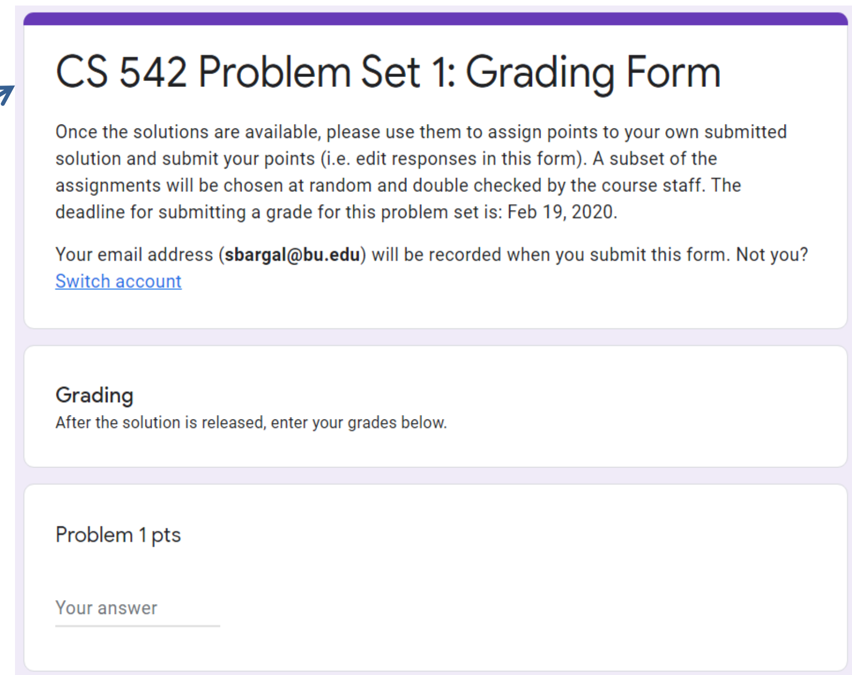
Important: This homework tests basic mathematical skills required for the course. Note that the order/points of questions do not always imply difficulty.

Submission: Please complete the homework and submit a PDF file (the only accepted file format). The *submission link* can be found here. Typed or scanned handwritten solutions are fine if they are easily readable. ***The deadline for this problem set is: Feb 12, 2020.***

Late Assignment Policy: Late problem sets will be levied a late penalty of 0.5% per hour (up to 72 hours). After 72 hours, no credit will be given.

Grading: The homework assignments in this course are self-graded. Once the solutions are available, please use them to assign points to your own **submitted** solution and submit the points. The *grading link* can be found here. A subset of the assignments will be chosen at random and double checked by the course staff. It is in your best interest to complete the assignments as they will prepare you for the exams. ***The deadline for submitting a grade for this problem set is: Feb 19, 2020.***

Total: 50 points.



CS 542 Problem Set 1: Grading Form

Once the solutions are available, please use them to assign points to your own submitted solution and submit your points (i.e. edit responses in this form). A subset of the assignments will be chosen at random and double checked by the course staff. The deadline for submitting a grade for this problem set is: Feb 19, 2020.

Your email address (**sbargal@bu.edu**) will be recorded when you submit this form. Not you? Switch account

Grading

After the solution is released, enter your grades below.

Problem 1 pts

Your answer

⋮

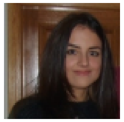
Boston University - Spring 2020

CS 542: Machine Learning

+ Add Syllabus

Course Information | Staff | Resources

Edit

	Name	Office Hours
	Sarah Adel Bargal	**When?** Mon 12-1:15, Fri 1:45-3 **Where?** MCS 207
	Isidora Chara Tourni	**When?** Wed 10:15-11:30, Thu 3:30-4:45 **Where?** EMA 302
	Hieu Le	**When?** Mon 1:15-2:30, Tue 2:15-3:30 **Where?** EMA 302

Reminder: Please let us know ASAP if

- You cannot make it to *any* of the posted office hours.
- You are not yet added to the Piazza course.
- There is a lecture you cannot attend -> for MT scheduling.