Applied Machine Learning!!!

W207 Section 9
Rasika Bhalerao
rasikabh@berkeley.edu

Aug 23: Welcome!
Nov 8 and 22: No classes

Schedule

Supervised learning methods

	Sync	Topic
2	Aug 30	Linear Regression / Gradient Descent
3	Sep 6	Feature Engineering
4	Sep 13	Logistic Regression
5	Sep 20	Multiclass classification / Eval Metrics
6	Sep 27	Neural Networks
7	Oct 4	KNN, Decision Trees, Ensembles

Unsupervised learning methods

	Sync	Topic
8	Oct 11	KMeans and PCA
9	Oct 18	Text Embeddings
10	Oct 25	CNNs
11	Nov 1	EDA, Real data, Baselines
12	Nov 15	Fairness / Ethics
13	Nov 29	Fancy Neural Networks
14	Dec 6	Final Presentations

Assignment Schedule

Due Date	Assignment
Aug 28	HW1
Sep 4	HW2
Sep 11	HW3
Sep 18	HW4
Sep 25	HW5
Oct 2	HW6
Oct 16	Group project baseline
Oct 23	HW8
Nov 6	HW9
Nov 20	HW10
Dec 4	Final project notebook + presentation

Don’t forget to sign up for final project groups by week 4!

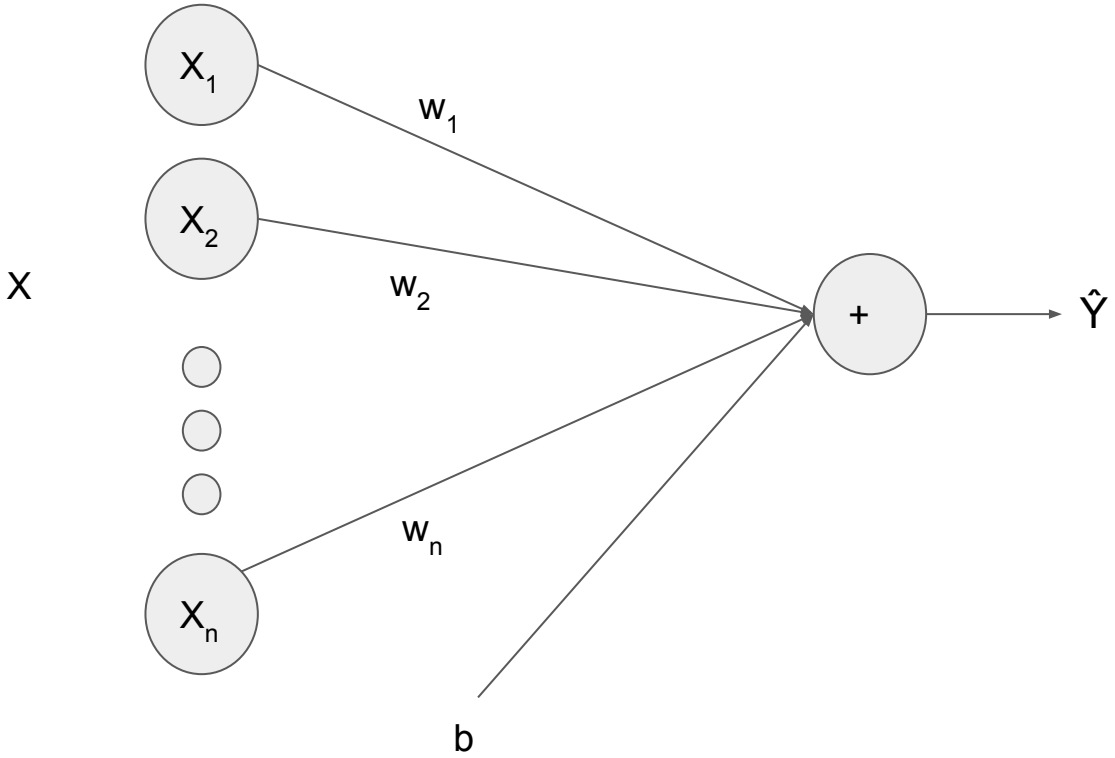
https://docs.google.com/document/d/1R3J_X1Rz6WP8eMQ2cyMC0wAr5iQdhMK_httdoNO6L0w/edit?usp=sharing

Behavior expectations

- Healthy disagreement is expected
- Be mindful of one another’s schedules
- Be a good listener
- Have fun in a professional manner
- Share related real-world experience
- Ask questions when something is confusing
- Keep it 100 but be respectful
- Be open-minded to new ideas in the real world and when coding
- On time for group meetings

What do we use linear regression for? What do we use gradient descent for?

They both look like this:



Which one can we solve for directly? Which one needs an iterative process? Why?

View options (at the top) → Annotate

Async Practice Quiz Questions (vote!)

Modern machine learning typically uses:	Logic	Numerical optimization
The input to the loss function is a mapping of parameters to values.	True	False
The loss function for any training data is minimized when its value is 0.	True	False
Gradient descent will always converge.	True	False
The gradient is computed separately for each parameter.	True	False

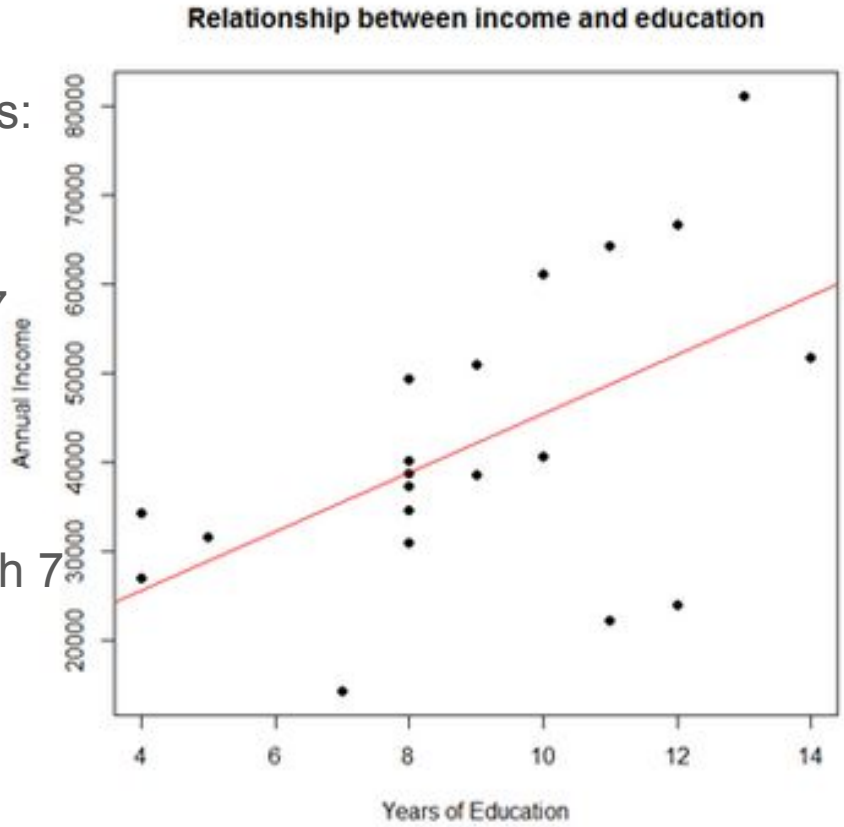
Based on the data and regression line, what is:

- the actual income for the individual with 7 years of education?

$5000

- the predicted income for an individual with 7 years of education?

$35000



What is the total cost for these three data points:

[(1,1), (2,2), (3,3)]

given the model $Y_i = \beta X_i$ and squared error cost function for:

$\beta = 0$: $((1-0)^2 + (2-0)^2 + (3-0)^2)^{½}$

$\beta = 0.5$: $((1-0.5)^2 + (2-1)^2 + (3-1.5)^2)^{½}$

$\beta = 1$: $((1-1)^2 + (2-2)^2 + (3-3)^2)^{½}$

Four key components for any gradient descent

- Model
  - Area = $x(5-x) = 5x-x^2$
- Parameters
  - x
- Cost function
  - $100 - (5x - x^2)$
- Objective
  - Minimize cost function

Example: what is the largest rectangular area we can enclose using a string of length 10?

In this example, our “training data” is just one point, (10, 100).

What are local minima? How can we avoid them?

When do we stop iterating?

When do we stop iterating?

Experts: when validation error isn’t improving

Hyperparameters

- Learning rate
- Batch size
- Epochs

What are they? How do we pick values for them?

What are the differences between using gradient descent for (nonlinear) **regression** and using it for **classification**?

Notebook!

To access later:
https://github.com/MIDS-W207/rasikabh/blob/main/live_sessions/Week2.ipynb