Linear Regression!

Class 4: AISC 1003

Course Logistics

Good question!!

Course Plan*

Week	Date	Topic	Activity
1	May 13, 2023	Intro to AI, ML, DS	Group Exercise (Risk and Applications)
2	May 20, 2023	Types of ML	Practice Exercise,
3	May 27, 2023	Linear Regression - OLS	Group Activity (Excel OLS)
4	Jun 3, 2023	Linear Regression - Gradient Descent	Indiv Assignment 1
5	Jun 10, 2023	Regularization	Group Exercise (Model building)
6	Jun 17, 2023	Cross Validation and Goodness of Fit	Group Assignment 2
7	Jun 24, 2023	Mid-Term	Mid-Term

*Course Plan is tentative and subject to change. All assessments are in-class assessments. No extensions will be provided.

Course Plan*

Week	Date	Topic	Activity
9	Jul 8, 2023	Classification, Sigmoid Function	Group Exercise (GoF)
10	Jul 15, 2023	Logistic Regression	Group Assignment 3
11	Jul 22, 2023	SVM	Indiv Assignment 4, Group Exercise (SVM)
12	Jul 29, 2023	SVM Kernels	Group Assignment 5 (SVM)
13	Aug 5, 2023	Sensitivity Analysis, Model Interpretability	Group Assignment 6 (Sensitivity Analysis)
14	Aug 12, 2023	Final	Final
15	Aug 19, 2023	~~Decision Trees, Ensembling~~	Individual Exercise

Instructor: Bhavik Gandhi

*Course Plan is tentative and subject to change. All assessments are in-class assessments. No extensions will be provided.

Grading & Class Structure

Grades will be absolute

Any integrity violations will get you 0 on the assessment

Bonus points can help you improve your grade

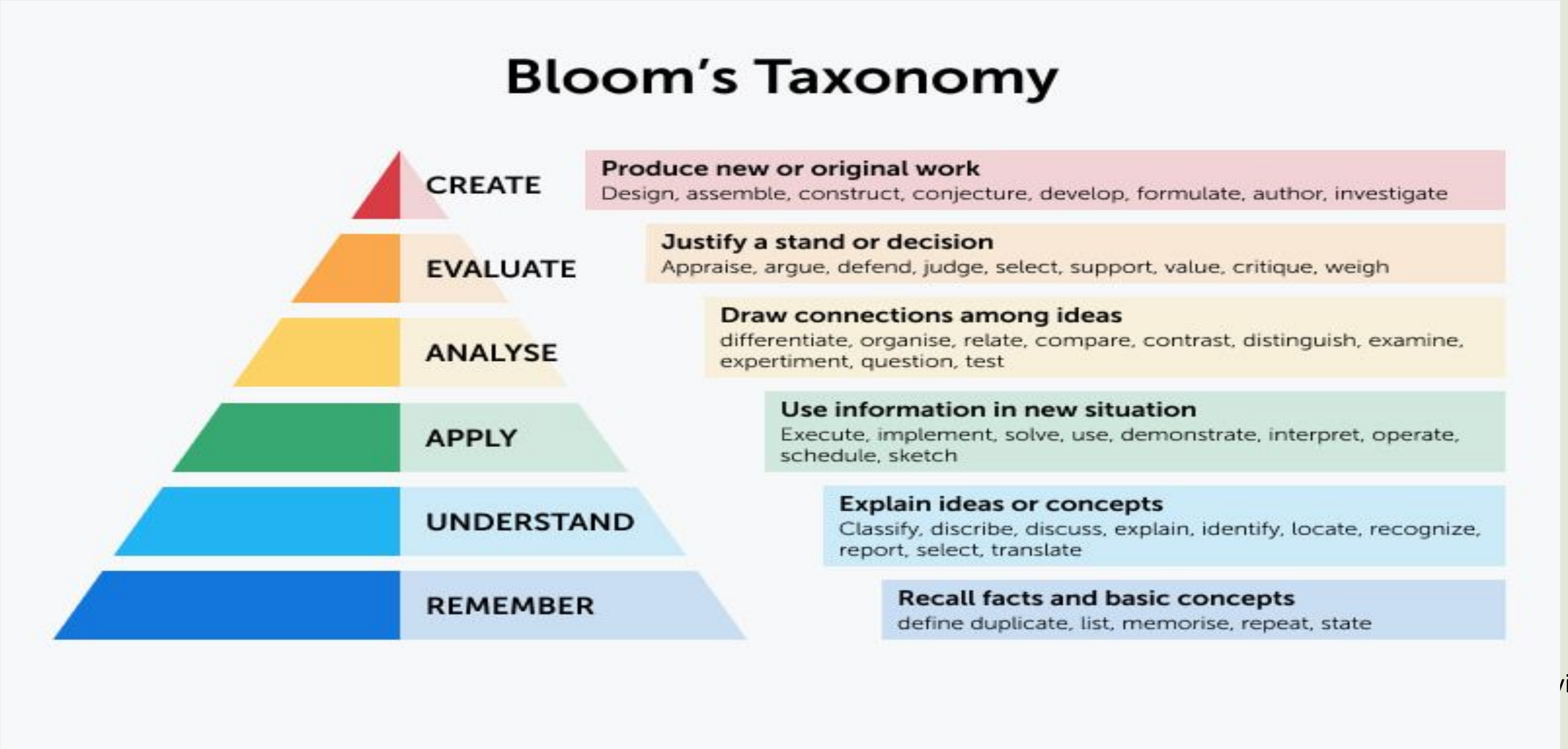
But you will still have to work! 😅

All submissions should be accompanied by an accurate task status report detailing which (and how well) the tasks/subtasks were completed and by who (in case of a group assignment) which can serve as a MECE table.

The class will usually be 1 - 1.5 hrs on instruction

The remaining 2-2.5 hrs will be on application and analysis of the material and in-class activities

Bloom’s Taxonomy



Slides, Learning Resources etc

Slides used here are only for reference, further asynchronous slides will be posted. However assessments will only be on the taught material (understanding, application, analysis, evaluation etc. - check Bloom's taxonomy)

Various concepts will be explained using whiteboard, so feel free to take notes

Additional resources will be posted in Moodle.

Rubrics may be used to evaluate assignments. The assignments will not have bonus pts but upto 10% bonus pts may be awarded for extraordinary work as per the instructor's discretion.

Class Decorum

No talking in the class even to yourself even if it is related to the subject matter except during breaks and group assignments/exercises.

Please step out for at least 15 minutes if you need to talk. If you are found talking you will be asked to step out and instruction to the class will be paused until you step out.

No language other than English in class even during breaks or group work. You need to get English and AI-ready. If found talking in another language you will be asked to leave the class until the next break and instruction to the class will be paused until you do.

No mobile phone ringing/vibrating in class. No recording in the class. If either of these happen you will be asked to leave the class until the next break and instruction will be paused until you do.

You can join in late in class, just come in and close the door. However attendance may be missed, you will miss on some content/instruction and you will not be allowed to join in late if an assessment has begun. Also, you will not get any make-up opportunities so be on time.

Email Writing

It is important to learn how to write good professional emails. Below are some useful instructions to bear in mind.

Please make sure any emails you write are to the point and clearly highlight what you expect from the person addressed.

Please make sure you do the groundwork and provide all the necessary details as opposed to expecting the person addressed to do the work.

Please make sure the subject of the email clearly identifies the objective of the email.

If it is an fyi email, please mention so right away. If you expect a certain action, please mention briefly and exactly what action do you expect and why is it objectively justified.

Academic Integrity

Academic Integrity is extremely important, honesty is highly valued in Canada

- Cannot accept any submissions over email
- If you face any issues submitting the assignment, follow the below steps
  - Take screenshots
  - Send an email to bhavik.gandhi@tbcollege.com with your complete assignment attached
  - Email itsupport@tbcollege.com or moodlesupport@tbcollege.com with these screenshots and get it resolved asap
  - Late submissions will incur a penalty so budget your time accordingly, genuine issues as verified by the it team or moodle team will be granted penalty-free extension
- Read the instructions and do not indulge in plagiarism. Anti-plagiarism softwares are very smart
- Please do not make me file ADRs 🙏 The risk is not worth the few extra pts you may get
- Not following instructions during test, even if you are not explicitly cheating, will still result in an ADR

Bonus Points!

- Total Bonus Points for the course will be capped at 20 (20%), total points capped at 100
- Attendance, Participation/Contribution: Upto 6 bonus points at the instructor’s discretion (includes participation in class, exercises, forums etc.). Will call it out in forums, exercise feedback or by email etc.
- Slide Corrections/Suggestions: Upto 4 bonus points at the instructor’s discretion, kindly email suggestions/corrections and save the response emails.
- Upto 10 bonus points distributed across assignments, exams and project.
- Bonus Point Assignment: Within 2 days of the last/W15 class each student will be required to submit how many bonus points they have and how they’d like their bonus points to be assigned. Failure to do so may lead to forfeiture of your bonus points. Set reminders now!

Group Exercises and Assignments!

- Points will be a combination of group performance and individual performance
- Individual responsibilities should be clearly outlined by the group in a MECE format in the assignment/exercise report
- The best submitted solutions at the instructor's discretion may be shared with the class. You may get some bonus pts for knowledge sharing is your submission in a graded assessment is selected for sharing. Please write to me if you have any concerns regarding this.

Success Factors

No one can teach you all of Machine Learning

There's just too much to teach, so the idea of this course is to give you a primer into these fields so that you can explore them correctly.

To succeed in this course, apart from following instructions and having a strong sense of integrity, you will need

1. Basic Mathematical Skills
2. Problem Solving
3. Time Management
4. Group Working Skills
5. Practice

Today's Outline!

- You will learn a new ML technique called Gradient Descent,
- Understanding the derivation of where it comes from, and
- How it compares to OLSR
- You should appreciate different types of data
- You will do a worked example of LR using GD

Attendance

I am present

Assignment 1

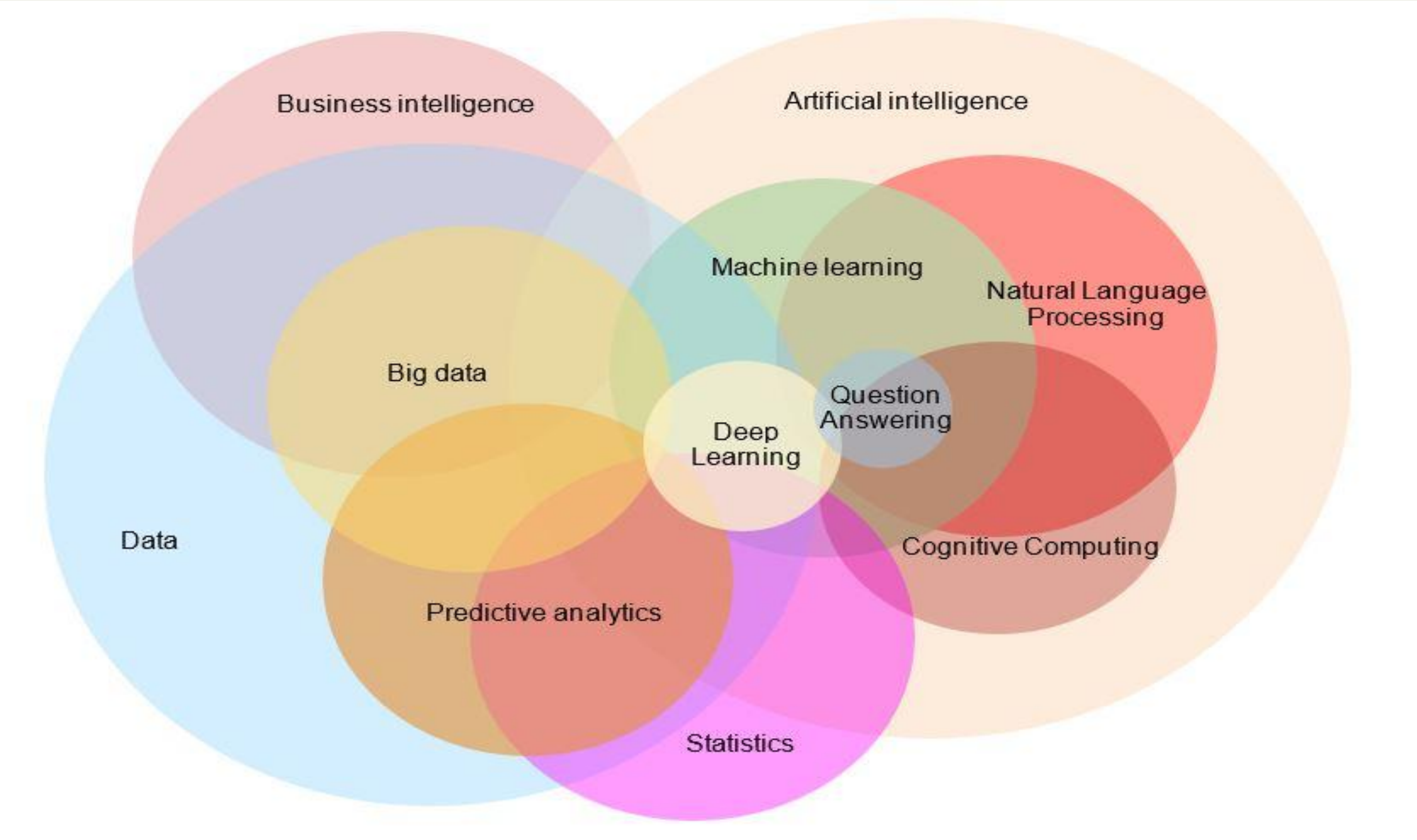
Yes let’s do it

Review

What did we talk about last time?

Intelligence, Learning & Challenges

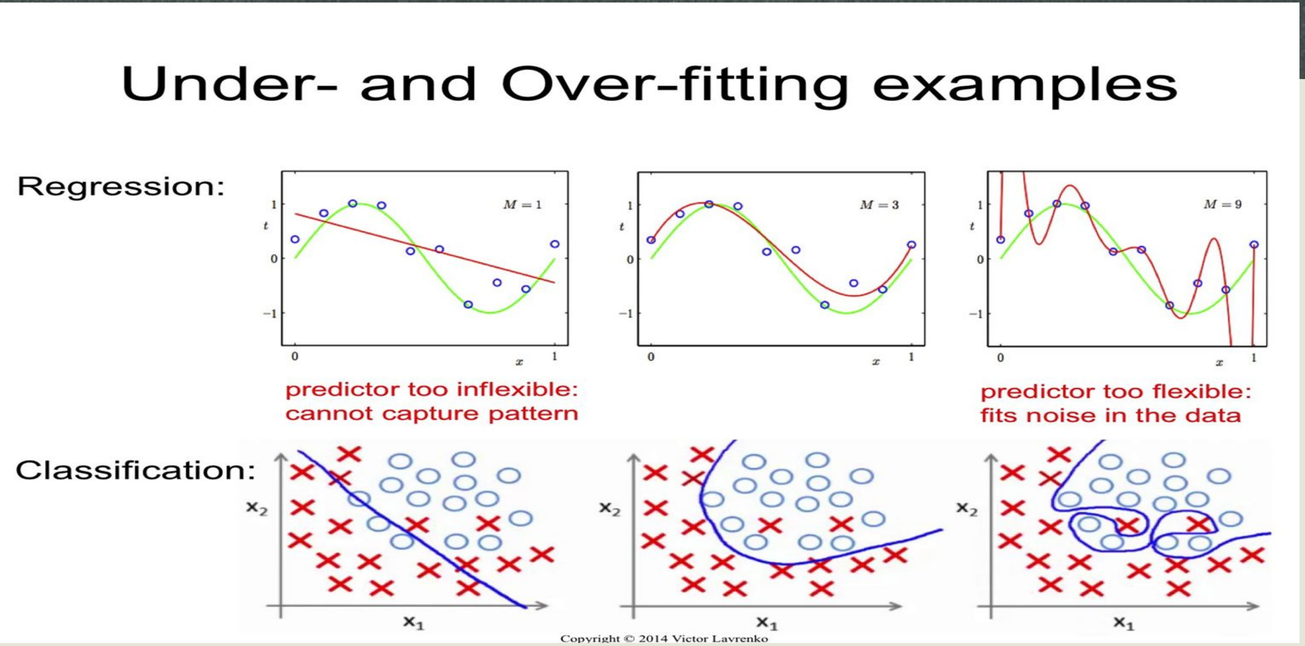
- Intelligence
- Machine Learning
- Human Learning
- AI & Data
- Challenges & Complexities
- AI Cycle
- Applications

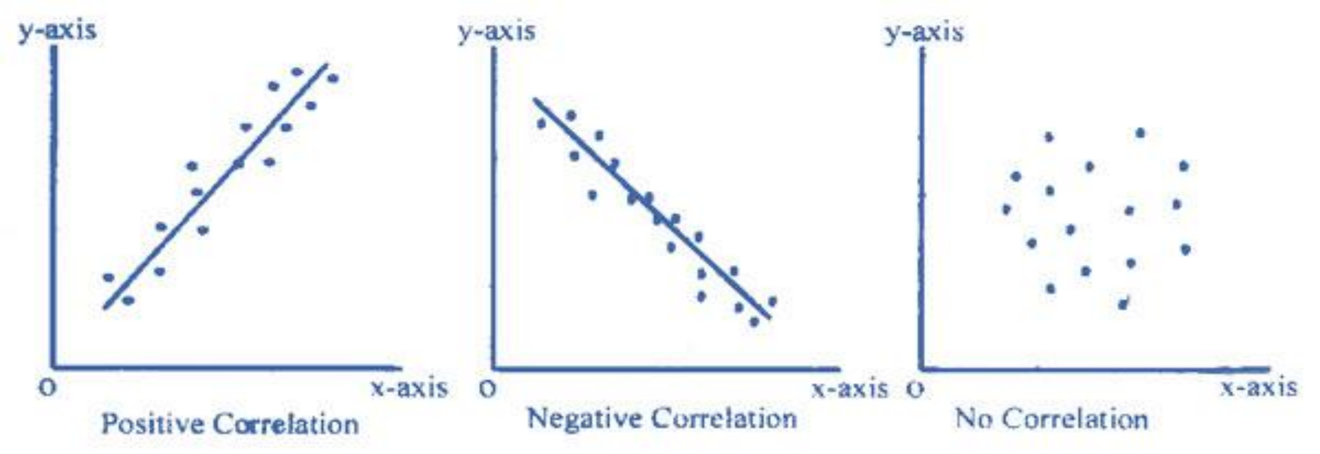


Instructor: Bhavik Gandhi

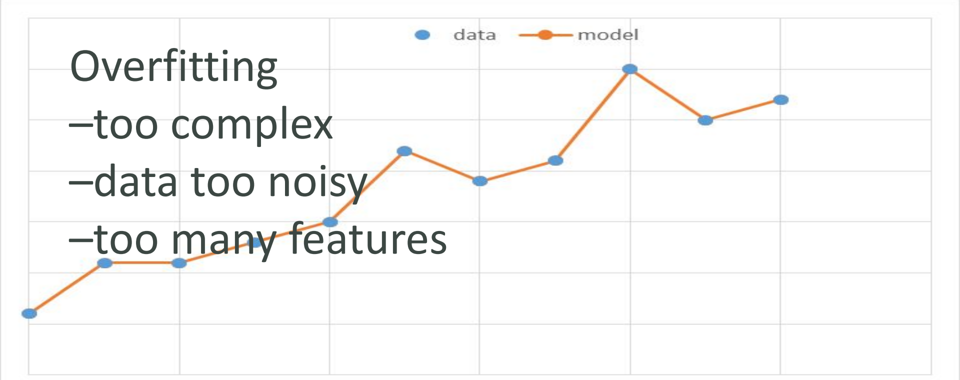
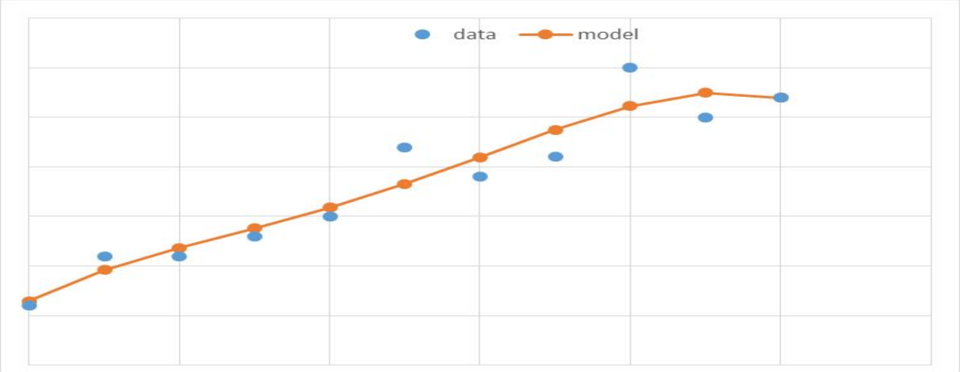
Supervised Learning & Linear Regression

- Types of Machine Learning
- Bias and Variance
- Cost Function & Cross Validation
- Regression vs Classification
- Linear Regression
- Correlation & R2
- Ordinary Least Squares
- Assumptions & Scaling





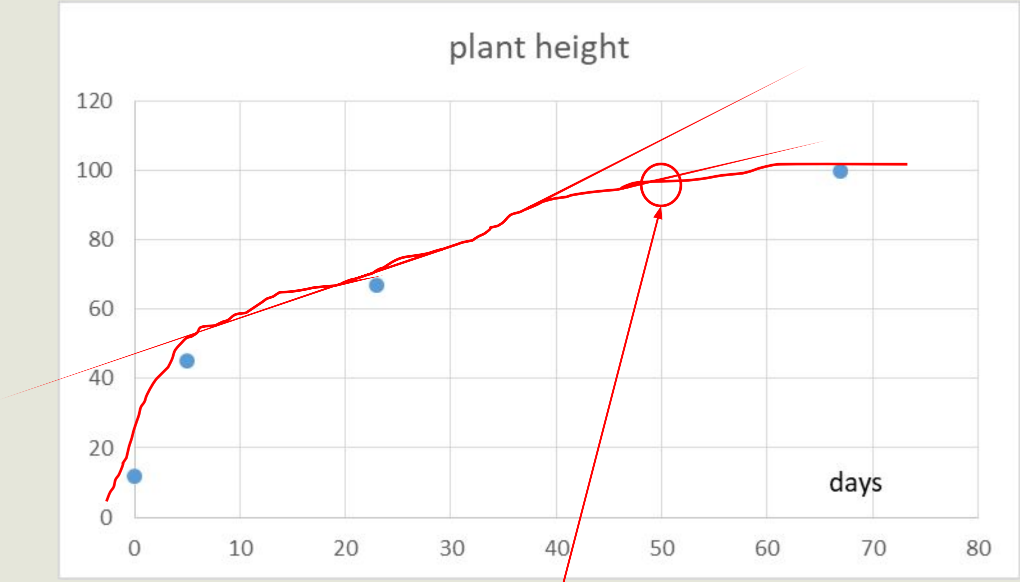
Bias vs Variance

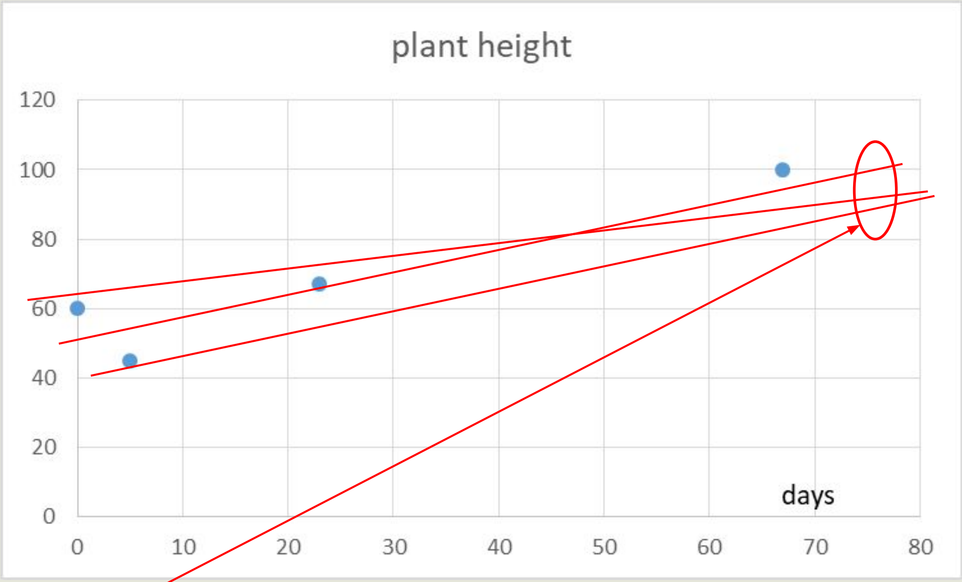
DATA FITTING		BIAS (shift)	
		High -doesn’t respect data, oversimplifies model	Low
VARIANCE (spread)	High –too much emphasis on data, poor prediction		Overfitting –too complex –data too noisy –too many features 
	Low	Underfitting –didn’t learn –not enough data –under sophisticated model 	

Cost Function

Cost of what?

Linear Regression

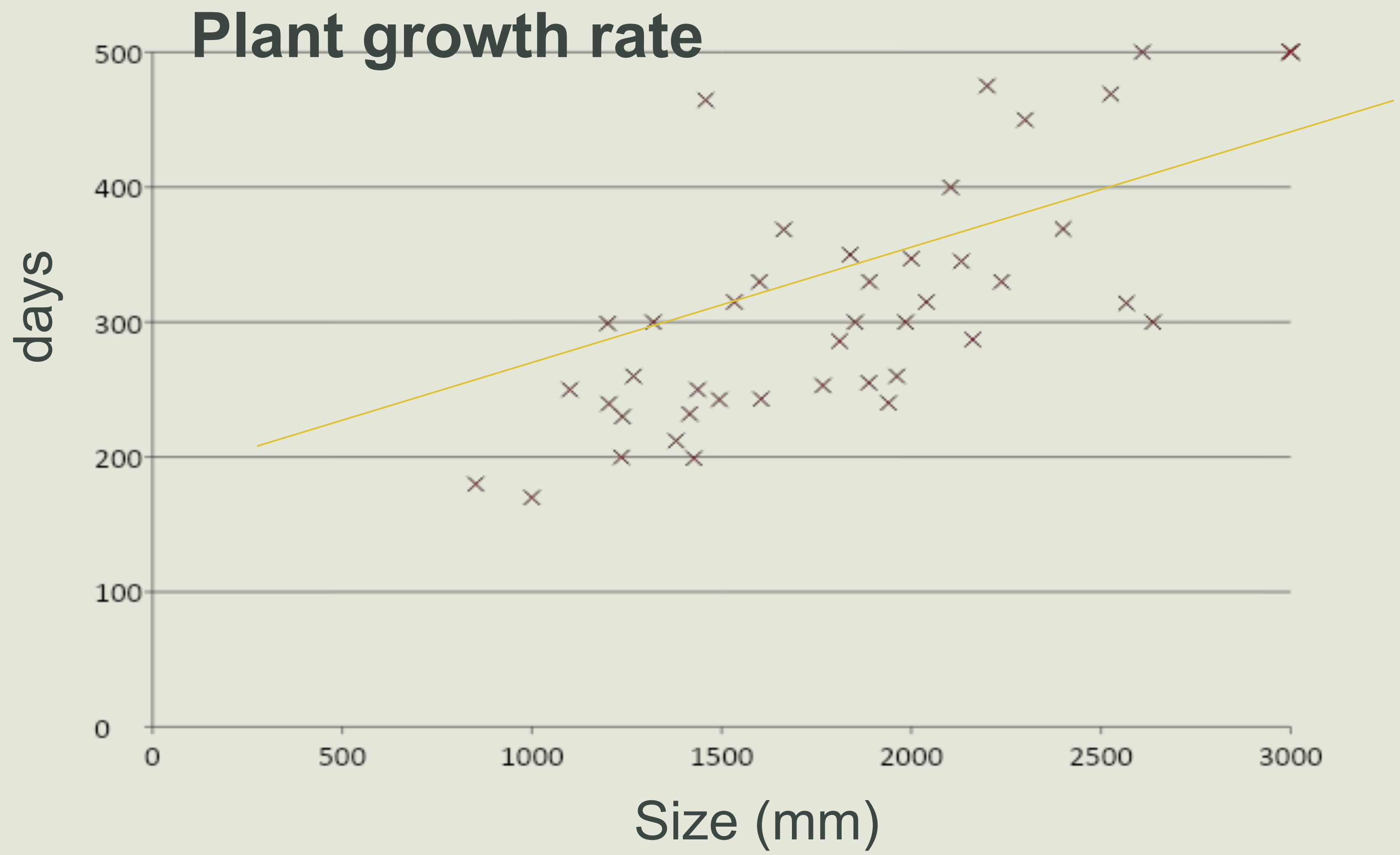




- How do we INTERPOLATE?
- How do we EXTRAPOLATE?

Instructor: Bhavik Gandhi

Linear Regression



Given this dataset, we hypothesize there is a linear relationship.

Y = m**X** + b, Y,X are **arrays**

Y = $(y_1, y_2, \ldots, y_n)$

X = $(x_1, x_2, \ldots, x_n)$

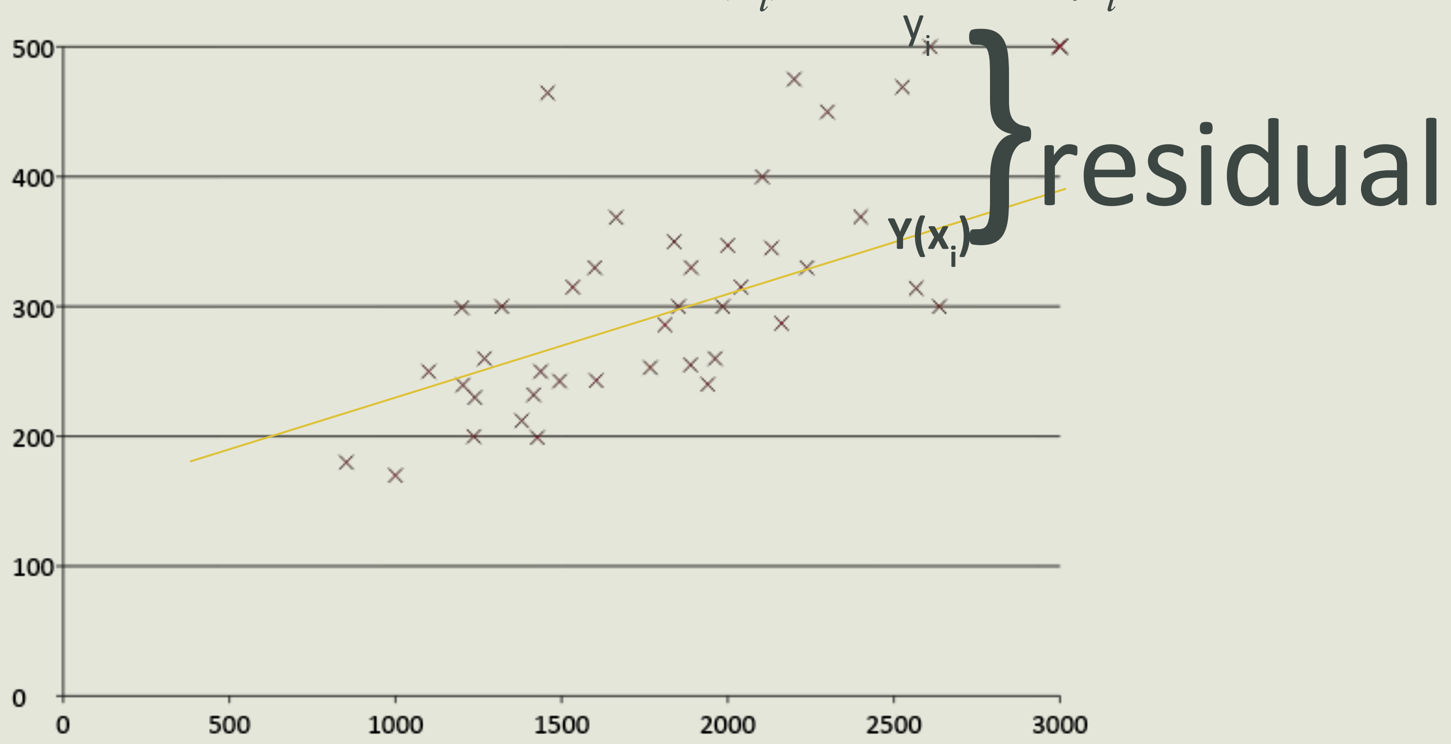
$Y \sim Y(x_i)$

$Y(x_i) = mX + b$

Design a cost function

$Y(x_i) = mX + b$

Choose m, b so that $Y(x_i)$ is close to y_i for each data point.



Cost Function:
$J(m,b) = (1/n)\ \Sigma\ (Y(x_i)-y_i)^2$

Goal:
Minimize J(m,b)

Instructor: Bhavik Gandhi

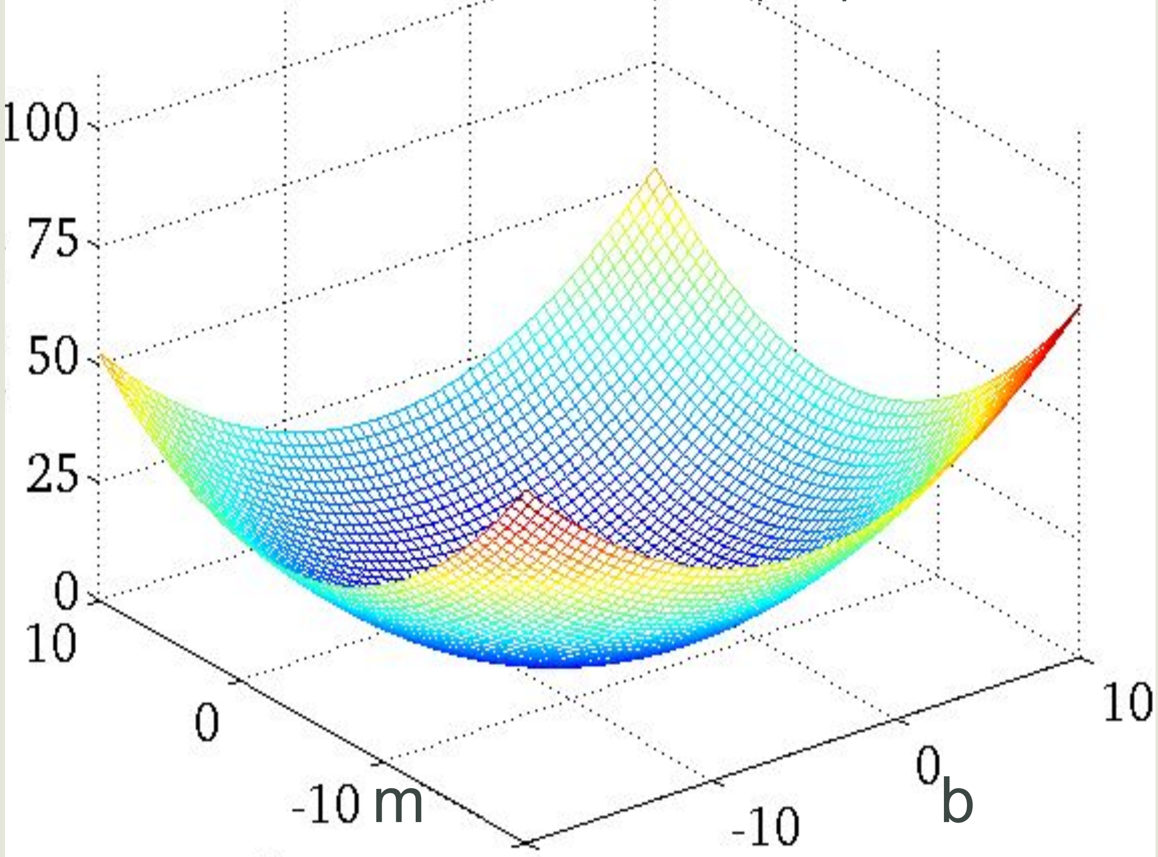
Minimize Cost

Yes let’s optimize!

Minimize Cost

Cost Function typically looks like this.

$$J(m,b) = (1/n)\ \Sigma\ (Y(x_i)-y_i)^2$$



Goal:
Minimize J(m,b)

Method:

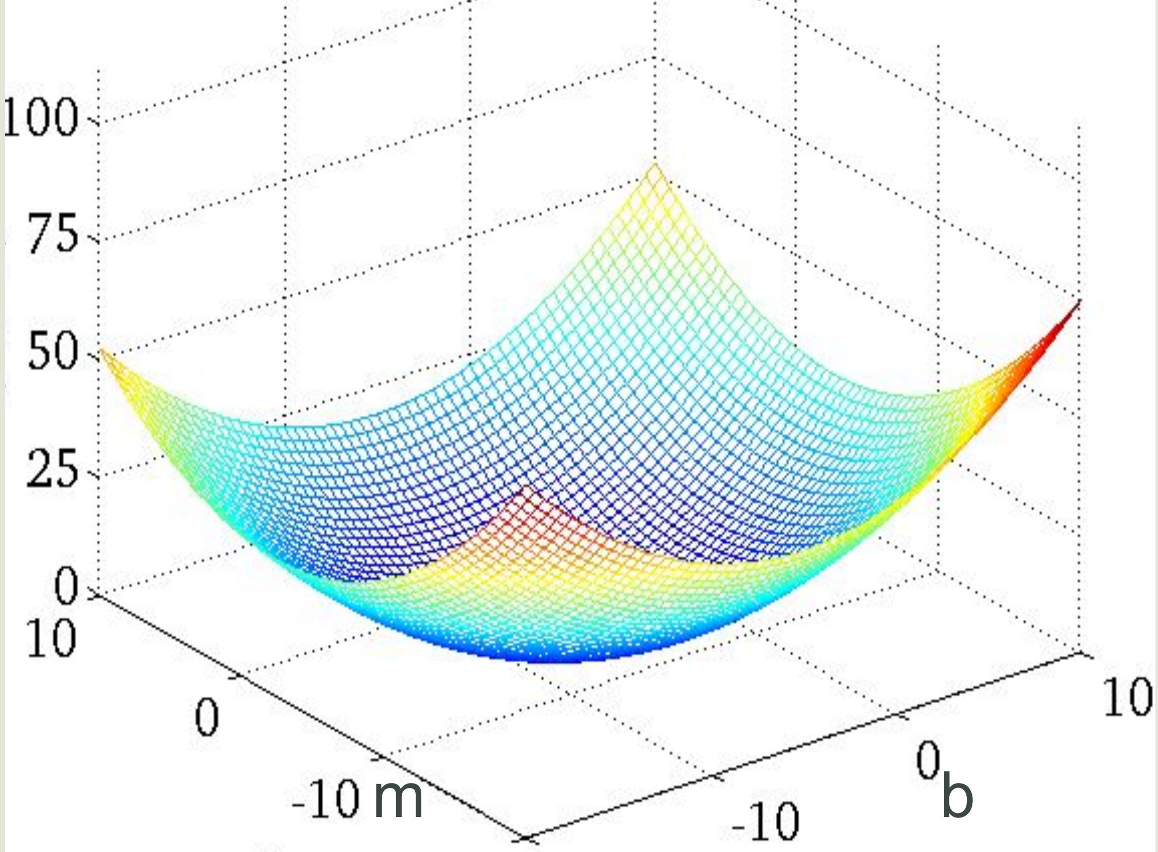
Start with some m_1, b_1

Keep changing m & b to reduce J until we hopefully end up at a minimum

What can go wrong?

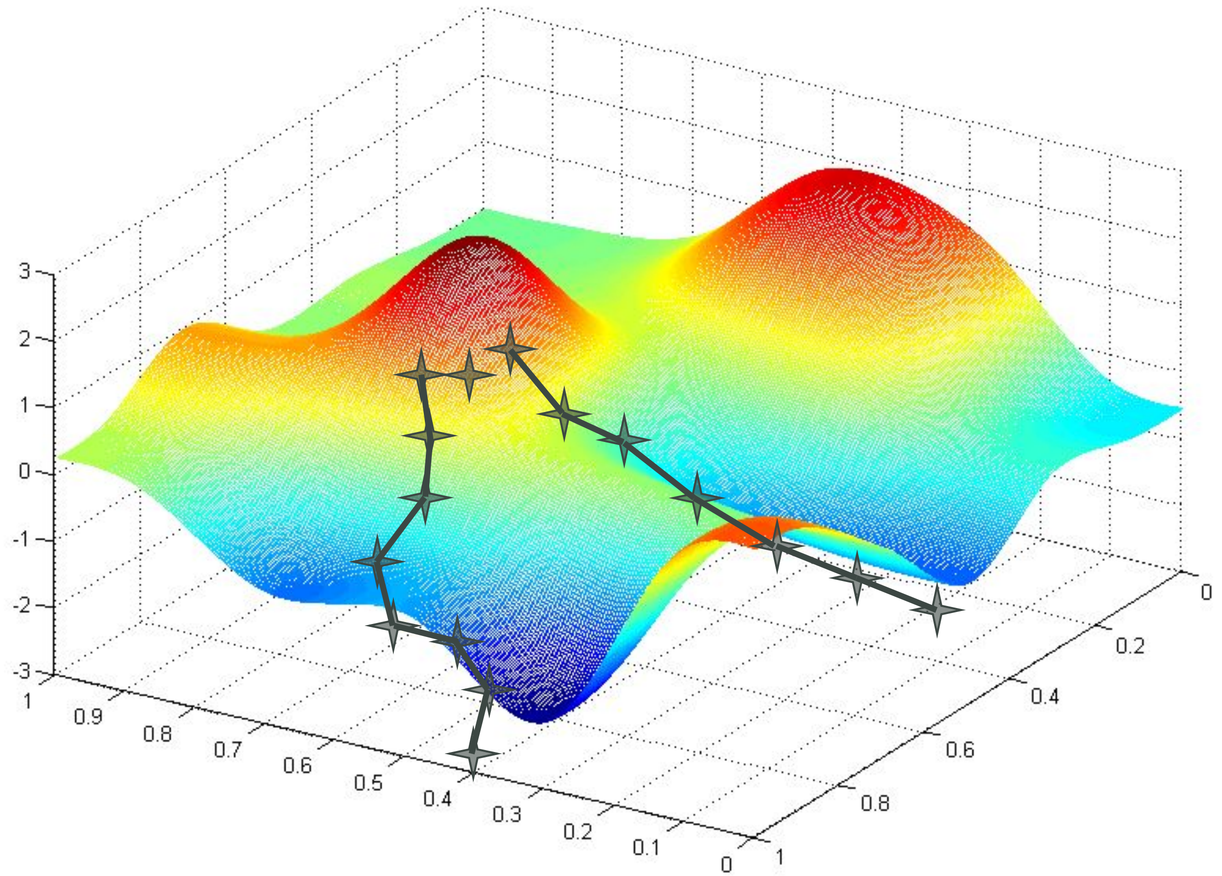
What can go wrong?

$$J(m,b) = (1/n)\ \Sigma\ (Y(x_i)-y_i)^2$$



- How to change m and b? Can we end up going in circles or in the wrong direction?
- How fast do we change them?
- Do we change them one by one or simultaneously?
- Will the function always look like this?
- Will it always have a single minima?
- Should the cost function only account for error?

Local vs Global **Extremum**



There may be more than one path:

What does this look like mathematically?

Instructor: Bhavik Gandhi

Learning Rate

Ohh, my rate of learning is amazing!

Rate of Approach

The gradient is simply the SLOPE of the surface.

It is the sensitivity to the parameters in question.

It is the partial derivative with respect to the inputs.

$J(m,b)_2 = J(m,b)_1 - \alpha \nabla J(m,b)_1$, ∇ = gradient,

α= rate of approach/learning rate (always +ve)

Think of the rate of approach or learning rate as the rate at which the model abandons old beliefs in favour of new data.

Rate of Approach

$J_2 = J_1 - \alpha \nabla J_1$, ∇= gradient, α= rate of approach

So we start by looking for the steepest descent path, then plug it in and make that the input for the next round and so on…

$$J_{i+1} = J_i - \alpha \nabla J_i$$

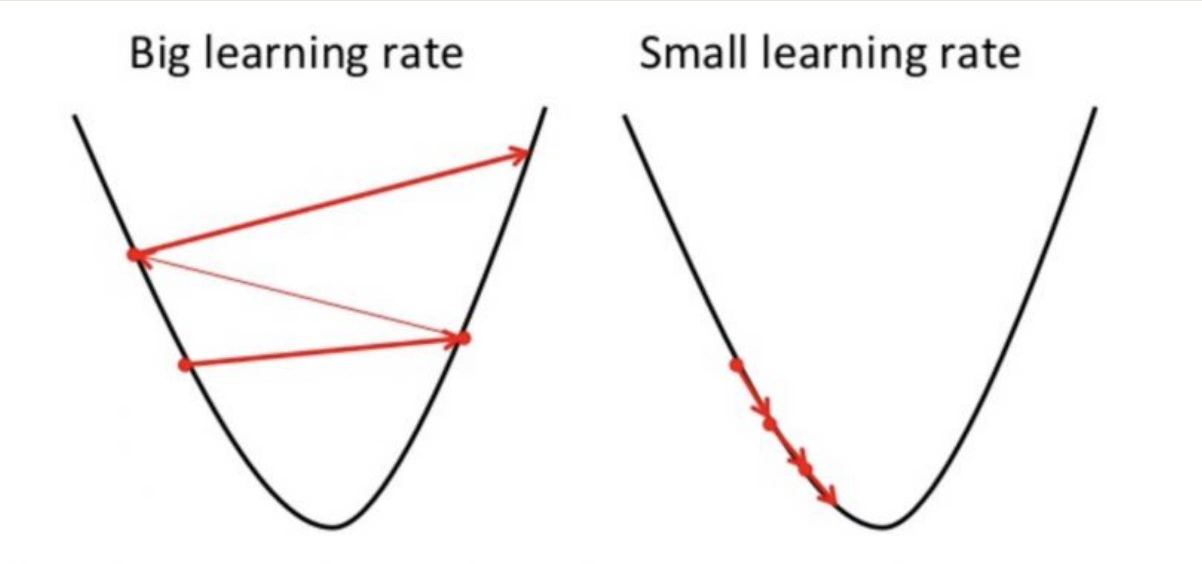
Since we started by finding the steepest path, we hope that the next jump will be smaller and smaller, til eventually we have no more gradient,

$$\nabla J_i = 0 \rightarrow J_{i+1} = J_i = J^* \text{ (optimal)}$$

Rate of approach

$J_2 = J_1 - \alpha \nabla J_{1}$, ∇ = gradient, α= rate of approach

If α is too small, the gradient descent will be small and slow (small steps). If it is too big, it will leap over the local minimum and jump out of the area of the curve:



Questions

What is important for a cost function?

What decides the magnitude and direction of the change of m and b?

Gradient Descent

Why not ascend?

Gradient Descent

We start as we did last time with our equation, of best fit,

$$y' = mx + b$$

We wish to minimize the MSE (mean squared error i.e. our residual distances):

$$\text{Loss} = (1/n)\Sigma(y - y')^2$$
$$= (1/n)\Sigma(y - (mx + b))^2$$

As before, we want to minimize loss wrt *m*, *b*,

$$\frac{\partial Loss}{\partial m} = 0, \qquad \frac{\partial Loss}{\partial b} = 0$$

Gradient Descent

Partial differentiation: $\frac{\partial f}{\partial x_i} =$? hold all other variables as constant, then differentiate.

Gradient Descent

$\text{Loss}=(1/n)\Sigma(y-(mx+b))^2$

$$\frac{\partial Loss}{\partial b} = \frac{-2\Sigma(y-mx-b)}{n} = \frac{-2\Sigma(y-y\prime)}{n} = 0$$

$$\frac{\partial Loss}{\partial m} = \frac{-2\Sigma x(y-mx-b)}{n} = \frac{-2\Sigma x(y-y\prime)}{n} = 0$$

Recall, $J_{i+1} = J_i - \alpha \nabla J_i$, so,

$$m_{i+1} = m_i - \alpha\frac{\partial L_i}{\partial m}$$

$$b_{i+1} = b_i - \alpha\frac{\partial L_i}{\partial b}$$

Gradient Descent

Loss=(1/n)Σ(y-(mx+b))^2

$$\frac{\partial Loss}{\partial b} = \frac{-2\Sigma(y-mx-b)}{n} = \frac{-2\Sigma(y-y\prime)}{n} = 0$$

$$\frac{\partial Loss}{\partial m} = \frac{-2\Sigma x(y-mx-b)}{n} = \frac{-2\Sigma x(y-y\prime)}{n} = 0$$

Recall, $J_{i+1} = J_i - \alpha \nabla J_i$, so,

$$b_{i+1} = b_i + 2\alpha\frac{\Sigma(y-m_i x-b_i)}{n}$$

$$m_{i+1} = m_i + 2\alpha\frac{\Sigma(yx-m_i x^2-b_i x)}{n}$$

Gradient Descent

So we now have a way to choose a learning rate, α, and a starting (m_0, b_0), then make new predictions and keep going until there is little change in m & b:

$$b_{i+1} = b_i + 2\alpha \frac{\Sigma(y - m_i x - b_i)}{n} = b_i + 2\alpha \frac{\Sigma(y - y')}{n}$$

$$m_{i+1} = m_i + 2\alpha \frac{\Sigma(yx - m_i x^2 - b_i x)}{n} = m_i + 2\alpha \frac{\Sigma x(y - y')}{n}$$

Key Assumptions

Assumptions again?

Key Assumptions

1. Linearity

The underlying relationship is LINEAR!

-slope doesn't depend on other variables

-other variables can each have an additive effect

-failure out of range can easily exist, even if ok in range

2. Independence:

Data is UNCORRELATED

-failure (e.g. auto-correlation lag) can result in skewed fits (under/overfitting in some regimes)

Key Assumptions

3. Errors (residuals) follow a Normal Distribution
   - failure could be data is not truly linear or have outliers
   - failure could mean unexplained influencing variables
   - failure could make prediction intervals very weak

4. Homoscedasticity:
   Errors have similar variance (no progression)
   - failure could lead to skewed weighting of some data
   - failure can cause poor estimation

Data Scaling

Does scaling help?

Scaling approaches

1. Standardization

1. Normalization

1. Scaling

Theory vs Practice

Are all assumptions satisfied? How much do we care?

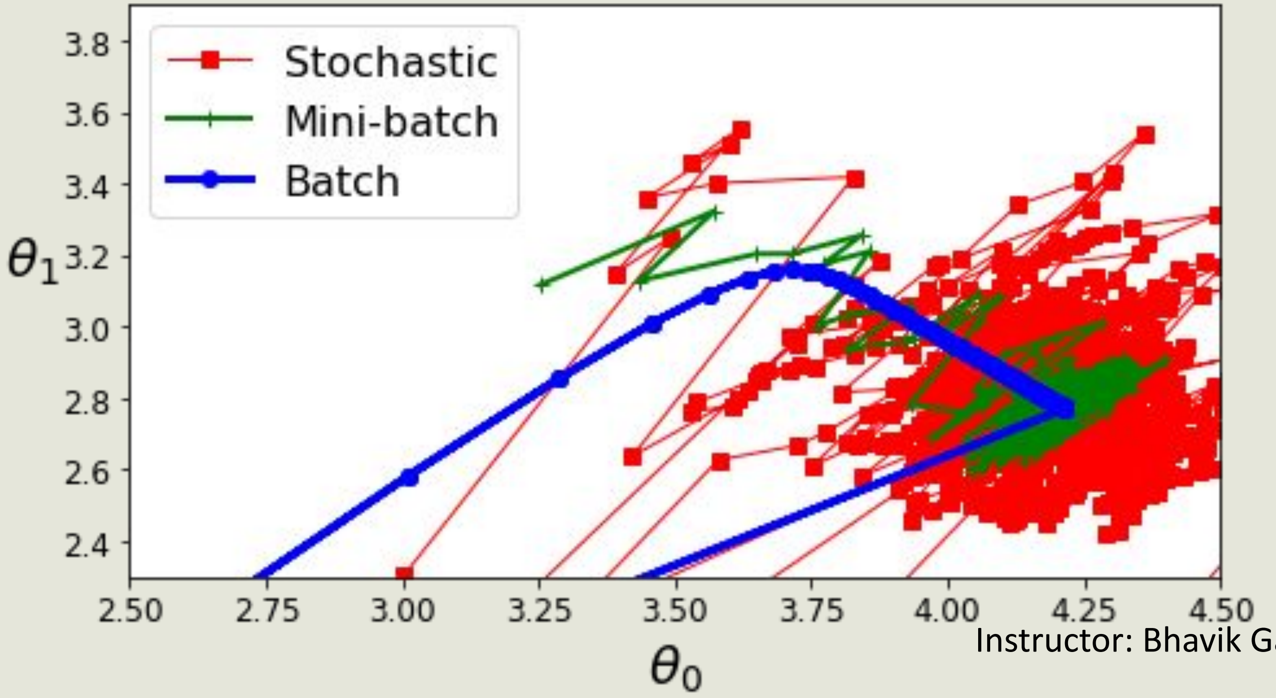
Do we ever reach the absolute minima? When do we stop?

Types of Gradient Descent

1. Batch Gradient Descent
2. Stochastic Gradient Descent
3. Mini Batch Gradient Descent

Gradient Descent is also leveraged

by Neural Networks/DL



Instructor: Bhavik Gandhi

Questions

Gradient Descent vs OLS?

Gradient Descent	OLS
Can scale with large data using stochastic or mini batch options	Does not scale well with large data
Can be used for non-linear cost functions	Cannot be used for non-linear cost functions
Practical challenges in achieving exact optima (but can come very close)	Optimal for convex functions with single global minima
Needs scaled data for optimal speed	Scaled data is only needed for better interpretability

Exercises, Assignments & Coding

Ok, let's understand that!

Quick Review

Ok, let's understand that!

Linear Regression

- Gradient Descent
- Learning Rate
- Hyperparameters
- Regularization

And we learnt Gradient Descent!! Yay!!

Ok, that was a lot of Maths! Let’s make sure we practice it!

Credits

https://towardsdatascience.com/linear-regression-simplified-ordinary-least-square-vs-gradient-descent-48145de2cf76

https://www.slideshare.net/kaz_yos/visual-explanation-of-ridge-regression-and-lasso

https://www.slideshare.net/DerekKane/data-science-part-xii-ridge-regression-lasso-and-elastic-nets

www.slideshare.net/sachinkumarsl/ridge-lasso-regression

https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression

https://www.danli.org/2021/06/06/hands-on-machine-learning/#chapter-4-training-linear-models