Regularization!

Class 5: 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

Instructor: Bhavik Gandhi

*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.
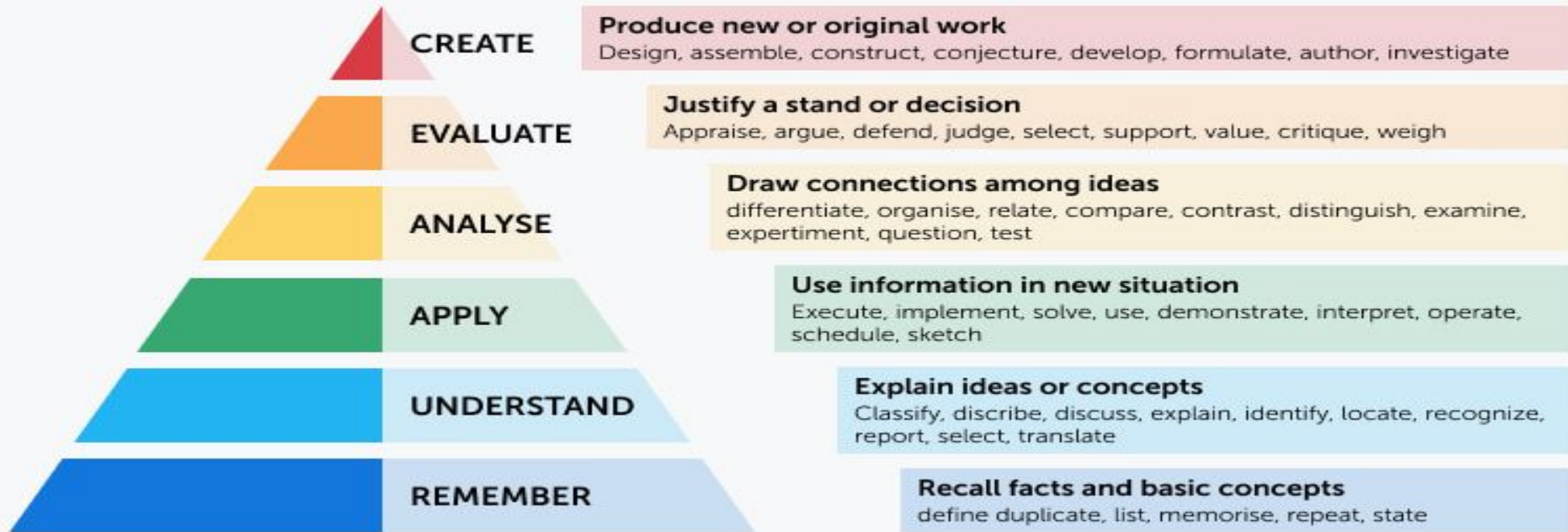
The class will usually be 1 - 1.5 hrs on instruction

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

Instructor: Bhavik Gandhi

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.

Instructor: Bhavik Gandhi

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.

Instructor: Bhavik Gandhi

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.

Instructor: Bhavik Gandhi

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

Instructor: Bhavik Gandhi

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!

Instructor: Bhavik Gandhi

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.

Instructor: Bhavik Gandhi

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, <u>apart from following instructions and having a strong sense of integrity</u>, 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 about Regularization, L1 and L2 norms

- You will learn about different types of data and learn how to deal with them

- You will learn some model tuning through data preprocessing and feature engineering

- You will do examples in Python

Instructor: Bhavik Gandhi

Seminars, Conferences et. al.

Data and the Metaverse
https://www.eventbrite.ca/e/data-and-the-metaverse-dsiutm-tickets-479515914007

DS ARES Seminar
https://www.eventbrite.ca/e/data-science-applied-research-and-education-seminar-rina-friedberg-tickets-483940096857

Responsible Data Science
https://www.eventbrite.ca/e/responsible-data-science-dsiutm-data-digests-data-and-students-tickets-467880803077

Toronto Health Datathon
https://www.eventbrite.ca/e/toronto-health-datathon-tickets-502120755697

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
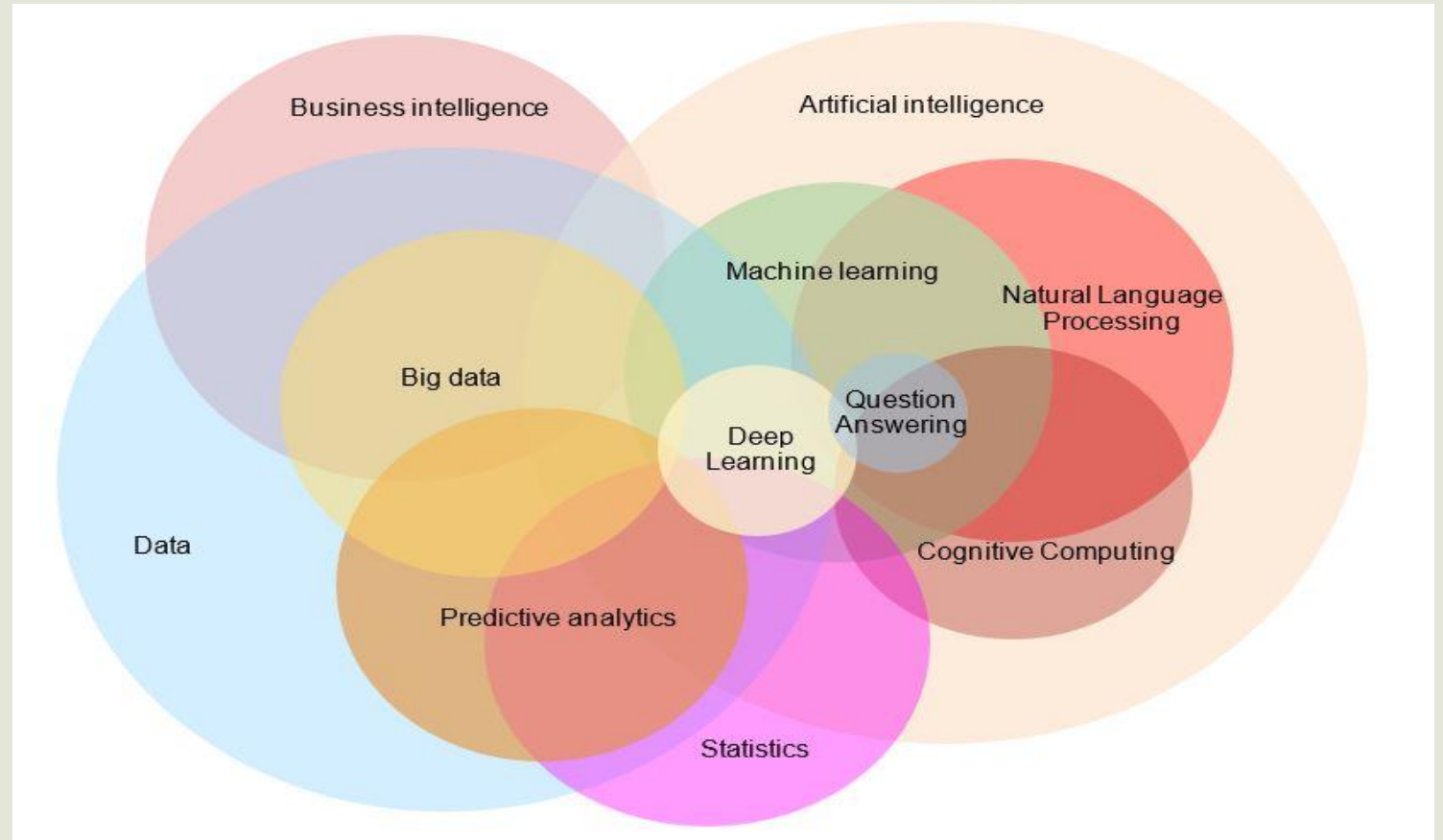
Supervised Learning & Linear Regression

- Types of Machine Learning

- Bias and Variance

- Cost Function & Cross Validation

- Regression vs Classification

- Linear Regression

- Correlation & R2

- Ordinary Least Squares

- Assumptions & Scaling
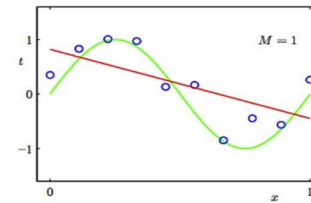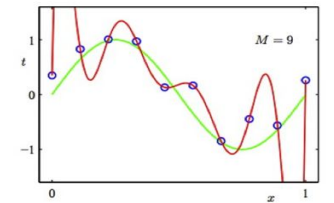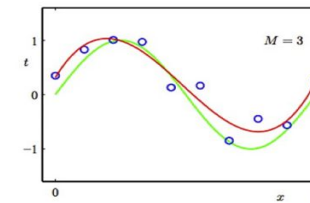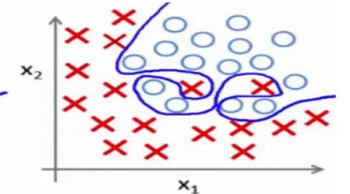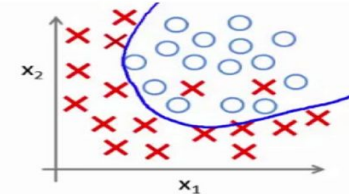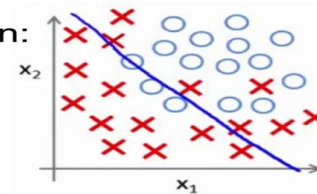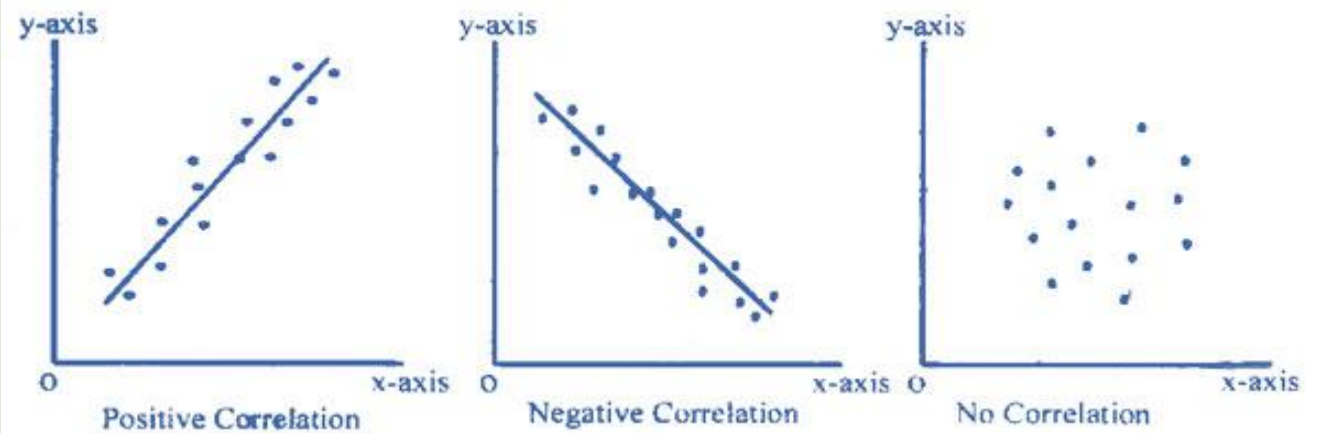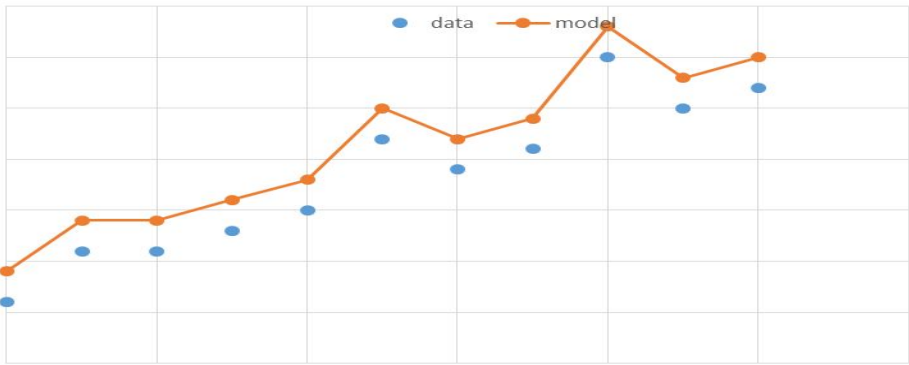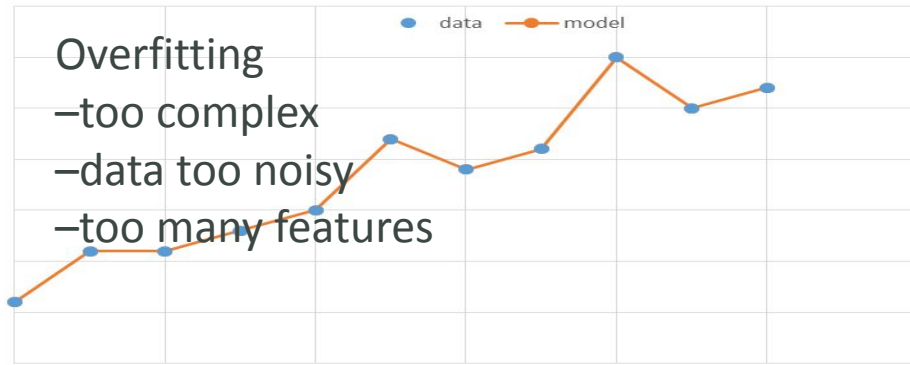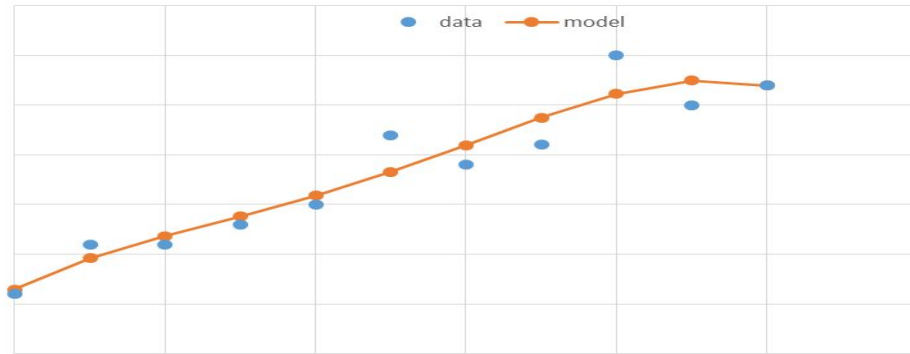
Bias vs Variance

| 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	

Gradient Descent

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?

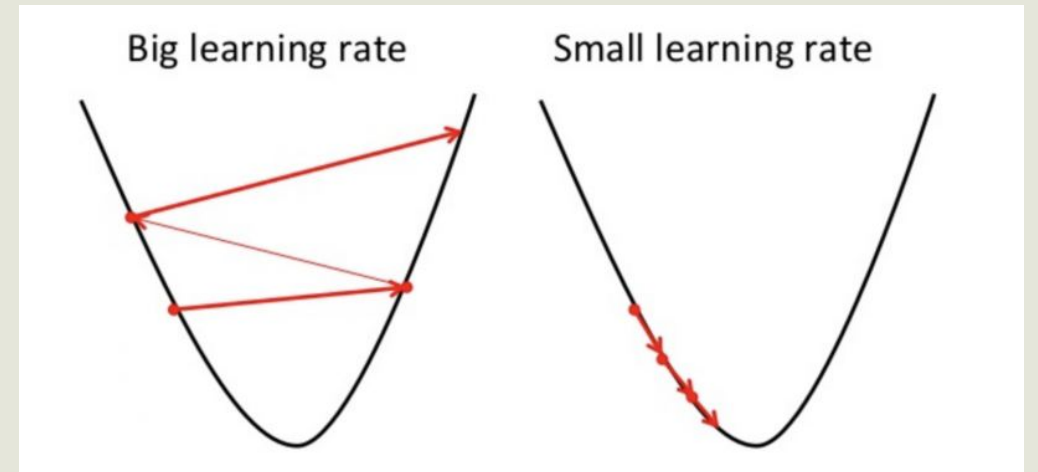IRL: Batch, Stochastic, Mini-batch

GD vs OLS

Hyperparameters, Validation
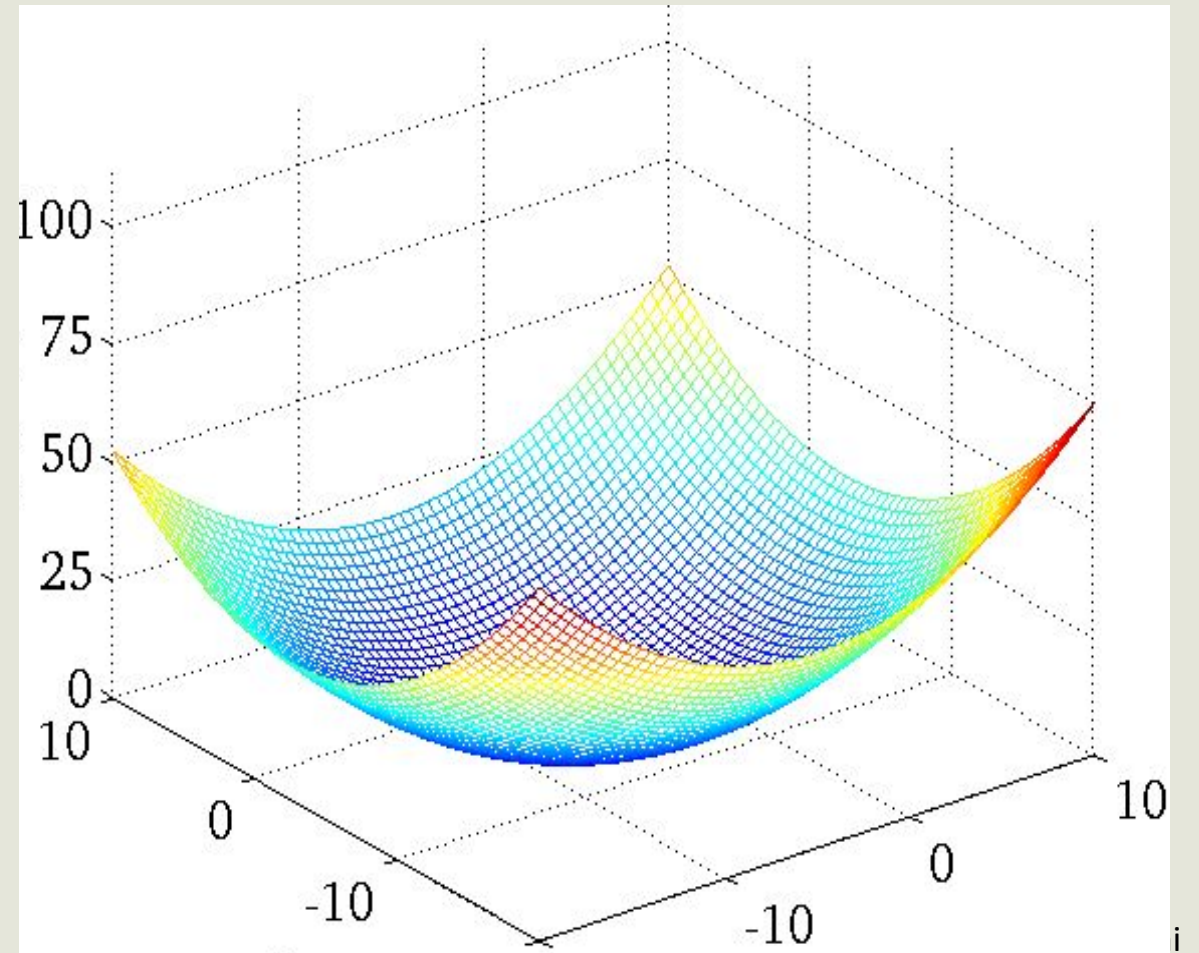
Gradient Descent

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

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

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

Big learning rate   Small learning rate
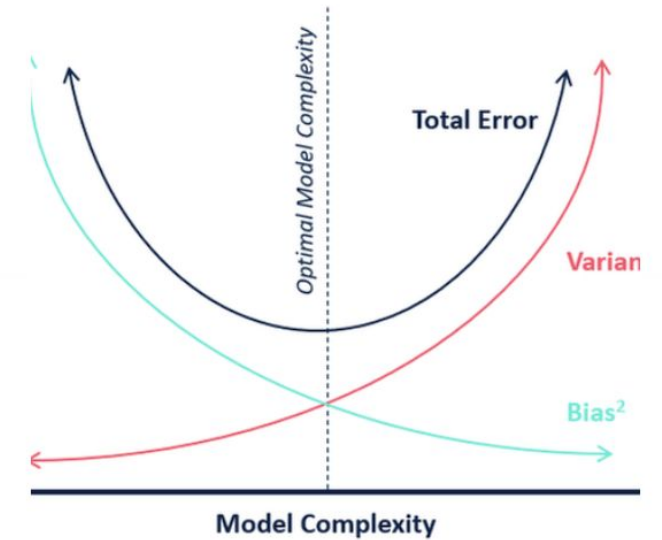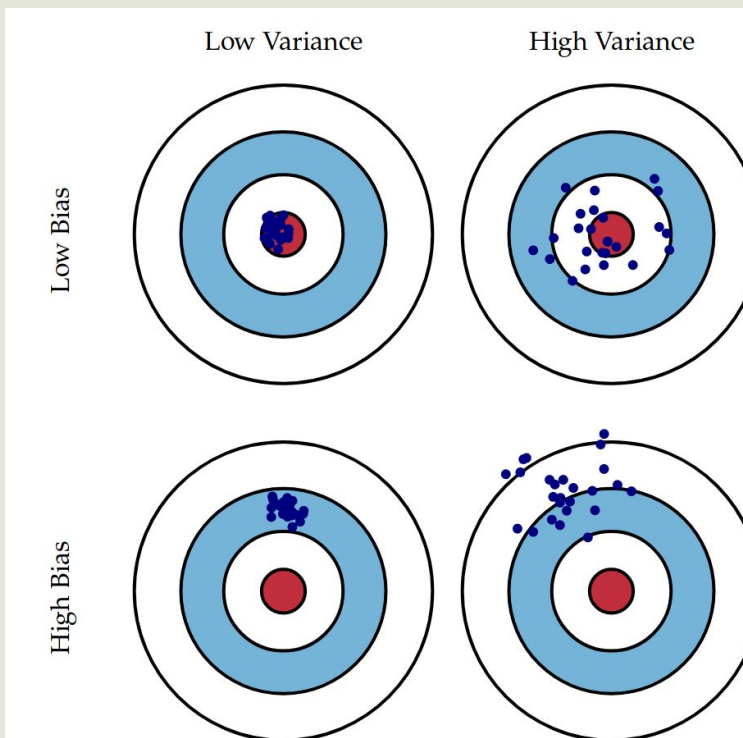
Instructor: Bhavik Gandhi

Regularization

Isn't it all regular?

Regularization

$$J(\lambda) = RMSE + \lambda Regularization(weights)$$

What is the regularization here?

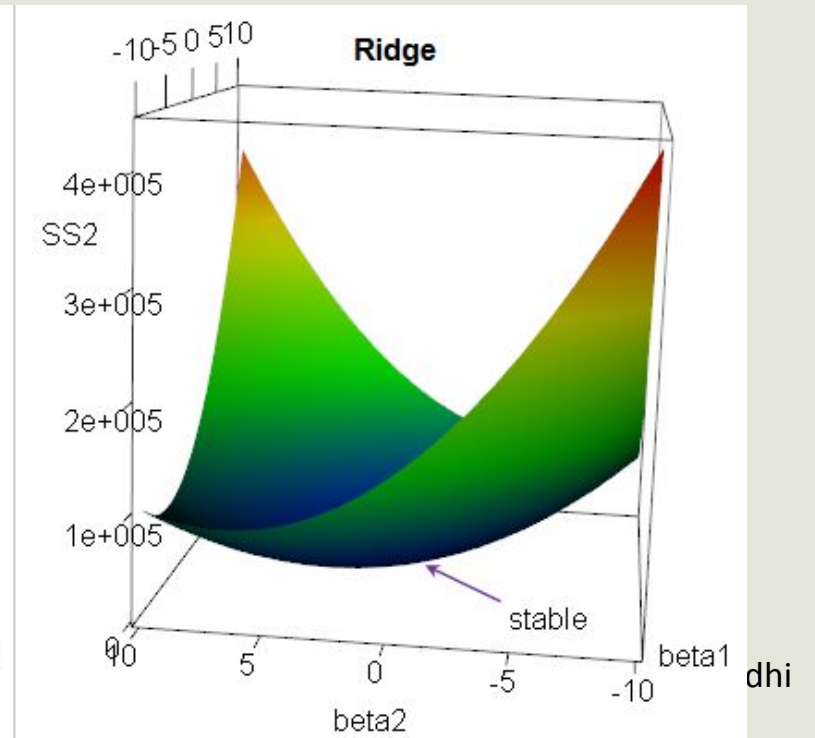It's a function on the weights

What functions can we have?

Instructor: Bhavik Gandhi
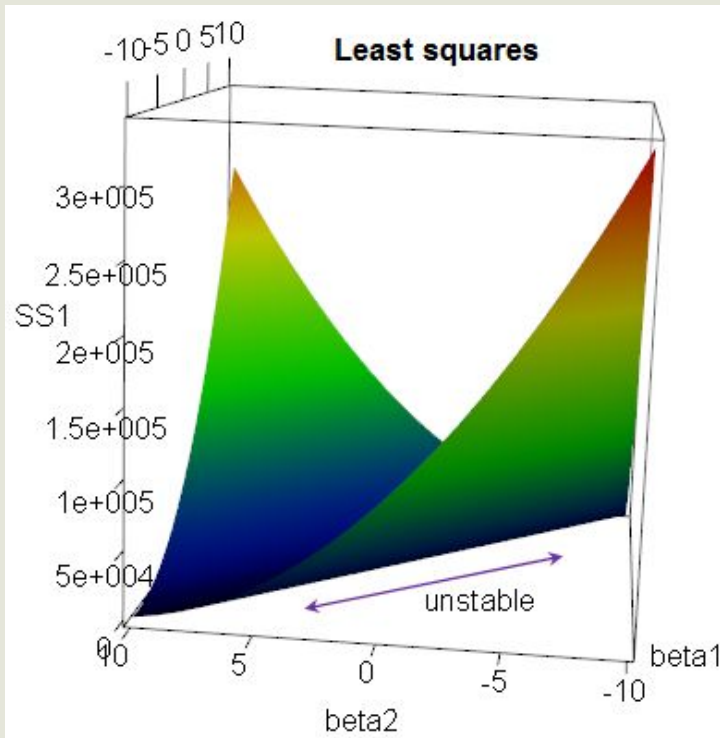
Ridge Regression

For Ridge regression

the function used is

a square of coefficients

The parameter space for

Ridge Regressions shows a

Ridge (hence the name)

It is also called as L2 regression

$$\operatorname*{arg\,min}_{\beta} \left[\sum_{i=1}^{n} \left(Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ji} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|^2 \right]$$

Ridge Regression

Least Absolute Shrinkage and Selection Operator

For LASSO regression the function used
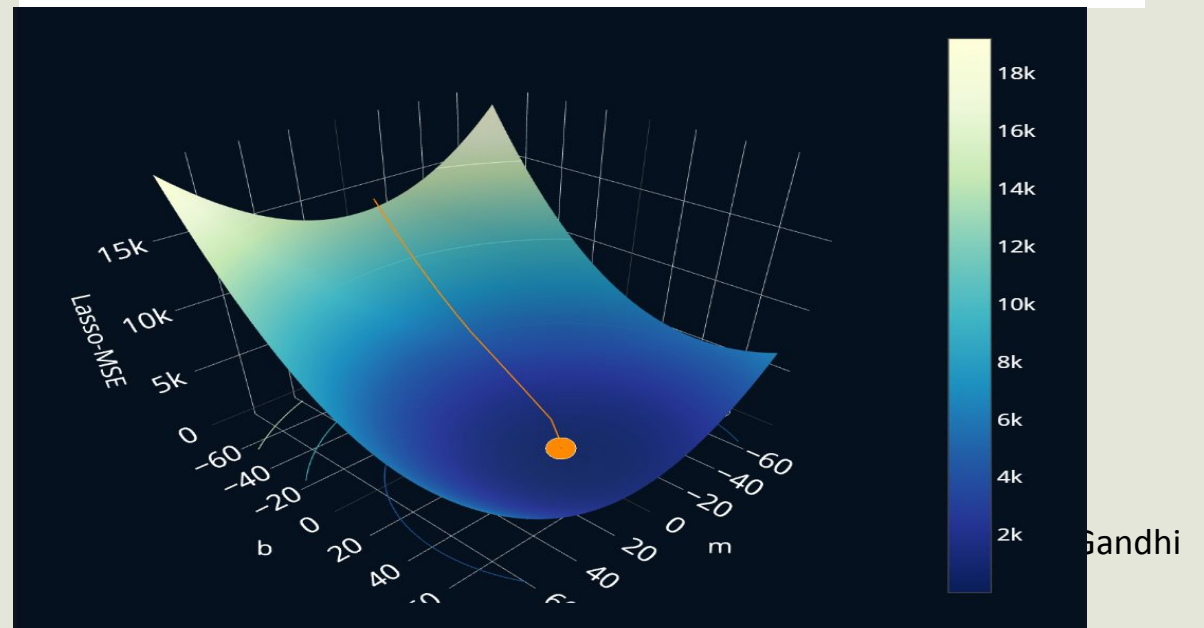
is an absolute of coefficients

It is also called as L1 Regression
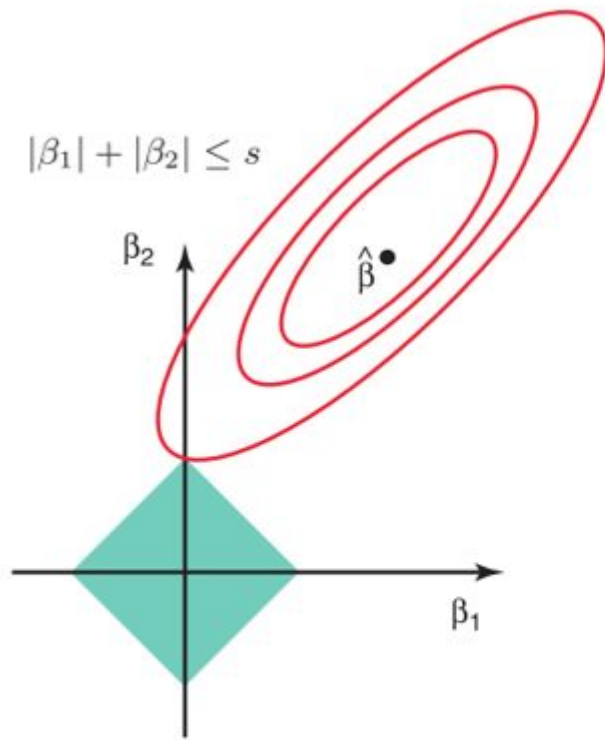
LASSO is also a good selection technique

i.e. LASSO is better at turning coefficients to 0
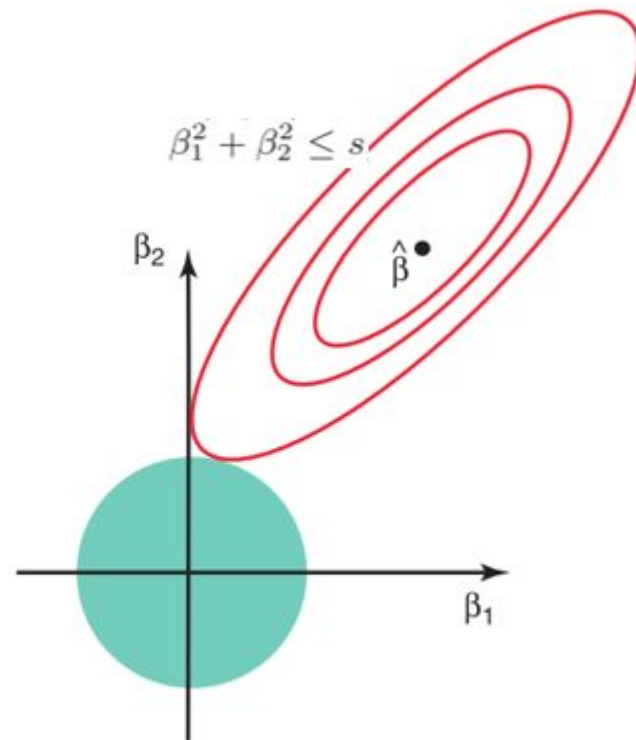
rather than reducing coefficients

$$\arg\min_{\beta} \left[\sum_{i=1}^{n} \left(Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ji} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right]$$
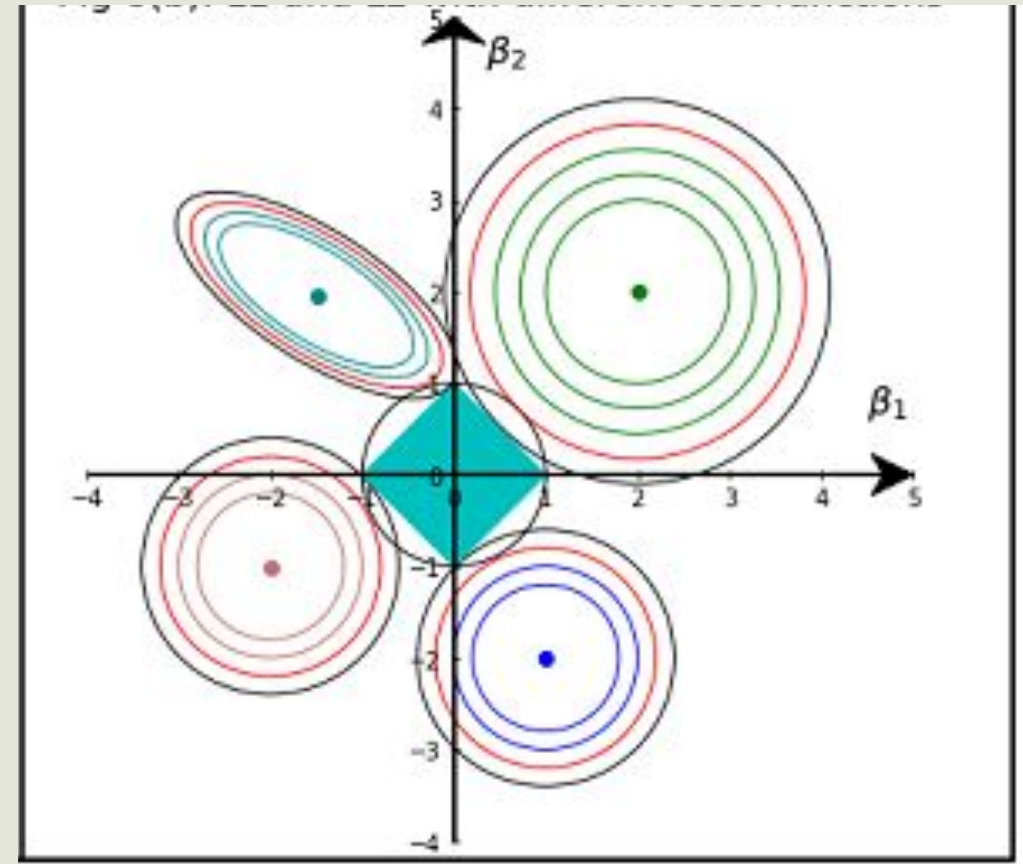
Gandhi

Ridge vs LASSO

Lasso Regression

$|\beta_1| + |\beta_2| \le s$

Ridge Regression

$\beta_1^2 + \beta_2^2 \le s$

Instructor: Bhavik Gandhi

Ridge vs LASSO

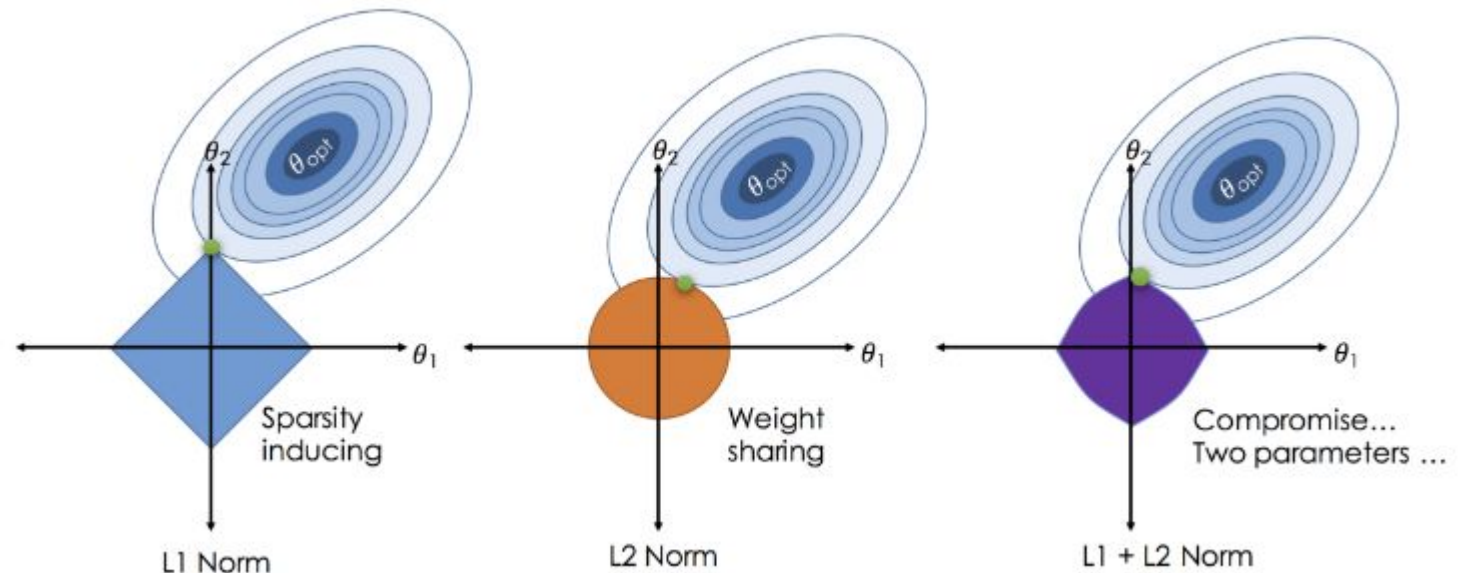| Ridge Regression	Lasso Regression
The penalty term is the sum of the squares of the coefficients (L2 regularization).	The penalty term is the sum of the absolute values of the coefficients (L1 regularization).
Shrinks the coefficients but doesn't set any coefficient to zero.	Can shrink some coefficients to zero, effectively performing feature selection.
Helps to reduce overfitting by shrinking large coefficients.	Helps to reduce overfitting by shrinking and selecting features with less importance.
Works well when there are a large number of features.	Works well when there are a small number of features.
Performs "soft thresholding" of coefficients.	Performs "hard thresholding" of coefficients.

Instructor: Bhavik Gandhi

Compromise between L1 & L2

Lasso and Ridge both have their own advantages so a compromise is Elastic Net

But the downside is there's another hyperparameter

Elastic Net

$$\arg\min_{\beta} \left[\sum_{i=1}^{n} \left(Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ji} \right)^2 + \lambda \sum_{j=1}^{p} \left(\alpha|\beta_j|^2 + (1-\alpha)|\beta_j| \right) \right]$$

θ_2 θ_{opt} θ_1

Sparsity inducing

L1 Norm

θ_2 θ_{opt} θ_1

Weight sharing

L2 Norm

θ_2 θ_{opt} θ_1

Compromise... Two parameters ...

L1 + L2 Norm

Questions

What is the difference between L1, L2 and ElasticNet?

Instructor: Bhavik Gandhi

Attendance

I am present!

Coding

Yay! Let's practice

Quick Review

Ok, let's understand that!

Linear Regression

- Regularization

- Ridge Regression

- LASSO

- ElasticNet

Instructor: Bhavik Gandhi

And we learnt Regularization!! Yay!!

Ok, that was a lot of stuff! Let's make sure we try it!