Artificial Intelligence and Expert Systems

Machine Learning (I)

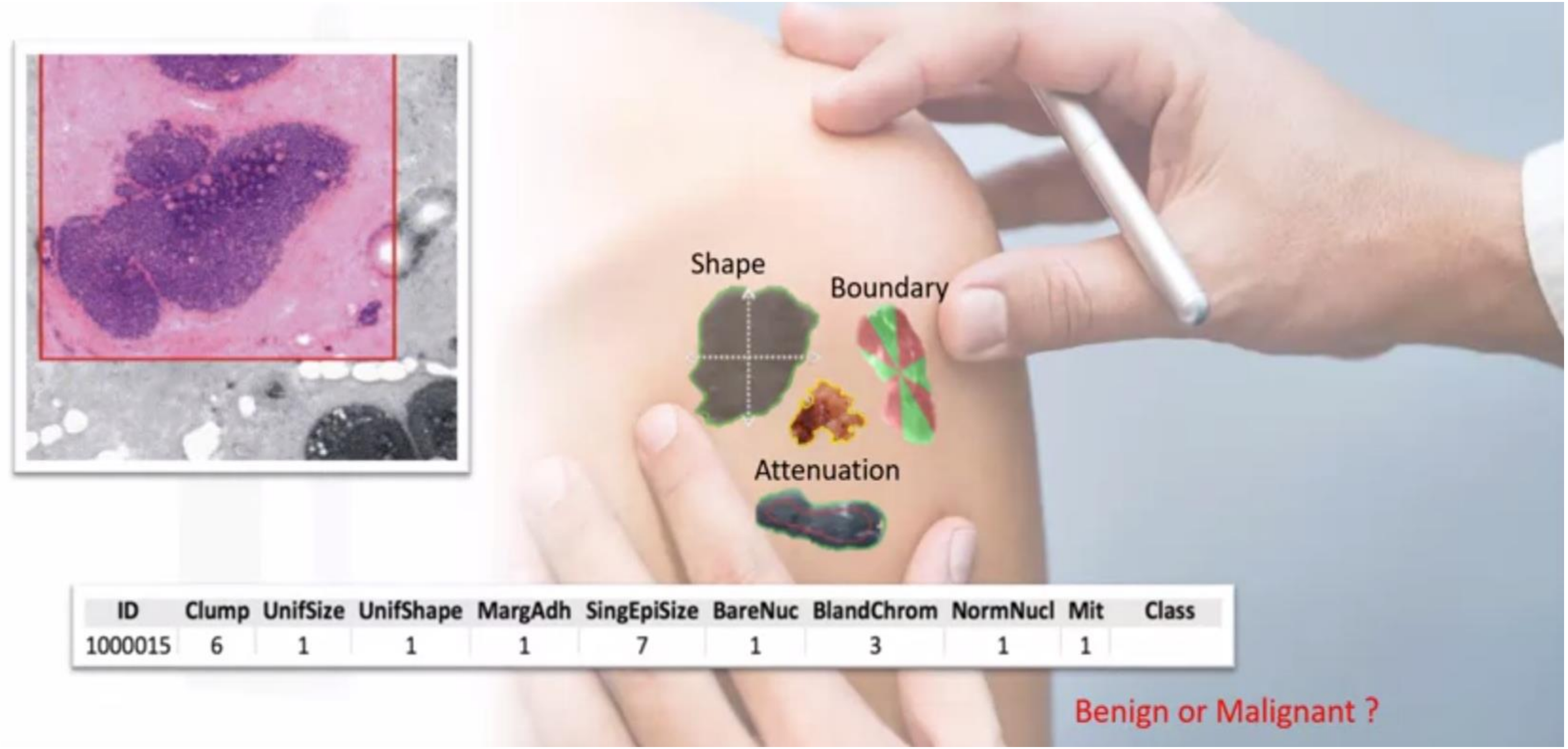
Lecture 1

Dr. Esmaeil Najafi, Javad Khorramdel
K.N. Toosi University of Technology

Today’s agenda

- Introduction to machine learning
- Simple linear regression
- Multiple linear regression
- Non-linear regression
- Evaluation approaches
- Evaluation metrics for regression
- Implementation of models in practice

Introduction to machine learning

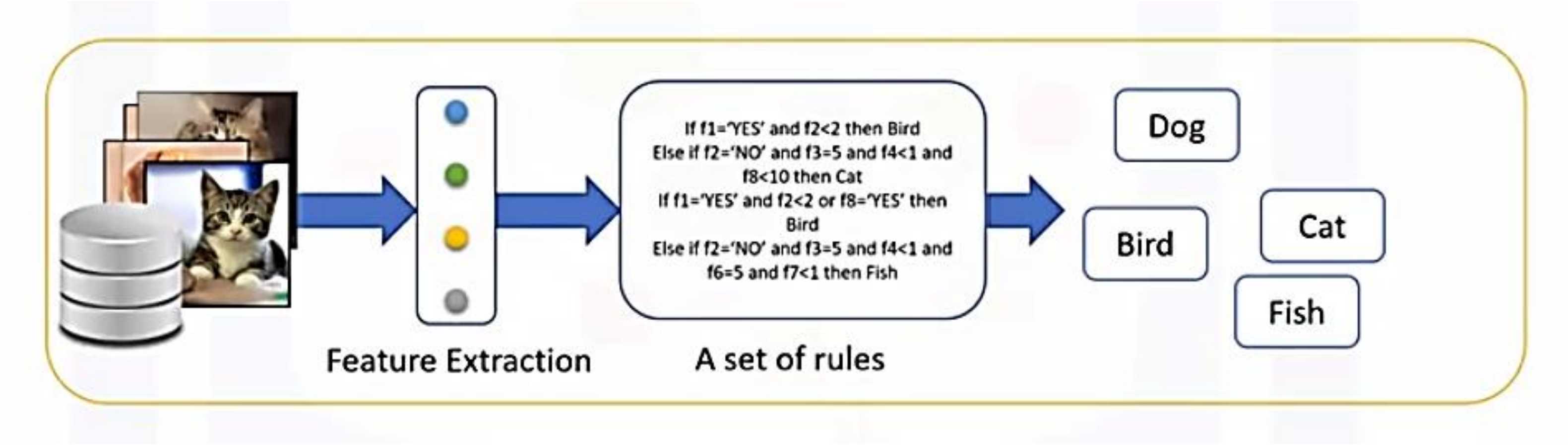


Introduction to machine learning

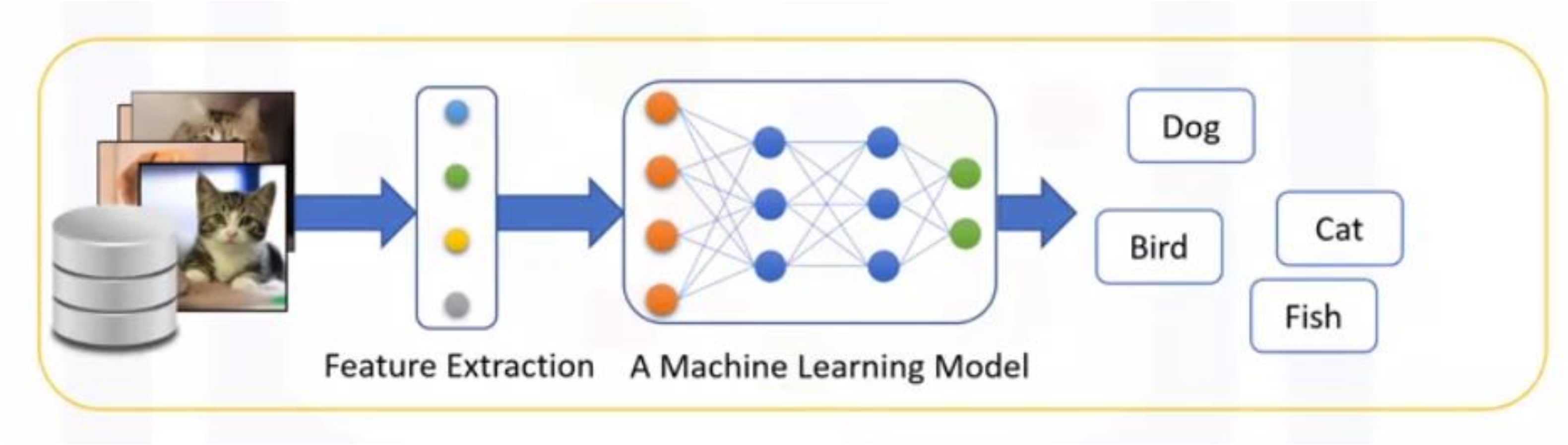
ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000015	6	1	1	1	7	1	3	1	1	?

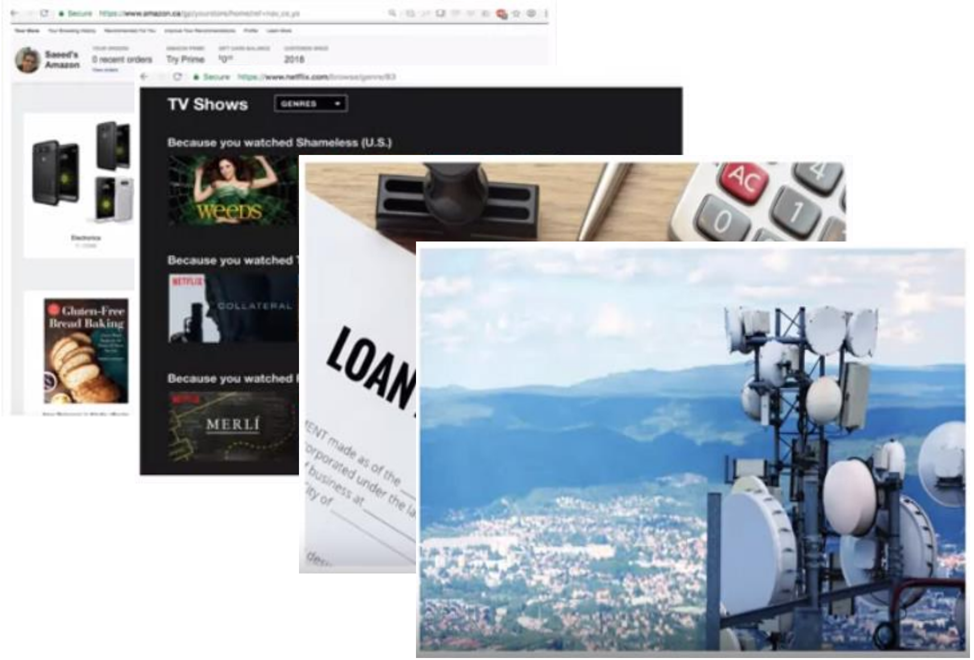
How machine learning works?



How machine learning works?



Machine learning applications



Major machine learning techniques

- Regression/Estimation
  - Predicting continuous values
- Classification
  - Predicting the item class/category of a case
- Clustering
  - Finding the structure of the data, summarization
- Associations
  - Associating frequent co-occurring items/events

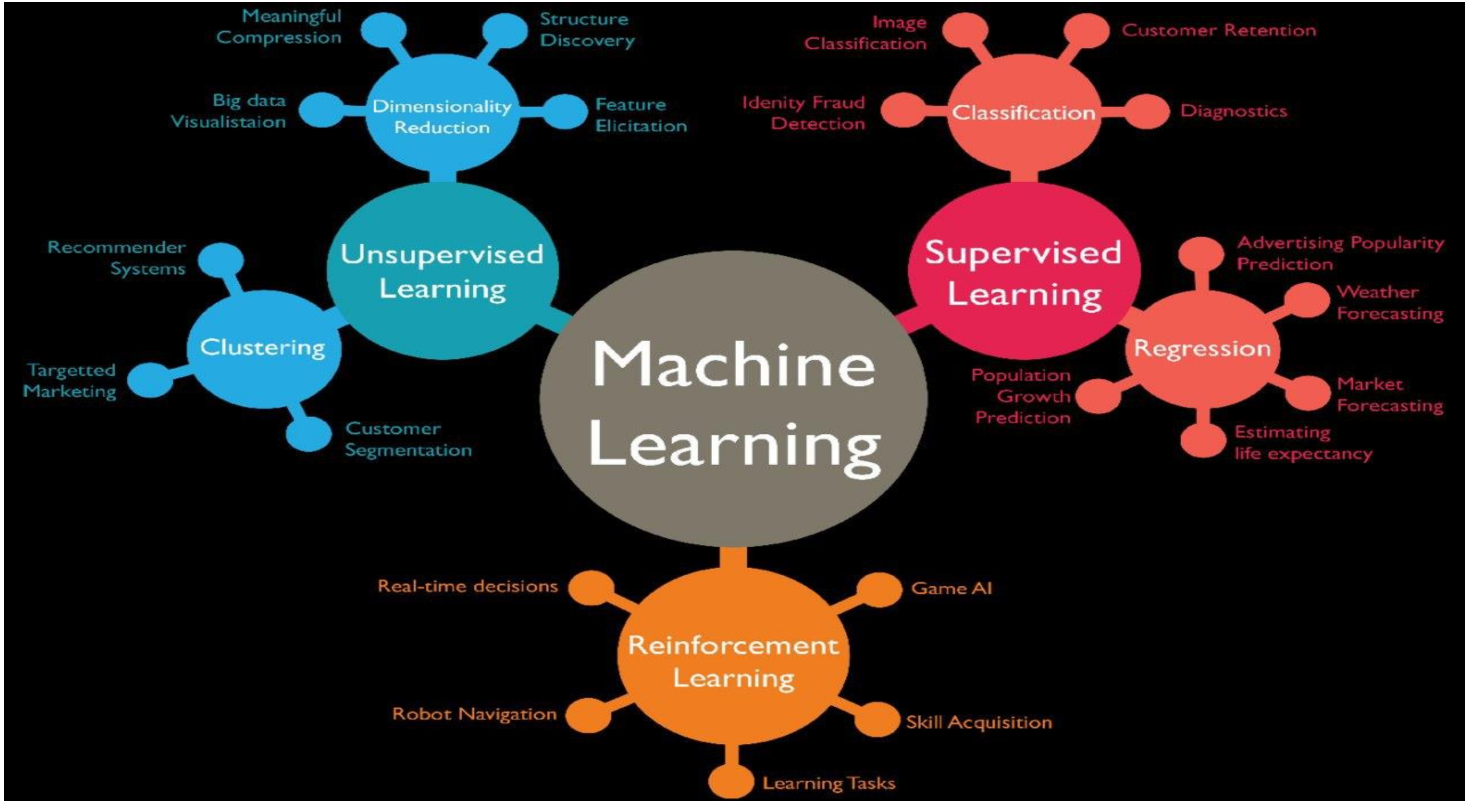
Major machine learning techniques

- Anomaly detection
  - Discovering abnormal and unusual cases
- Sequence mining
  - Predicting next events; click-steam (Markov Model, HMM)
- Dimension reduction
  - Reducing the size of data (PCA)
- Recommendation systems
  - Recommending items

Major machine learning techniques

- Which Machine Learning technique is proper for grouping of similar cases in a dataset, for example to find similar patients, or for customers segmentation in a bank?
  - a) Clustering
  - b) Classification
  - c) Regression
  - d) Recommender system

Machine Learning



Difference between artificial intelligence, machine learning and deep learning

- Artificial intelligence:
  - Computer vision
  - Language processing
  - Creativity
  - Etc.
- Machine learning:
  - Classification
  - Clustering
  - Neural network
  - Etc.

- Revolution in ML:
  - Deep learning

Python libraries for machine learning

NumPy

SciPy

matplotlib

pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

scikits learn

More about scikits learn

- Free software machine learning library
- Classification, clustering and regression algorithm
- Works with NumPy and SciPy
- Great documentation
- Easy to implement

Supervised vs Unsupervised

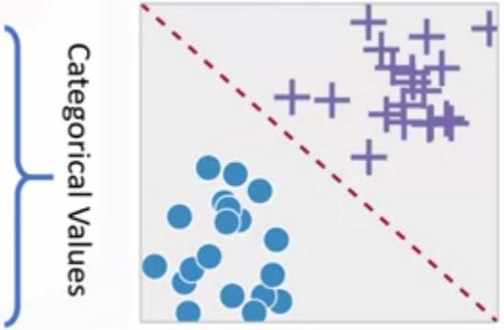
- Supervised:
  - We “teach the model” then with that knowledge, it can predict unknown or future instances.
  - But how exactly do we teach a model?
    - ✓ We teach the model with a labeled dataset.

ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

Supervised vs Unsupervised

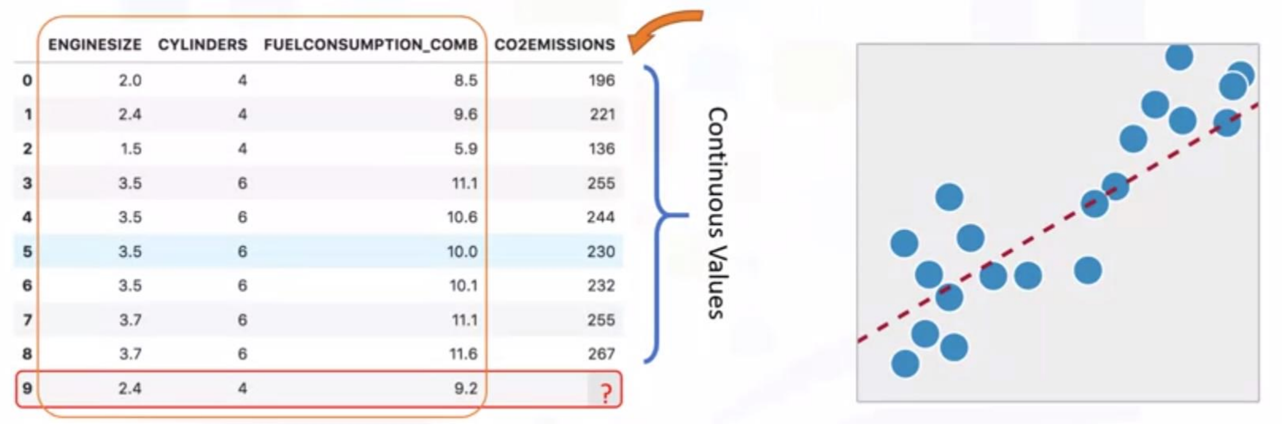
- Supervised:
  - Classification:
    - ✓ Classification is the process of predicting discrete class labels or categories.

ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign



Supervised vs Unsupervised

- Supervised:
  - Regression:
    - ✓ Regression is the process of predicting continuous values.



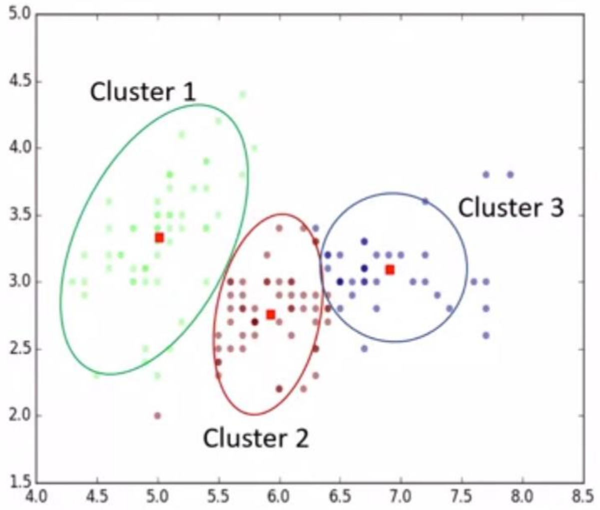
	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Supervised vs Unsupervised

- Unsupervised:
  - The model works on its own to discover information that may not be visible to human eye.
  - Has more difficult algorithms than supervised learning.
  - Unsupervised learning techniques:
    - ✓ Dimension reduction
    - ✓ Density estimation
    - ✓ Market basket analysis
    - ✓ Clustering

Supervised vs Unsupervised

- Unsupervised:
  - Clustering:
    - ✓ Clustering is grouping of data points that are somehow similar. Some applications are:
      - Discovering structure
      - Summarization
      - Anomaly detection



Supervised vs Unsupervised

- Supervised
  - Classification:
    - ✓ Classifies labeled data
  - Regression:
    - ✓ Predicts trend using previous **labeled** data
  - Has more evaluation methods than unsupervised methods
  - Controlled environment

- Unsupervised
  - Clustering:
    - ✓ Finds patterns and groupings from **unlabeled** data
  - Has fewer evaluation methods than supervised methods
  - Less controlled environment


BREAK TIME

What is regression?

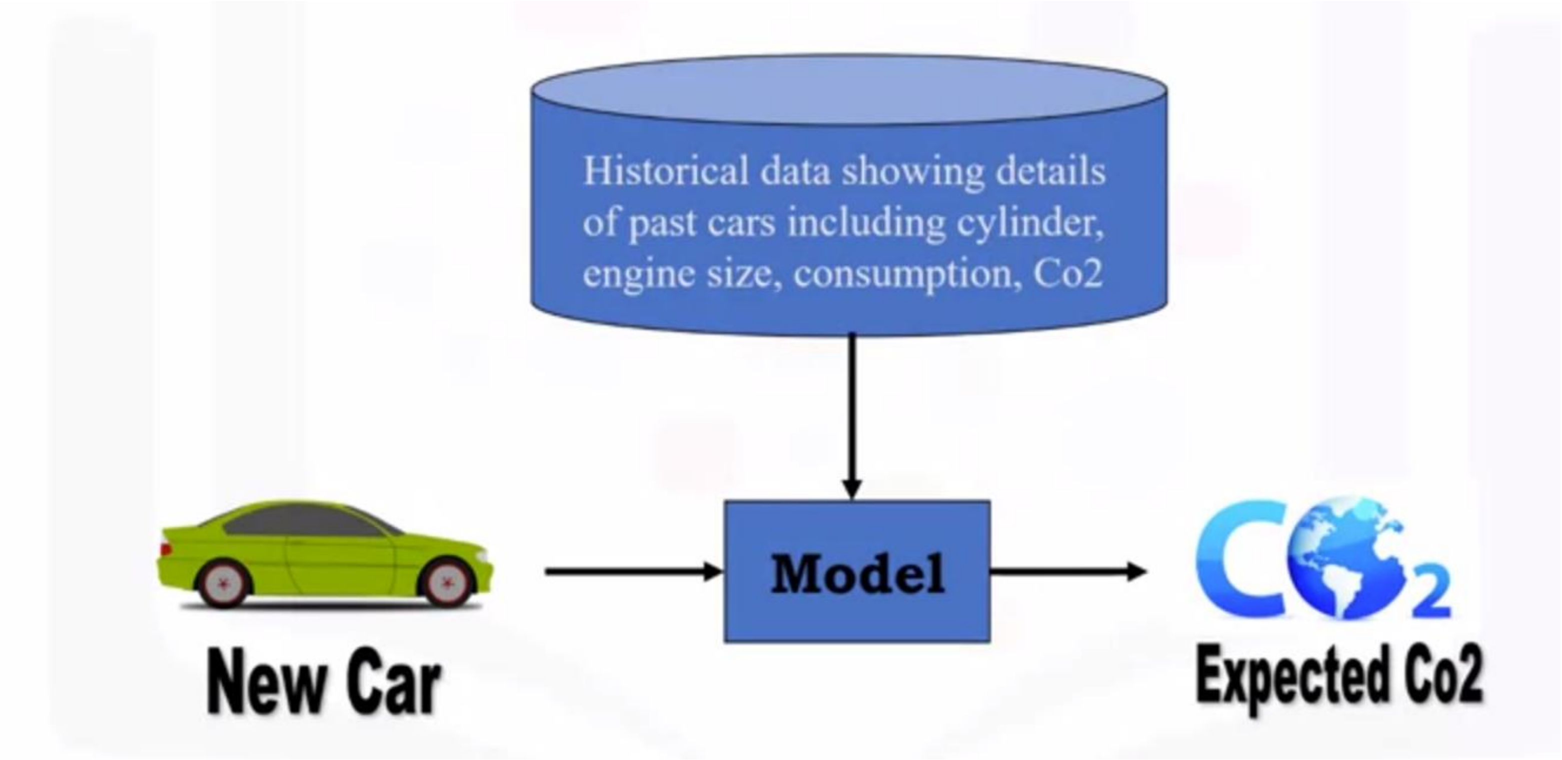
X: Independent variable

Y: Dependent variable

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Regression is the process of predicting a continuous value

What is a regression model?



What is a regression model?

- Simple regression:
  - Simple linear regression
  - Simple non-linear regression
- Multiple regression:
  - Multiple linear regression
  - Multiple non-linear regression

Predict CO2 emission vs EngineSize of all cars

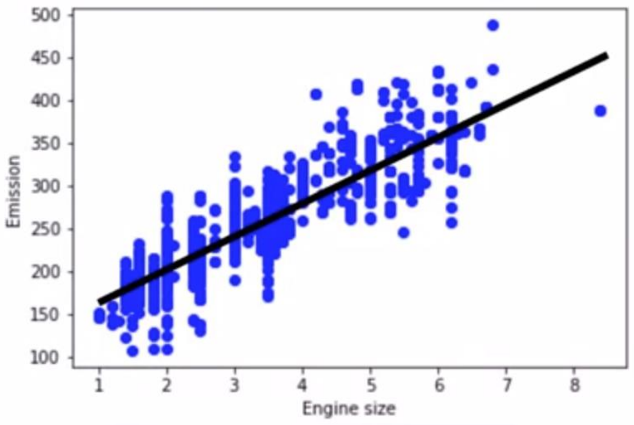
Predict CO2 emission vs EngineSize and Cylinders of all cars

Simple linear regression

X Y

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

$$Y = w_1 X + w_0$$



Simple linear regression

$x = 5.4 \;\; independent\ variable$

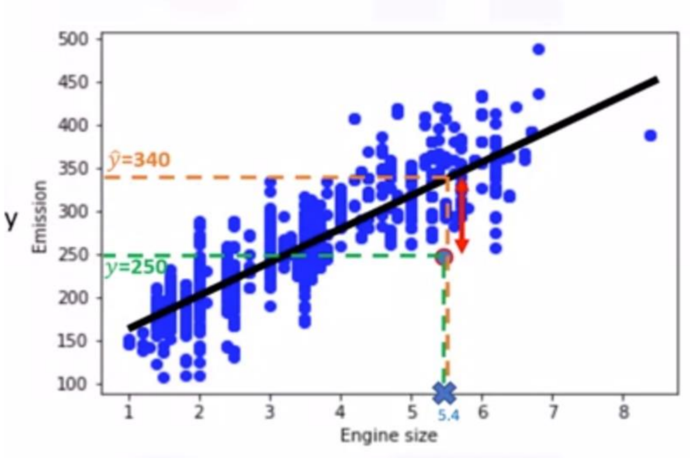
$y = 250 \; actual\ CO_2\ emission$

$\hat{y} = w_1 x + w_0$

$\hat{y} = 340 \;\; the\ predicted\ value\ for\ x_1$

$Error = y - \hat{y}$=250-340=-90

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2$$



Simple linear regression

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$$[x_1 \;\; 1]\begin{bmatrix} w_1 \\ w_0 \end{bmatrix} = y_1$$

$$X = \begin{bmatrix} 2 & 1 \\ 2.4 & 1 \\ 1.5 & 1 \\ 3.5 & 1 \\ 3.5 & 1 \\ 3.5 & 1 \\ 3.5 & 1 \\ 3.7 & 1 \\ 3.7 & 1 \end{bmatrix} \qquad W = \begin{bmatrix} w_1 \\ w_0 \end{bmatrix} \qquad Y = \begin{bmatrix} 196 \\ 221 \\ 136 \\ 255 \\ 244 \\ 230 \\ 232 \\ 255 \\ 267 \end{bmatrix}$$

$$X.W = Y \rightarrow (X^T X)W = X^T.Y \rightarrow W = (X^T X)^{-1} X^T Y$$

Multiple linear regression

X Y

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

$$Y = w_0 + w_1.ENGINESIZE + w_2.CYLINDERS + w_3 FUELCONSUMPTION_{COMB}$$

$$Y = w_0 + w_1.x_1 + w_2.x_2 + w_3.x_3$$

$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} \quad Y = W^T X$$

Multiple linear regression

	X			Y
	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} \quad Y = W^T X$$

$$X = \begin{bmatrix} 1 & 2 & 4 & 8.5 \\ 1 & 2.4 & 4 & 9.6 \\ 1 & 1.5 & 4 & 5.9 \\ 1 & 3.5 & 6 & 11.1 \\ 1 & 3.5 & 6 & 1.6 \\ 1 & 3.5 & 6 & 10 \\ 1 & 3.5 & 6 & 10.1 \\ 1 & 3.7 & 6 & 11.1 \\ 1 & 3.7 & 6 & 11.6 \\ 1 & 2.4 & 4 & 9.2 \end{bmatrix} \quad W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} \quad Y = \begin{bmatrix} 196 \\ 221 \\ 136 \\ 255 \\ 244 \\ 230 \\ 232 \\ 255 \\ 267 \end{bmatrix}$$

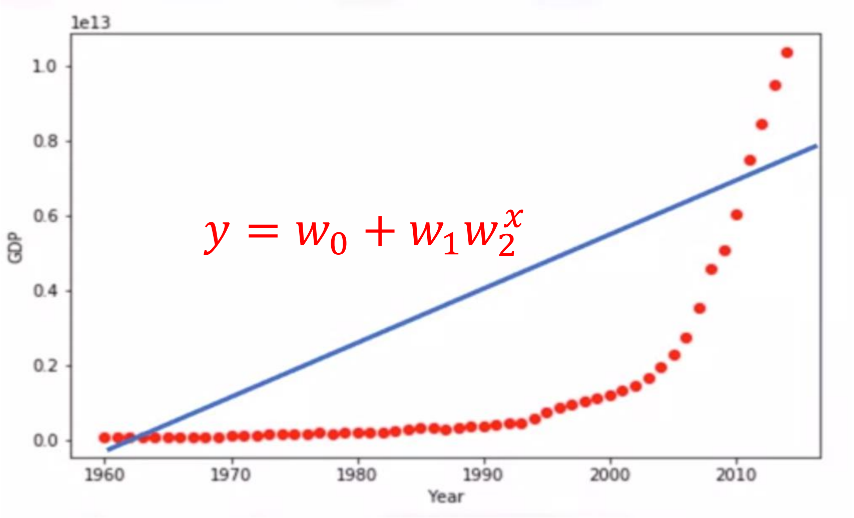
$$X.W = Y \rightarrow (X^T X)W = X^T.Y \rightarrow W = (X^T X)^{-1} X^T Y$$

Linear regression

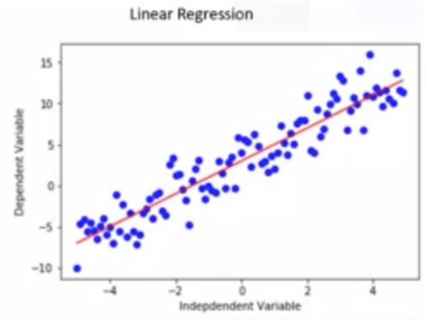
- How to estimate W?
  - Ordinary least square
    - ✓Linear algebra operation
    - ✓Takes a long time for large dataset (10K+row)
- Optimization algorithms

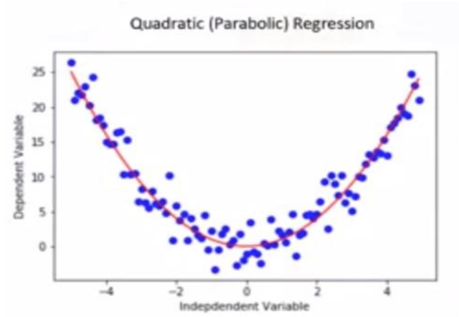
Non-linear regression

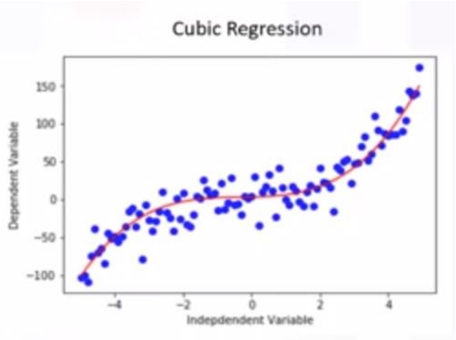
	Year	Value
0	1960	5.918412e+10
1	1961	4.955705e+10
2	1962	4.668518e+10
3	1963	5.009730e+10
4	1964	5.906225e+10
5	1965	6.970915e+10
6	1966	7.587943e+10
7	1967	7.205703e+10
8	1968	6.999350e+10
9	1969	7.871882e+10
...	.....	....

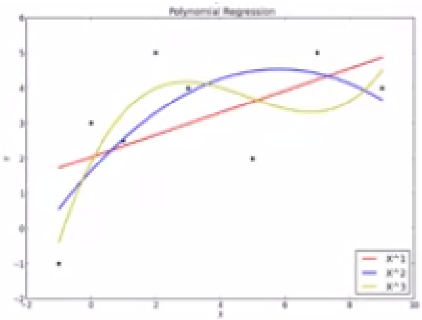


Non-linear regression



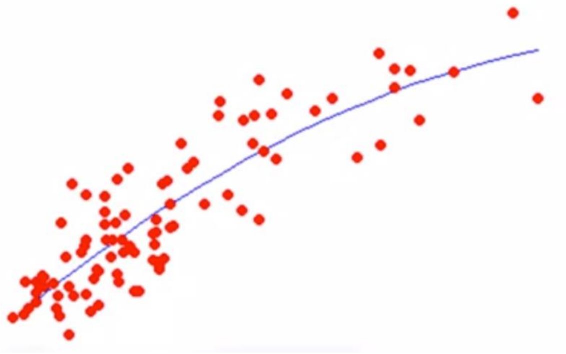






Non-linear regression

- What is polynomial regression?
  - Some data can be modeled by a polynomial regression
  - For example:
    - ✓ $y = w_3 x^3 + w_2 x^2 + w_1 x^1 + w_0$
  - A polynomial regression model can be transformed into linear regression model.
    - ✓ $x_1 = x, x_2 = x^2, x_3 = x^3$
      - $y = w_3 x_3 + w_2 x_2 + w_1 x_1 + w_0$



Non-linear regression

- Which sentence is not true about polynomial regression?
  - polynomial regression fits a curve line to your data.
  - Quadratic and cubic regression lines are a type polynomial regression
  - A polynomial regression model cannot be expressed as linear regression
  - Polynomial regression models can fit using the method of least squares.

Non-linear regression

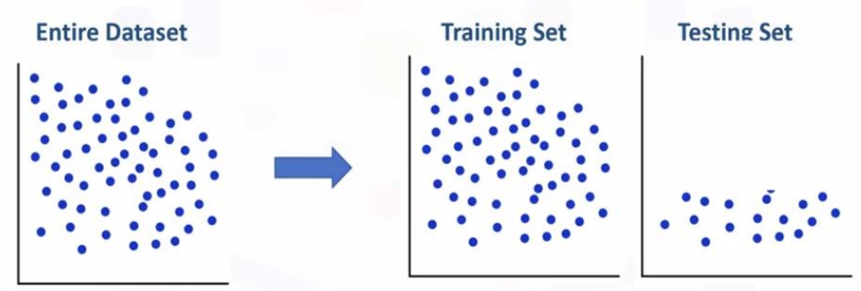
- What is non-linear regression?
  - To model non-linear relationship between the dependent variable and a set of independent variables.
  - $\hat{y}$ must be a non-linear function of the parameter w, not necessarily feature x.
    - ✓ $y = w_0 + w_1^2 x$
    - ✓ $y = w_0 + w_1 w_2^x$
    - ✓ $y = \log(w_0 + w_1 x + w_2 x^2)$
    - ✓ $y = \frac{w_0}{1+w_1^x}$

Linear vs non-linear regression

- How can I know if a problem is linear or non-linear in an easy way?
  - Inspect visually
  - Based on accuracy

Regression evaluation approaches

- Train and test on a same dataset
  - High training accuracy
  - Low out of sample accuracy



	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	212

Train

Test

What is training & out of sample accuracy?

- Train accuracy
  - Training accuracy isn’t a necessarily a good thing.
  - Result of over-fitting:
    - ✓ Over-fit: the model is overly trained to the dataset, which may capture noise and produce a non-generalized model
- Out of sample accuracy
  - It’s important that our model has a high, out of sample accuracy.
  - But, how can we improve out of sample accuracy?

Regression evaluation approaches

- Which of the sentences are not true?
  - Having a high training accuracy may result in an ‘over-fit’ of the data.
  - Doing a train and test on the same dataset will cause very high out-of-sample accuracy.
  - If a model is overly trained to the dataset, it may capture noise and produce a non-generalized model.

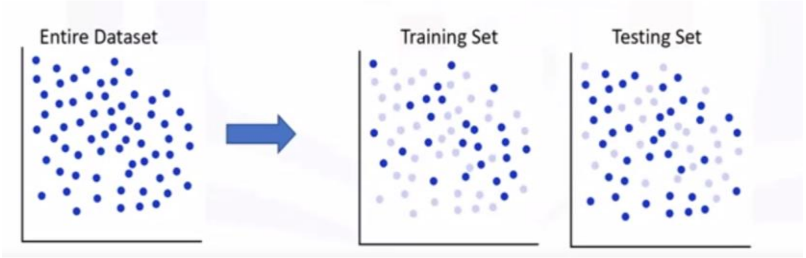
Regression evaluation approaches

- Train/test split
  - More exclusive
  - More accurate on out of sample accuracy
  - Highly dependent on which datasets the data is trained and tested

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	212

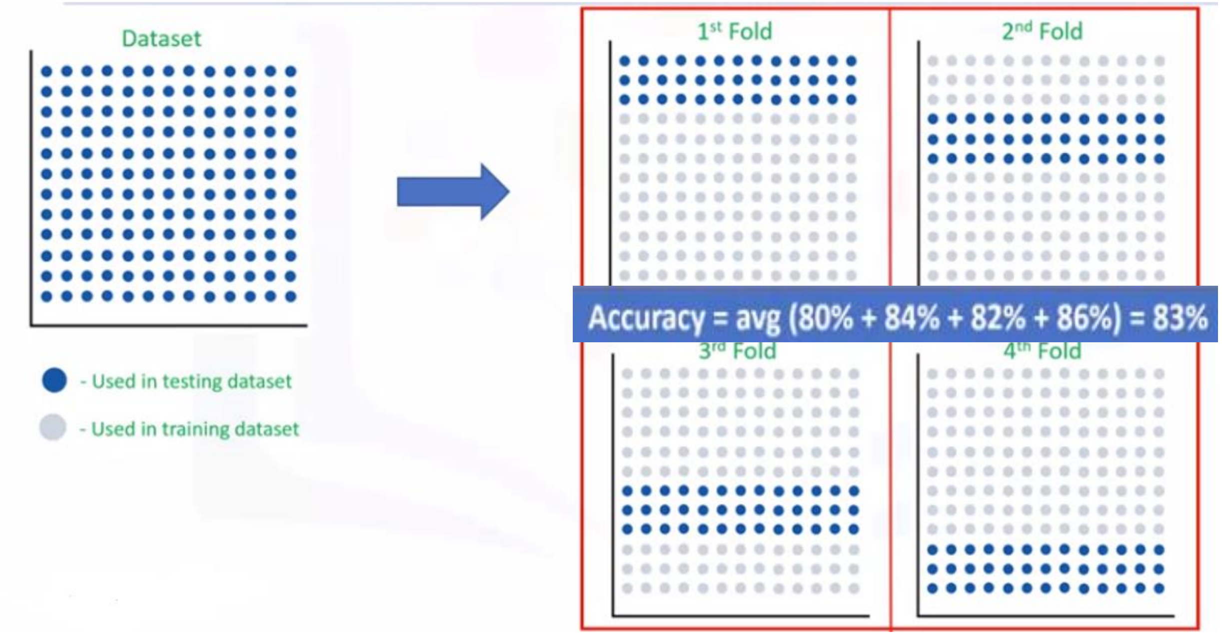
Train

Test

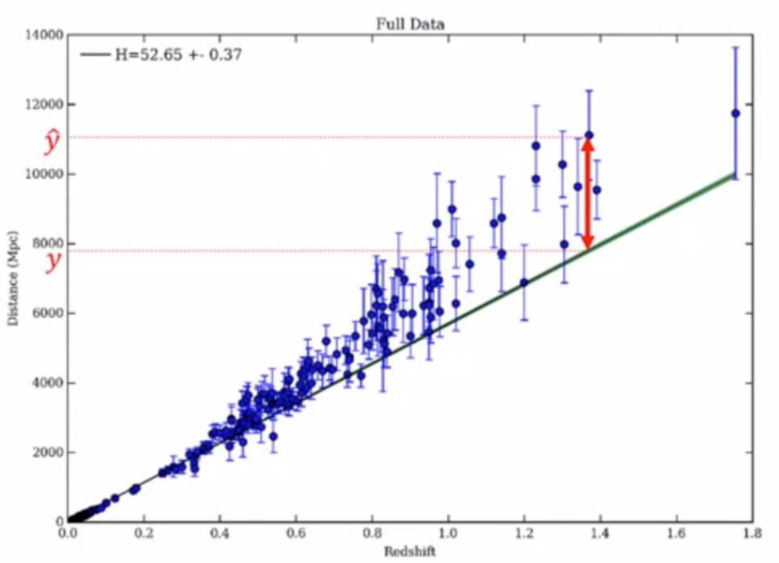


Regression evaluation approaches

- K-fold cross validation



Regression evaluation metrics



$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

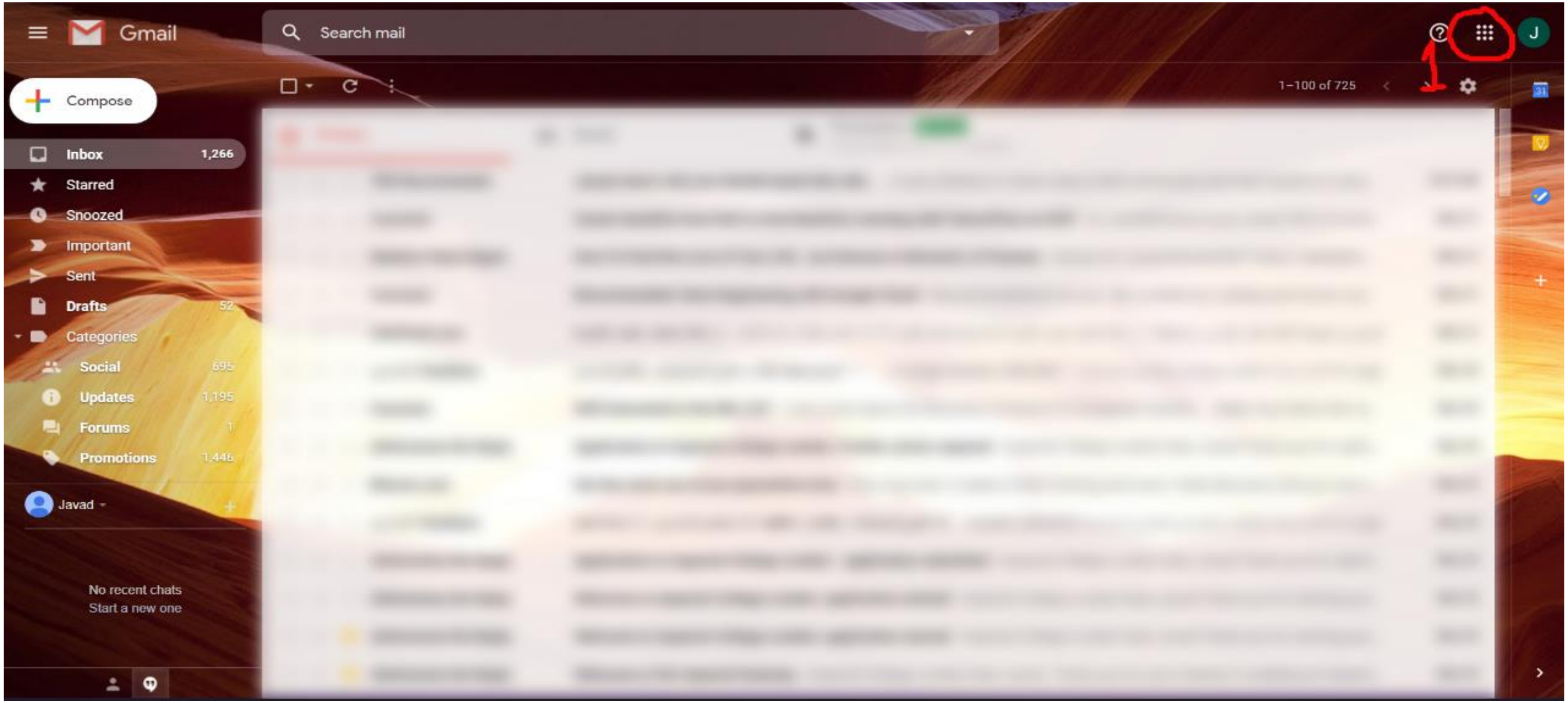
$$RAE = \frac{\sum_{j=1}^{n}|y_j - \hat{y}_j|}{\sum_{j=1}^{n}|y_j - \bar{y}|}$$

$$RSE = \frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2}$$

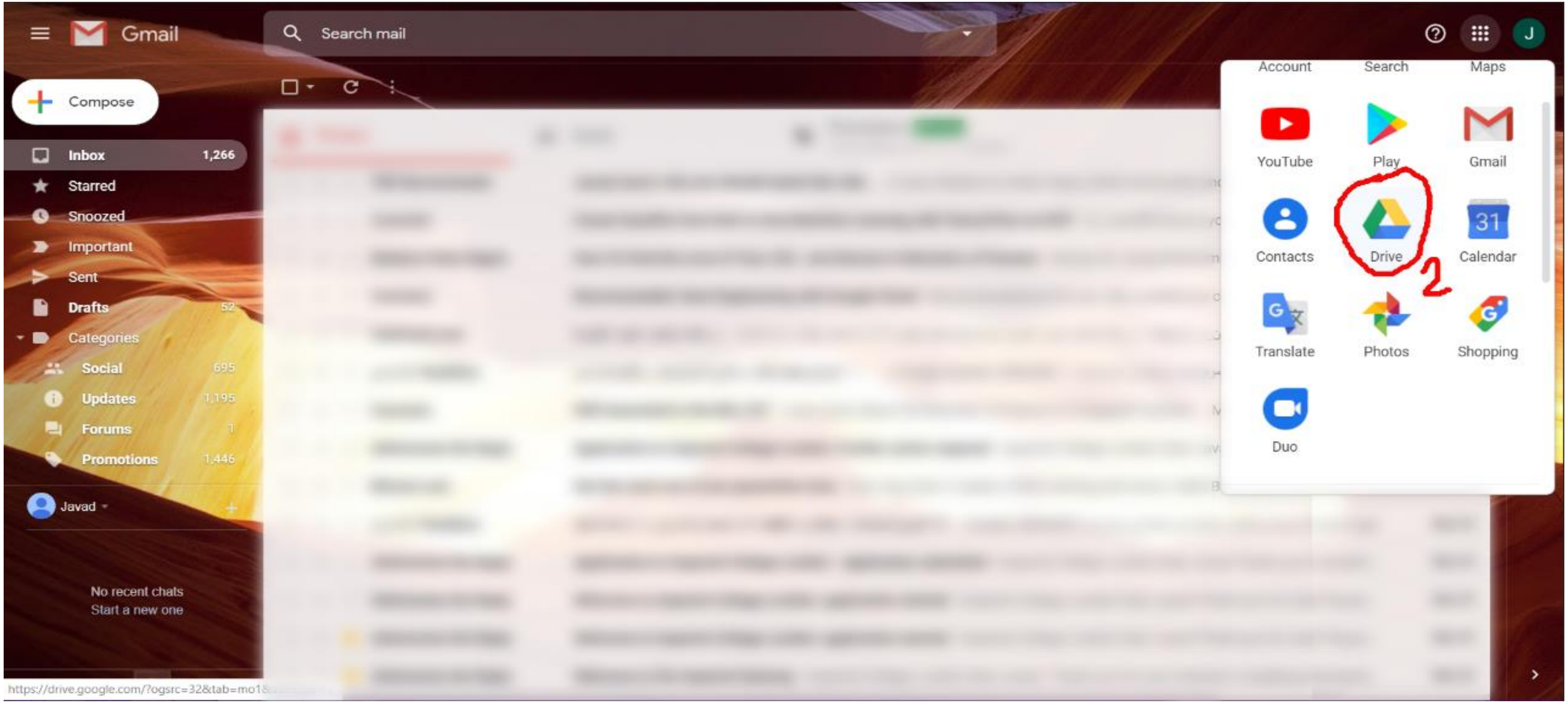
$$R^2 = 1 - RSE$$

Google Colaboratory

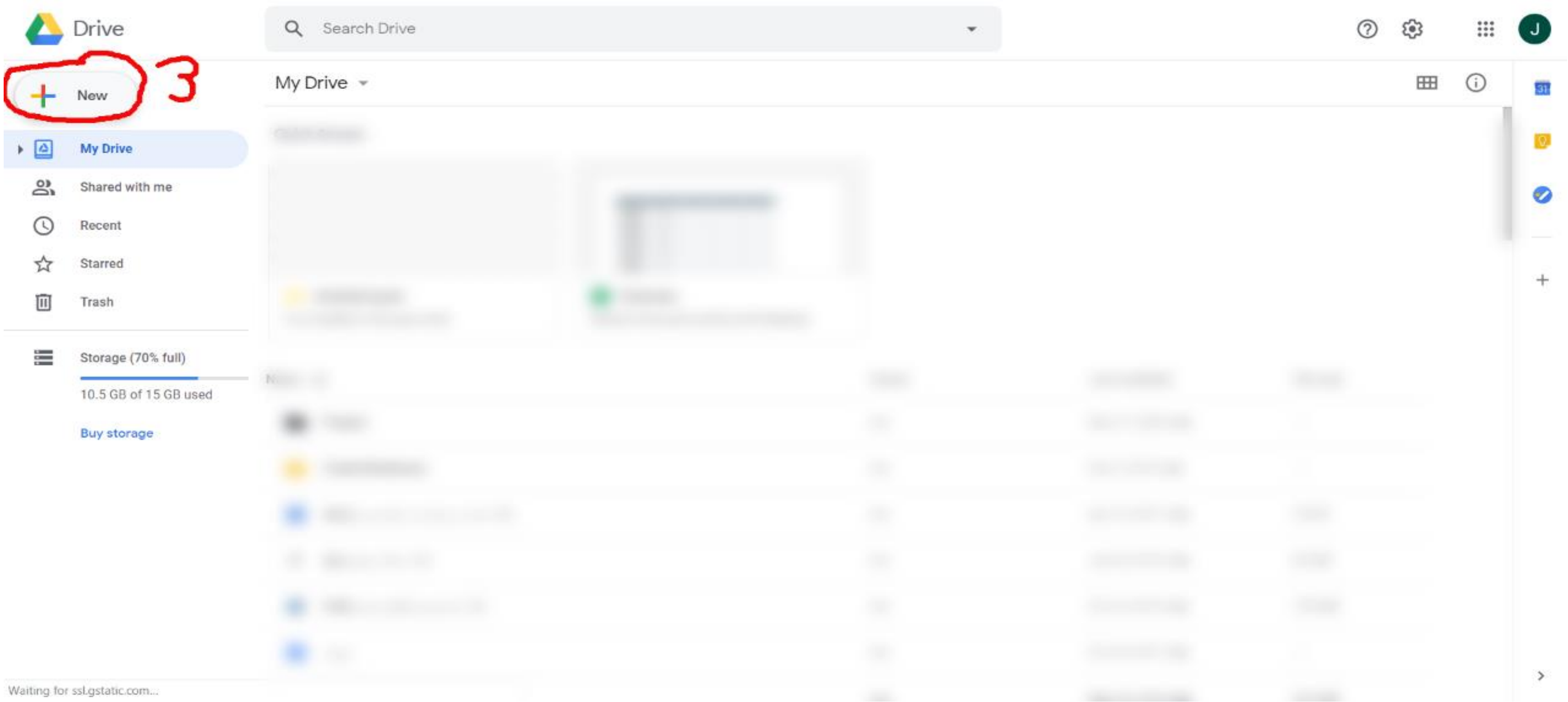
Google Colaboratory



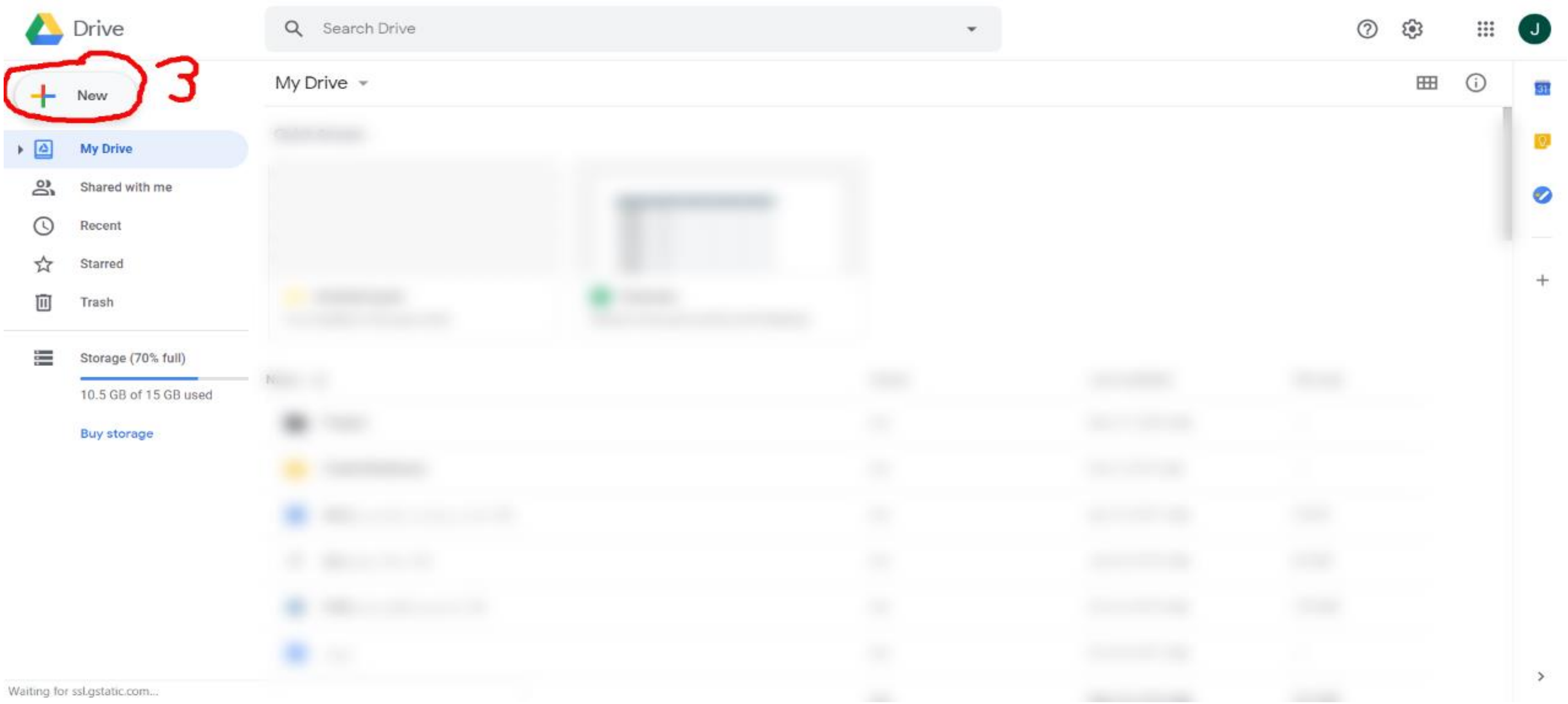
Google Colaboratory



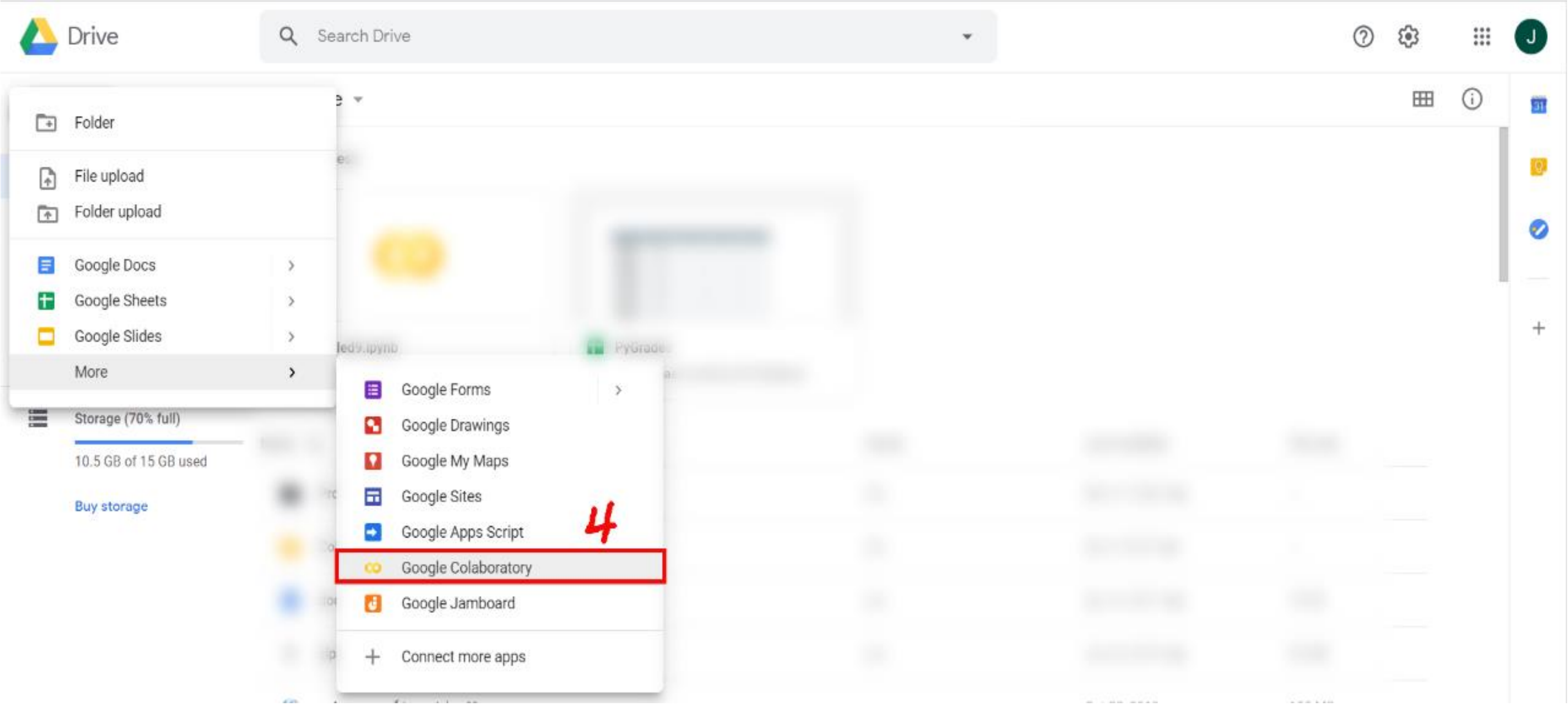
Google Colaboratory



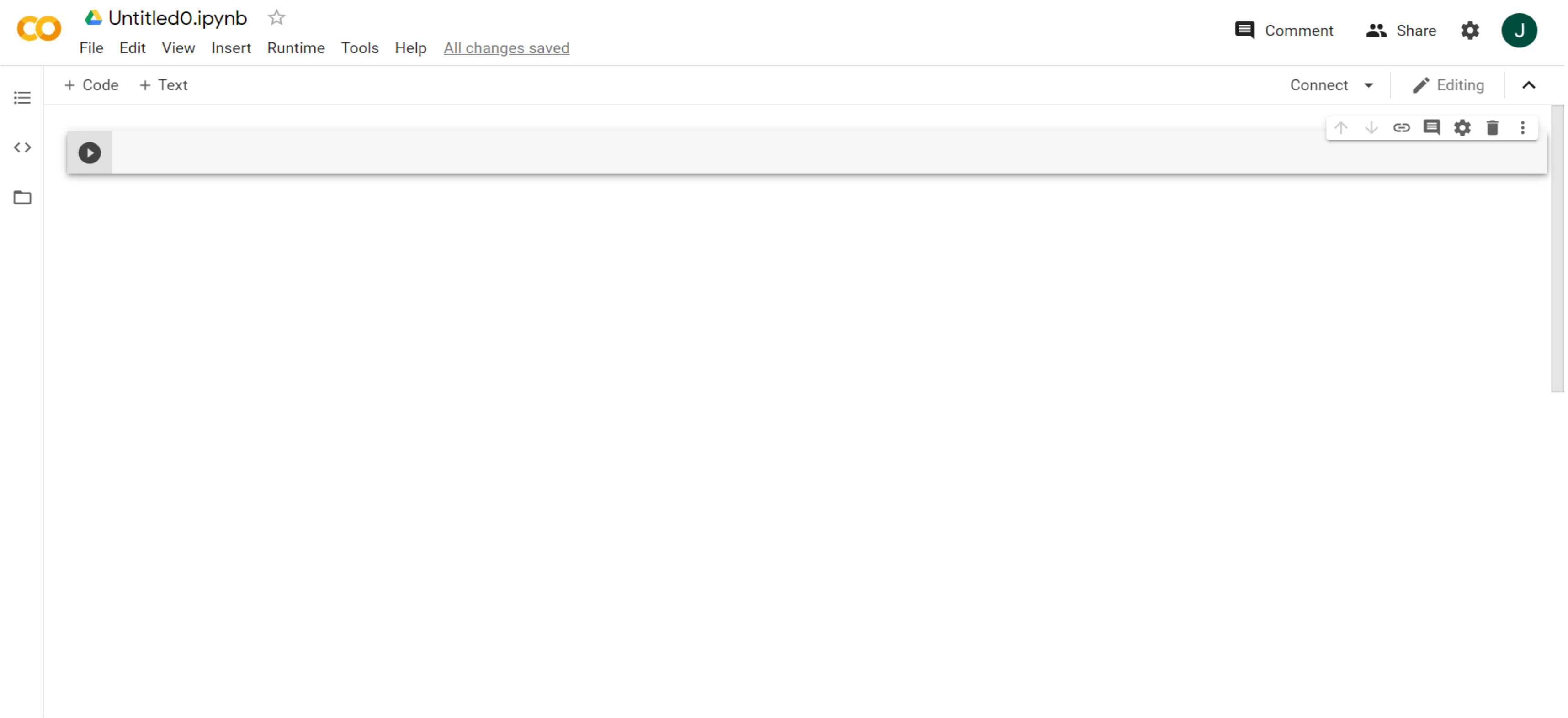
Google Colaboratory



Google Colaboratory



Google Colaboratory



- Assignment 1 has been uploaded on the website and due date is 1 week.
  - http:\\www.courses.kntu.ac.ir

Contact us

- Najafi.e@kntu.ac.ir
- J.khorramdel96@gmail.com