CSE 445/1-1/21.05.2025/

Absent

Attendance - 51.

Assignment - 209.

> 3 out of 4

i. ML Project base Code

ii. Project Proposal

III. VIVA on Project progress

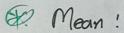
iv. Video demonstration of project.

Midtorm - 254.

Final - 254.

Project - 251.

1-2/26.05.2025/



- Anithmetit mean!

⇒ direct average of a data set, is found by adding all numbers in the data set and then dividing by the number of values in the set.

> Harmonic Mean:

- defined as the reciprocal of the average of the reciprocals of the data svalues.

(A) Median:

- is the middle value when a data set is ordered from least to greatest.

& Mode:

- is the number that occurs mest often in a data set.

& Conditional Probability:

- is a measure of the probability of an event occurring, given that another event 6 has already occurred.

Tom Mitchell!

- a computer program is said to learn from experience E with sa respect to some class of tasks T and performance measure P, if its performance at task in T, as measured by P, improves with experience E.

& Anthun Sumuel:

- Machine learning is the field of study that gives the computer the ability to learn without being explicitly programmed.

The difference between traditional programming and the ML approach! atte

Slide-9,10

ML Approach:

- allow fine tuning and make long list of rule.
- can adapt in changing environment.
- provide insight from large amount of data.
- solve complex problem.

(Deep learning!

- subset of ML
- automatic features entraction
- function like human brain.
- more data, better prediction.



AI > ML > DL

Narmow AZ:

- train for specific tack.

Panametric knowledge:

- learn from data.

L-3/28.05.2025/

& Supervised Learning:

- trained with labeled data.
 - > Linear Regression
 - Logistic Regnession
 - SVM
 - Decision Tree
 - NN

& Unsupervised Learning:

- data has no label.
- model try to make group based on familian features
- → Used in
 - clustering
 - anomaly detection
 - association mining
 - data prieprocessing
 - => k-mean ND
 - PCA ICA

Descrised Learning:

- partially labeled data
- model categoriaise the data based on familian feature and then asked for the label to human.

Reinforcement Learning:

- first provide a output, based on the feedback of the output, it try to re-learn its state for the data

- Batch Learning!
 - learning is not possible after deployment.
 - For new data, we need to train from screatch.

& Online Learning:

- continue learning after deployment as new obota comes.
 - use parallel computing, no down time. time.

Instance Based Leavening:

- memorize known data and try to match with these KNN
- mudel Based Learning 1
 - more generalize it divide the data arrea and build a model, then product the output.
- 3 V's of Big Data:
 - Volume: amount of data
 - Variety: data to for each possibility.
 - Velocity: how the data change over time

Overfitting:

- very good on training data
- Verry bad on test data.
- model need to be simple nespect to the data.
- need to tune the hyperparameters to control the negularization.

Data Split!

- Treaining
- Validation
- Test
- > Good mode: 60-20-20
- => an okay model: 70-15-15
- =) banely acceptable model 1 80-10-10

allowed unly when we have millions of intence in the dataset.

L-4/02.06.2025/

Problem with the housing data:

- price changes over time.
 - influencing features also change

Might be a good Mexanch arrea.

Clasification problem:

- predict the class of the data

Regnession problem:

- predict continuou varciable.
- > one value > univariate regression problem
 multiple value > multivariate regression problem.

- RME: Ruot mean squared error

- MAF: Mean absolute error

RMSE:

- straight line distance

- Euclidean norm on L2 norm

& MAE:

- city block distance

- Manhattan + norm on L1 norm

differences available

> Describe Code from Colab:

(1) Correlation between attributes:

> 1:

-if n goes up. y goes up too

- both one in same dinection

=) -1:

- be n and y in opposite direction.

⇒) ≈0:

- no relation between them.

L-5/04.06.2025/

@ Enplained Code from colab.

5 Steps:

- (i) Download the data
- (i) Buick look at Data Structure
 - plot histogram
- (11) Data viunalization:
 - seatten plut
 - ban plot
- (1) I reproces the data
 - nemove oudlier
 - nemore on manage me NaN value
 - + Manuform the data
 - use encoder -scaling
- Split the data : stratified sampling
- (i) Select model and train.
 - validate wing validation set
 - (vii) Test the model:
 -wing test set.

L-06/16.06.2025/

- 2 What is one-not encoder?
- From similar correlation, we need to take only one, as the effect is same.
- During encode, we need to maintain the order.
- Remove outlier before apply normalization.

 on use standardization.
- What the fit and fit transform do in the custom transformer method?

L-07/18.062025/

(2) Confusion Matrix:

TP => True Positive => Actual true & Predicted True.

TN => True Negative => Actual Negative & Predicted Negative

FP => Talse Positive => Actual Negative & Predicted Builtive

FN => Talse Negative => Actual Positive & Predicted Negative

FN => Talse Negative => Actual Positive & Predicted Negative

Precision & Reent