Computational Statistics and NLP Lab

Supervised By: Abu Omayed

Contact (what's app): 01813131900

Email: abuomayed@gmail.com

LinkedIn: abu_omayed

GitHub: https://github.com/AbuOmayed

Class time: Fryday and Monday at 11P.M

The Class Roadmap contains Basic Python, Machine Learning, Deep learning and Natural Language Processing.

Python and Data Analysis

Python fundamental part 1:

a. Python data structure
b. Python Variable
c. Python data type
d. list
e. Faction

Python fundamental's part 2:

1. Operator:
   i) Arithmetic operator
   ii) Relational operator
   iii) Logical operator
2. Conditional Control Statement
   i) IF statement
   ii) IF….else statement

Advanced Python:

1) Loop control statement

i) For loop
ii) While loop

2) User define function
3) Lambda function

Introduction to NumPy

i) What is NumPy?
ii) NumPy List
iii) Basic statistics in NumPy (Mean, median, mode etc)
iv) Matrix in NumPy

Statistics in Python

i) What is statistics?
ii) What can statistics do?
iii) Types of statistics?
iv) Types of data?
v) Measure of Centre.
vi) Measure of spread.
vii) Standard deviation.
viii) Mean absolute deviation.
ix) Quartiles.
x) IQR (Inter Quartile range)
xi) Statistical graph.

Introduction to Pandas

i) What is Pandas?
ii) Relation and defences between pandas and Microsoft excel.
iii) Create a data frame.
iv) Import a csv file.
v) Pandas basic function.
vi) Summary statistics.
vii) Visualizing data frame.
(1) Histogram
(2) Bar plot
(3) Line plot
(4) Scatter plot

Intermediate pandas:

i. Aggregate function
ii. Group By function
iii. Pivot Table
iv. Explicit Indexes
v. loc and iloc function.
vi. Joining data with pandas:
   a) Inner join
   b) Left join
   c) Right join
   d) Outer join
   e) Concatenates.
   f) Melt

Data Visualization with matplotlib:

- What is matplotlib?
- Introducing the Pyplot interface.
- Customizing plot.
- Histogram, Bar plot, scatter plot in matplotlib.
- Adding title and labels.
- Quantitative comparison
- Plotting time series data.
- Save figure in different file format.

Data Visualization with seaborn

- What is seaborn?
- Why is seaborn useful?
- Basic graph in seaborn. (Histogram, bar plot, scatter plot)
- Using pandas with seaborn.
- Adding third variable with Hue function.
- Different types of plots
- Changing plot style and colour.
- Adding titles and labels.

Explanatory data analysis (EDA)

i) What is EDA?
ii) Data Validation.

iii) Data Summarization.
iv) Missing data Handling.
v) Converting and analysing categorical data.
vi) Converting string to number.
vii) Handling outliers.
viii) Patters over time.
ix) Correlation by graph.
x) Factor relationships and distributions.

Working with categorical data

i) What is categorical data?
ii) Categorical data in pandas.
iii) Grouping data by categories in pandas.
iv) Setting categorical variables.
v) Updating categories.
vi) Cleaning and accessing data.
vii) Visualizing categorical data.

Statistics

Probability & Statistics

- Random variables, expectation, variance
- Distributions (Normal, Binomial, Poisson)
- Bayes theorem (critical for Naive Bayes)
- Hypothesis testing

Machine Learning Algorithms

Supervised Learning

- Linear regression
- Logistic regression
- KNN
- Decision Tree
- Random Forest
- Naive Bayes
- SVM + Kernels
- Gradient Boosting (XGBoost, LightGBM, CatBoost)

Unsupervised Learning

- K-means
- Hierarchical clustering

- PCA
- ICA
- LDA

Deep Learning

Neural Network

- What is a neuron?
- Perceptron
- Single-layer NN
- Multi-layer NN
- Backpropagation implementation

Convolutional Neural Networks (CNN)

- Convolution operation
- Filters, kernels
- Padding, stride
- Pooling
- Feature maps
- Flattening
- Fully Connected layers

Recurrent Neural Networks (RNN)

- RNN architecture
- Vanishing/exploding gradient
- LSTM
- GRU

Natural Language Processing (NLP)

Text Preprocessing Essentials

- Lowercasing
- Tokenization
- Stop word removal
- Lemmatization
- Stemming
- N-grams
- TF-IDF representation

Word Embeddings

- Word2Vec (CBOW, Skip-gram)
- GloVe
- FastText
- Distributional semantics

Attention Mechanisms

- Global attention
- Local attention
- Self-attention
- Key-query-value mechanism

Transformer Architecture

paper: Attention is All You Need

- Multi-head attention
- Positional encoding
- Encoder and decoder
- Feed-forward network
- Residual connections
- Layer normalization

Encoder-only models

- BERT
- RoBERTa
- ALBERT

Decoder-only models (LLMs)

- GPT series
- LLaMA
- DeepSeek
- Phi

Encoder-decoder (Seq2Seq)

- T5
- BART
- mT5
- Whisper (speech)

Large Language Model

Multi modal model development