

Digital Skills Training for Student

Machine Learning with Python

1. About Course

Machine Learning is broad and fast-growing sub-field of Artificial Intelligence. This course introduces students to the basic concepts and techniques of Machine Learning. And this course also contains Basic of Python Programming to contain control structure, conditional statement, function Sequence Data type and numpy that make students to more skilled.

2. Objective of Course

Machine Learning is broad and fast-growing sub-field of Artificial Intelligence. This course introduces students to the basic concepts and techniques of Machine Learning. The objective of this course is to develop the skills required for Machine Learning Technologies with use of Python to analyze data and solving ML problems like Regression and Classification using machine learning algorithms.

After completing the module, the learner will be able to:

- Understand the basic concepts of Python language.
- Understand the basics of Machine Learning & their types.
- Understand various learning models, methods and applications under supervised and unsupervised learning.
- Understand data preprocessing for Machine Learning.
- Solve real world problems through machine learning implementation leading to predictions.

3. Job Roles of Course

After successful completion of the qualification the candidates shall be employed in the industries for following occupations:

- Junior Machine Learning Engineer
- Machine Learning Developer
- Machine Learning Researcher

Basics of Python Programming

1. Session 1: Python Basics

- About Python
- Python Output/print function
- Python Data Types
- Python Variables
- Python comments
- Python Keywords and Identifiers
- Python User Input
- Python Type conversion
- Python Literals

2. Session 2: Python Operators + if-else + Loops

- Start of the session
- Python Operators
- Python if-else
- Python Modules
- Python While Loop
- Python for loop

3. Session 3: Python Strings

- Introduction
- Break, continue, pass statement in loops
- Strings
- String indexing
- String slicing
- Edit and delete a string
- Operations on String
- Common String functions

Python Data Types

1. Session 4: Python Lists

- Introduction
- Array vs List
- How lists are stored in a memory
- Characteristics of Python List

- Code Example of Lists
 - ❖ *Create and access a list*
 - ❖ *append(), extend(), insert()*
 - ❖ *Edit items in a list*
 - ❖ *Deleting items from a list*
 - ❖ *Arithmetic, membership and loop operations on a List*
 - ❖ *Various List functions*
 - ❖ *List comprehension*
 - ❖ *2 Ways to traverse a list*
 - ❖ *Zip() function*
 - ❖ *Python List can store any kind of objects*
- Disadvantages of Python list

2. Session 5: Tuples + Set + Dictionary

- **Tuple**
 - ❖ *Create and access a tuple*
 - ❖ *Can we edit and add items to a tuple?*
 - ❖ *Deletion*
 - ❖ *Operations on tuple*
 - ❖ *Tuple functions*
 - ❖ *List vs tuple*
 - ❖ *Tuple unpacking*
 - ❖ *Zip () on tuple*
- **Set**
 - ❖ *Create and access a set*
 - ❖ *Can we edit and add items to a set?*
 - ❖ *Deletion*
 - ❖ *Operations on set*
 - ❖ *set functions*
 - ❖ *Frozen set (immutable set)*
 - ❖ *Set comprehension*
- **Dictionary**
 - ❖ *Create dictionary*
 - ❖ *Accessing items*
 - ❖ *Add, remove, edit key-value pairs*
 - ❖ *Operations on dictionary*
 - ❖ *Dictionary functions*

- ❖ *Dictionary comprehension*
- ❖ *Zip() on dictionary*
- ❖ *Nested comprehension*

3. Session 6: Python Functions

- Create function
- Arguments and parameters
- args and kwargs
- How to access documentation of a function
- Variable scope
- Nested functions with examples
- Returning of function
- Advantages of functions
- Lambda functions

Numpy

1. Session 7: Numpy Fundamentals

- Numpy Theory
- Numpy array
- Matrix in numpy
- Numpy array attributes
- Array operations
- Scalar and Vector operations
- Numpy array functions
 - ❖ *Dot product*
 - ❖ *Log, exp, mean, median, std, prod, min, max, trigo, variance, ceil, floor, slicing, iteration*
 - ❖ *Reshaping*

Pandas

1. Session 8: Pandas Series

- What is Pandas?
- Introduction to Pandas Series
- Series Methods
- Series with Python functionalities
- Boolean Indexing on Series
- Plotting graphs on series

2. Session 9: Pandas DataFrame

- Introduction Pandas DataFrame
- Creating DataFrame and read_csv()
- DataFrame attributes and methods
- Dataframe Math Methods
- Selecting cols and rows from dataframe
- Filtering a Dataframe
- Adding new columns
- Dataframe function – astype()

Data Visualization

1) Session 10: Plotting Using Matplotlib

- Get started with Matplotlib
- Plotting simple functions, labels, legends, multiple plots
- About scatter plots
- Bar chart
- Histogram
- Pie chart
- Changing styles of plots

Linear Regression

1) Session 11: Introduction to Machine Learning

- About Machine Learning (History and Definition)
- Types of ML
 - ❖ *Supervised Machine Learning*
 - ❖ *Unsupervised Machine Learning*
 - ❖ *Semi supervised Machine Learning*
 - ❖ *Reinforcement Learning*
- Batch/Offline Machine Learning
- Disadvantages of Batch learning
- Online Machine Learning
 - ❖ Importance
 - ❖ When to use and how to use
 - ❖ Learning Rate
 - ❖ Out of core learning
 - ❖ Disadvantages

- Batch vs Online learning
- Instance based learning
- model-based learning
- Instance vs model-based learning
- Challenges in ML
 - ❖ Data collection
 - ❖ Insufficient/Labelled data
 - ❖ Non-representative data
 - ❖ Poor quality data
 - ❖ Irrelevant features
 - ❖ Overfitting and Underfitting
 - ❖ Offline learning
 - ❖ Cost
- Machine Learning Development Life-cycle

2) Session 12: Simple Linear regression

- Introduction and Types of Linear Regression
- Simple Linear Regression
- Intuition of simple linear regression
- Code example
- How to find m and b?
- Simple Linear Regression model code from scratch
- Regression Metrics
 - ❖ *MAE*
 - ❖ *MSE*
 - ❖ *RMSE*
 - ❖ *R2 score*
 - ❖ *Adjusted R2 score*

3) Session 13: Multiple Linear Regression

- Introduction to Multiple Linear Regression (MLR)
- Code of MLR
- Mathematical Formulation of MLR
- Error function of MLR
- Minimizing error
- Error function continued
- Code from scratch

Feature Selection

1) Session 14: feature Selection Part 1

- What is Feature Selection?
- Why to do Feature Selection?
- Types of Feature Selection
- Filter based Feature Selection
 - ❖ *Duplicate Features*
 - ❖ *Variance Threshold*
 - ❖ *Correlation*
 - ❖ *ANOVA*
 - ❖ *Chi-Square*
 - ❖ *Advantages and Disadvantages*

2) Session 15: Feature Selection Part 2

- Wrapper method
- Types of wrapper method
- Exhaustive Feature Selection/Best Subset Selection
- Sequential Backward Selection/Elimination
- Sequential Forward Selection
- Advantages and Disadvantages

3) Session on Feature Selection part 3

- Recursive Feature Elimination
- Advantages and Disadvantages

4) PCA

- Dimensionality reduction
- Scikit Learn's PCA class
- PCA variance Ratio
- Randomized PCA
- Incremental PCA
- PCA on MNIST dataset
- Code with example

Regularization

1) Session 16: Regularization Part 1 | Bias-Variance Tradeoff

- Why we need to study Bias and Variance
- Expected Value and Variance
- Bias and Variance Mathematically

2) Session on Regularization Part 1 | What is Regularization

- Bias Variance Decomposition
- Diagram
- Analogy
- Code Example
- What is Regularization?
- When to use Regularization?

3) Ridge Regression Part 1

- Types of Regularization
- Geometric Intuition
- Sklearn Implementation

4) Session 16 : Lasso Regression

- Intuition
- Code example
- Lasso regression key points

5) Why Lasso Regression creates Sparsity?

6) ElasticNet Regression

- Intuition
- Code example

K Nearest Neighbors

1) Session 17: K nearest Neighbors Part 1

- KNN intuition
- Code Example
- How to select K?
- Decision Surface
- Overfitting and Underfitting in KNN
- Limitations of KNN

2) Classification Metrics Part 1

- Accuracy
- Accuracy of multi-classification problems
- How much accuracy is good?
- Problem with accuracy
- Confusion matrix
- Confusion matrix of multi-classification problems
- When accuracy is misleading

3) Classification Metrics Part 2

- Precision
- Recall
- F1 score
- Multi class Precision and Recall
- Multi class F1 score

Model Evaluation & Selection

1. Session 18: ROC Curve in Machine Learning

- ROC AUC Curve and its requirements
- Confusion matrix
- True Positive Rate (TPR)
- False Positive Rate (FPR)
- Different cases of TPR & FPR

2. Cross Validation

- Why do we need Cross Validation?
- Cross Validation
 - ❖ Leave One Out Cross Validation (LOOCV)
 1. Advantages
 2. Disadvantages
 3. When to use
 - ❖ K-Fold Cross Validation
 4. Advantages
 5. Disadvantages
 6. When to use
 - ❖ Stratified K-Fold CV

3. Hyperparameter Tuning

- Parameter vs Hyperparameter
- Why the word "hyper" in the term
- Requirements
 - ❖ Grid Search CV
 - ❖ Randomized Search CV
- Can this be improved?

Naive Bayse

1. Session 19: Crash course on Probability Part 1

- 5 important terms in Probability
 - ❖ Random Experiment
 - ❖ Trials
 - ❖ Outcome
 - ❖ Sample Space
 - ❖ Event
- Some examples of these terms
- Types of events
- What is probability
- Empirical vs Theoretical probability
- Random variable
- Probability distribution of random variable
- Mean of 2 random variable
- Variance of Random variable

2. Crash course on Probability Part 2

- Venn diagrams
- Contingency table
- Joint probability
- Marginal probability
- Conditional probability
- Intuition of Conditional Probability
- Independent vs Dependent vs Mutually Exclusive Events
- Bayes Theorem

3. Session 20: Naive Bayes

- Intuition
- Mathematical formulation
- How Naive Bayes handles numerical data
- What if data is not Gaussian
- Naive Bayse on Textual data

Logistics Regression

1. Session 21: Logistic Regression

- Introduction
- Some Basic Geometry
- Classification Problem
 - ❖ Basic Algorithm
 - ❖ Updation in Basic Algorithm
- Sigmoid Function
- Maximum Likelihood
- Log Loss
- Gradient Descent
- Summary

Decision Trees

1. Session 22: Decision Tree

- Introduction
- Intuition behind DT
- Terminology in Decision Tree
- The CART Algorithm - Classification
- Splitting Categorical Features
- Splitting Numerical Features
- Understanding Gini Impurity?
- Geometric Intuition of DT

Ensemble Methods – Introduction

1. Session 23: Ensemble Learning

- Intuition
- Types of Ensemble Learning
- Why it works?
- Benefits of Ensemble
- When to use Ensemble

2. Bagging Part 1 – Introduction

- Core Idea
- Why use Bagging?
- When to use Bagging?

- Code Demo
- 3. Bagging Part 2 – Classifier**
 - Intuition through Demo app
 - Code Demo
- 4. Bagging Part 3 – Regressor**
 - Core Idea
 - Code Demo
- 5. Session 24: Random Forest**
 - Introduction to Random Forest
 - Bagging
 - Random Forest Intuition
 - Why Random Forest Works?
 - Bagging vs. Random Forest
 - Feature Importance

Gradient Boosting and XGBoost

- 1. Session 25: Gradient Boosting**
 - Boosting
 - What is Gradient Boosting
 - How
 - What
 - Why
- 2. Gradient Boosting**
 - How Gradient Boosting works?
 - Intuition of Gradient Boosting
 - Function Space vs. Parameter Space
 - Direction of Loss Minimization
 - How to update the function
 - Iterate
 - Another perspective of Gradient Boosting
 - Difference between Gradient Boosting and Gradient Descent

Introduction to XGBoost

- Introduction
- Features
 - ❖ Performance
 - ❖ Speed

❖ Flexibility

XGBoost for Classification

- Classification Problem Statement
- Step-by-Step Mathematical Calculation

Session 26 : DNN:

- MLP and Backpropagation
- Regression MLP
- Implementing MLP with Keras
- Fine tuning NN hyperparameters
- Activation function
- Batch normalization
- Monte Carlo dropout
- Tensor Flow's API
- Dataset with Keras
- Deep Computer Vision using CNN

Session 27: Evaluation
