

Mirco lecture Glossary

This document provides a glossary of topics covered in the MachineLearnAthon micro-lectures. For each topic, it specifies the corresponding covered micro-lecture and slide numbers where further information can be found. All entries are arranged alphabetically to ensure easy navigation and quick reference for learners and educators.

Topic	Micro Lecture	Slide
Local Interpretable Model-Agnostic Explanation (LIME)	Interpretability part 4	5
"Best" straight line with given values	Simple and multivariate linear regression	4
(S)ARIMA	Forecasting Methods	4
Accumulated Local Effects (ALE) Plot	Interpretability part 3	5-7
Advanced Widgets in Orange	No Code	12
Advantages	LightGBM	4
Analysis of the impact of the size and number of rooms on the price of the apartment	Simple and multivariate linear regression	12, 13
Application example	Random Forest	7
Applications of neural networks for regression	Neural Networks	63-68
Approaches to explaining AI	Interpretability part 1	15
Architecture and hyperparameters	Neural Networks	52-61
Assessment of the model	Regression Analysis Guide	9, 10
Assumptions of multivariate regression	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	2, 4, 5, 6, 7
Barplots	Data visualization	6
Basic images features	Feature Extraction	4
Bias in Machine Learning	Fairness in Machine Learning	6
Bias Mitigation Strategies	Fairness in Machine Learning	7
Binning	Data Preprocessing	22
Branches and merging	Introduction to GIT	8
Calculating Splits and Growing the Tree	XGBoost	7
Can I trust the Model based on Accuracy?	Interpretability part 1	9,10, 11
Case study	LLM & Ethics	39, 40
Characteristics of the explanation provided by LIME	Interpretability part 4	6
Class weighting	Imbalanced Dataset	9
Classification metrics - Accuracy	Evaluation Metrics	5-7
Classification metrics - Precision, Recall	Evaluation Metrics	8, 9

Clustering	Clustering	6
clustering applications	Clustering	8
clustering in customer analysis	Clustering	9
Clustering in Image Processing	Clustering	11
clustering in medical diagnostics	Clustering	10
Clustering in Text Analysis	Clustering	12
Color features	Feature Extraction	5-7
Combining datasets	Data Preparation (for tabular data)	11-12
Comparison of the Models	Low Code 1	12
Comparison with other Low code /No code tools	No Code	13
Comparison with other methods	Random Forest	9
Connection to a database	Data Preprocessing	6
Core Libraries in Orange	No Code	13
Correlation analysis	Data Preprocessing	18
Create the Model	Low Code 1	13
Creating a multivariate linear regression model	Simple and multivariate linear regression	9
Creating Random Sample	Low Code 1	8
Cross Validation	Time Series Forecasting	9
Data Import and Selection	Data Preprocessing	4
Data Imputation	Data Preparation (for tabular data)	6
Data Loading and Preprocessing	No Code	6
Data normalization	Data Preparation (for tabular data)	13-14
Data Preparation	Data Preprocessing	10
Data Preprocessing Workflow	Data Preprocessing	3
Data Selection and cleaning	Data Preprocessing	9
Data transformation	Data Preprocessing	12
Data type conversion	Data Preprocessing	11
Data types	Data Preparation (for tabular data)	7-9
Data visualization	Data visualization	4
Decision Trees	Random Forest	4
Decision Trees as Base Learners	XGBoost	3
Deep Learning: When to use ML? When to use DL?	Classification#2_Types_of_Classification_Models	19
Definition of large language models	LLM & Ethics	8
Definition of natural language models	LLM & Ethics	6
Development Environments	Installing Python	6
Differences(Time series forecasting vs Regression)	Time Series Forecasting	8
Different data splits	Data Preprocessing	24
Dimensionality reduction	Feature Extraction	17

Distance measures	Clustering	13-16
Distribution plots - Boxplots	Data visualization	8-9
Drawbacks of lime	Interpretability part 4	13
Dummy Variable Added	Regression Analysis Guide	8
Dummy variables	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	10, 11, 12, 14
Edge detection	Feature Extraction	11, 12
Elements in cv algorithms	Computer Vision	65-74
Embedded algorithms	Feature Engineering	14
Embeddings	LLM & Ethics	16, 17
Ensemble Model	Forecasting Methods	8
Estimation of multivariate linear regression model	Simple and multivariate linear regression	10, 11
Ethical and Practical challenges	Fairness in Machine Learning	8
EU Definition of trustworthy AI	LLM & Ethics	38
Evaluation Interpretability	Interpretability part 1	13
Evaluation metrics	Imbalanced Dataset	6
Evaluation metrics in machine learning	Evaluation Metrics	4
Evaluation the Model	Low Code 2	8
Example of Clasification Models	Low Code 1	6
Exclusive Feature Bundling	LightGBM	7
Exploratory data analysis	Data Preprocessing	16
Extracting statistical features	Feature Extraction	16
Fairness in Machine Learning	Fairness in Machine Learning	4, 5
Feature Extraction from images	Feature Extraction	3
Feature extraction from tabular data	Feature Extraction	15
Feature Extraction from text data	Feature Extraction	18
Feature Extraction using convolution	Feature Extraction	13
Feature Importance	Interpretability part 3	8-10
Feature Selection	Feature Engineering	5,6
Feature Selection types	Feature Engineering	7,8
Fidelity-Interpretability Trade-Off	Interpretability part 4	7
Filter	Computer Vision	21 - 40
Filtering	Feature Extraction	9,10
Finalizing the Model	Low Code 2	9
Fisher Score	Feature Engineering	9,10
Forecast Horizons	Time Series Forecasting	6
Forecast of Y values	Regression Analysis Guide	13

From language to vectors	LLM & Ethics	13
Functionality of Random Forest	Random Forest	6
Generalization	Overfitting	4
Get the Data	Low Code 1	7
GitHub	Introduction to GIT	10
Global machine learning interpretability techniques	Interpretability part 2	8
Global Surrogate Models	Interpretability part 3	11-13
Google Colab	Installing Python	8
Google Collab and GIT	Introduction to GIT	11
Gradient Boosting	Forecasting Methods	9
Gradient Boosting Classifier	Low Code 1	16
Gradient-based One-Side Sampling	LightGBM	9-17
Graph of multivariate linear regression model	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	3
Greedy Bundling	LightGBM	8
Grid Search	Hyperparameter Tuning	6
Hard vs soft clustering	Clustering	7
Heatmaps	Data visualization	10
Histogram-based Algorithm	LightGBM	18
How can LLMs handle textual data/inputs?	LLM & Ethics	12
How to choose your models?	Classification#2_Types_of_Classification_Models	22
How to evaluate your classification model?	Classification#3_Evaluating_Classification_Models	5, 6, 7, 8, 9
How to measure Calibration? Reliability Diagram	Classification#3_Evaluating_Classification_Models	16-23
Hyperband Optimization	Hyperparameter Tuning	8
Hyperparameter Tuning	Hyperparameter Tuning	5
Hyperparameters	Hyperparameter Tuning	4
Impact of LLMs	LLM & Ethics	5
Important concepts in computer vision	Computer Vision	20
Important Outputs of setup()	Low Code 1	11
Importing data from a flat file	Data Preprocessing	5
Individual conditional expectation	Interpretability part 2	13,14
Interactions between variables	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	13
Introduction to GIT	Introduction to GIT	5
Introduction to Linear Regression	Simple and multivariate linear regression	2
Jupyter Notebook	Installing Python	7

Key Git Commands	Introduction to GIT	7
Key XGBoost Concepts	XGBoost	8
Legal Issues GDPR	Interpretability part 1	14
LightGBM	Forecasting Methods	10
LIME Image	Interpretability part 4	10,1 1
LIME-Key Ideas	Interpretability part 4	8,9
Linear Models in 1D	Classification#2_Types_of_Classification _Models	6
Linear Models in 2D	Classification#2_Types_of_Classification _Models	7
Linear Models in 3D	Classification#2_Types_of_Classification _Models	8
Linear vs Non-Linear Models	Classification#2_Types_of_Classification _Models	21
Lineplots	Data visualization	7
LLM architecture	LLM & Ethics	23
LLM Examples	LLM & Ethics	26- 30
LLM Training	LLM & Ethics	24
LLMs and Higher Education	LLM & Ethics	36, 37
Loss function	LightGBM	6
Low-Code approach	Low Code 1	2
Machine Learning	Introduction to ML	4
Machine Learning applications becoming pervasive	Interpretability part 1	5
Machine Learning Methods for Time Series	Forecasting Methods	5
Machine Learning Pipeline with PyCaret	Low Code 2	2
Main Component	LightGBM	5
Main findings from aggregated results	LLM & Ethics	31
Merging data tables	Data Preprocessing	7
Minimizing Residuals with Gradient Descent	XGBoost	5
Missing Data	Data Preparation (for tabular data)	4-5
Model Building	No Code	9
Model Evaluation	No Code	10
Modelling preparation	Data Preprocessing	20
Modern CV	Computer Vision	76- 85
More advanced techniques of text feature extraction	Feature Extraction	19
Multivariate coefficients of determination	Simple and multivariate linear regression	14

Multivariate regression	Simple and multivariate linear regression	8
Neural Network Basics	Neural Networks	4-28
Neural Network Training	Neural Networks	30-50
No-Code approach	No Code	2
Non-Linear Models	Classification#2_Types_of_Classification_Models	9-18
Orange Data Mining	No Code	3
Orange ML: Overview and Features	No Code	4
Orange Workflow	No Code	5
outcomes	Data Preprocessing	25
Overfitting	Overfitting	5-8
Oversampling the minority class	Imbalanced Dataset	7
Overview of XGBoost and Boosting	XGBoost	2
Padding	Computer Vision	42-48
Partial dependency plot	Interpretability part 2	9-12
Plot structure	Data visualization	11
Plotting the Model	Low Code 2	4, 5, 6, 7
Pooling	Computer Vision	58-63
Predicting with the Model	Low Code 2	10
Probabilistic Classification	Classification#3_Evaluating_Classification_Models	10, 11, 12, 13
Probabilistic decision tree	Classification#3_Evaluating_Classification_Models	15
Probabilistic Linear Classifier	Classification#3_Evaluating_Classification_Models	14
Proper ML Pipeline	Classification#1_What_is_Classification	24, 25
Pull Request	Introduction to GIT	9
PyCaret	Low Code 1	3, 4, 5
Python modules	Feature Extraction	14
Quantity sold = f(advertising costs)	Regression Analysis Guide	11
Random Forest	Random Forest	5
Random Forest Classifier	Low Code 1	14
Random Search	Hyperparameter Tuning	7
Regression	Time Series Forecasting	7
Regression - Least Squares Method	Simple and multivariate linear regression	5, 6, 7
Regression metrics	Evaluation Metrics	10
Regression metrics - Mean Absolute Error (MAE)	Evaluation Metrics	11

Regression metrics - Mean Squared Error (MSE)	Evaluation Metrics	12
Regression metrics - Root Mean Squared Error(RMSE)	Evaluation Metrics	13
Regularization, Overfitting, and Learning Rate	XGBoost	4
Reinforcement Learning	Introduction to ML	5,12-13
Relation of NLP, foundation model, and LLM	LLM & Ethics	9
Removing unseen data	Low Code 1	9
Repositories - local and remote	Introduction to GIT	6
Residuum	Regression Analysis Guide	12
Ridge Classifier	Low Code 1	15
Sampling	Data Preprocessing	21
Save/Load Model for Production	Low Code 2	11
Scatterplots	Data visualization	5
Semi-Supervised Learning	Introduction to ML	5
Sequential forward selection	Feature Engineering	11,12
Setting up the PyCaret environment Now	Low Code 1	10
Signal decomposition	Data Preprocessing	17
Significance of individual variables	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	8
Significance of the whole model	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	9
Some Properties of Interpretations	Interpretability part 1	12
Splitting data for model training and evaluation	Data Preprocessing	23
Starting with a Single Leaf Node	XGBoost	6
Steps of regression analysis	Regression Analysis Guide	2, 3, 4, 5, 6, 7
Strengths and Weaknesses of Random Forest	Random Forest	8
Stride	Computer Vision	50-56
Submodular selection (SP-LIME)-Key Ideas	Interpretability part 4	12
Subplots	Data visualization	12
Supervised Classification	Classification#1_What is Classification	6
Supervised Learning	Introduction to ML	5-10,13

Supervised ML Pipeline	Classification#1_What_is_Classification	13-23
Supervised Regression	Classification#1_What_is_Classification	7, 8
Support Vector Machine	Forecasting Methods	6
Support Vector Regression	Forecasting Methods	7
Taxonomy of Interpretability methods	Interpretability part 2	7
Texture features	Feature Extraction	8
The meaning of simple linear regression	Simple and multivariate linear regression	3
Time Series	Time Series Forecasting	4
Time Series Cross Validation	Time Series Forecasting	10
Time Series pattern	Time Series Forecasting	5
Tokenize language	LLM & Ethics	14, 15
Trade off between Accuracy and Interpretability	Interpretability part 2	6
Transformation	No Code	7
Tuning the Model	Low Code 2	3
Types of Deep Learning Architectures	Classification#2_Types_of_Classification_Models	20
Types of joins	Data Preprocessing	8
Types of Machine Learning Tasks	Classification#1_What_is_Classification	4, 5, 9, 11
Uncertainty Quantification	Time Series Forecasting	11
Underfitting	Overfitting	9-10
Undersampling the majority class	Imbalanced Dataset	8
Unsupervised	No Code	11
Unsupervised Learning	Introduction to ML	5,11, 13
Use of Python	Installing Python	5
Visualization	No Code	8
What do machines see?	Computer Vision	10-18
What does it look like to interact with LLMs?	LLM & Ethics	10
What is a foundation model?	LLM & Ethics	7
What is an imbalanced Dataset?	Imbalanced Dataset	4
What is Computer Vision?	Computer Vision	5, 6, 7
What is interpretability?	Interpretability part 1	4
What is ML Model?	Classification#2_Types_of_Classification_Models	5
What is version control?	Introduction to GIT	4
Why explainability	Interpretability part 1	6
Why Explainability: Verify the ML Model / System	Interpretability part 1	8

Why is an imbalanced Dataset a Problem?	Imbalanced Dataset	5
Why is the model making these predictions?	Interpretability part 1	7
Why Python?	Installing Python	4
Why we need data preprocessing?	Data Preprocessing	2
Why we need version control?	Introduction to GIT	3
Wrapper algorithms	Feature Engineering	13
XGBoost Efficiency Features	XGBoost	9
XGBoost in Classification Problems	XGBoost	11
XGBoost in Regression Problems	XGBoost	10
	Classification#1_What is Classification	6,7,8
	Classification#1_What is Classification	10
	Classification#1_What is Classification	12
	Data preprocessing	19
	Introduction to ML	5, 10
	Interpretability part 2	5