

# Contents of Machine Learning

## Part 1: Classical Machine Learning

- **Theory**
  - ML as part of Artificial Intelligence
  - Types of ML: supervised learning, unsupervised learning, reinforcement learning, regression, classification
  - Terminology:
    - task data; data frame (rows and columns, samples and features); input features and output class; train-test evaluation, data and algorithm bias; multi-class, binary, multi-label classification
  - Linear, non-linear and inseparable tasks
  - Batch and incremental learning
  - Data and compute scale: effect of training data and computing size growth
  - Machine Learning project steps: data collection, analysis and preprocessing, model training, model testing
  - Ethical issues
  - Bias / Variance trade-off / overfitting and underfitting / regularization
  - Evaluation
  - Cost functions
  - Gradient Descent
- **ML algorithms and methodology**
  - Univariate and multiple Linear Regression
  - Logistic Regression
  - Missing values and imputation
  - Scaling
  - Pipelines, column transformers
  - Hyperparameter optimization, gridsearch, randomizedsearch
  - Baseline algorithms
  - Evaluation metrics
  - K-nn Classification
  - Decision Tree Learning
  - SVM
  - Random Forest classifier
  - SGDClassifier
  - Ensemble methods
- **Datasets**
  - real\_estate.csv
  - titanic.csv
  - TOEFL.csv
  - adult\_small.csv
- **Python (Pandas and sklearn, for complete list: check the notebooks)**

- Data analysis and visualization: `df.head()`, `df.tail()`, `df.sample()`, `df.describe()`, `df.info()`, `df.corr()`, `df.plot()`, `df.duplicated()`, `df.isnull()`, `pd.plotting()`
- Sklearn experiments: `train_test_split()`, `make_pipeline()`; `GridSearchCV()`, `learning_curve`, `validation_curve`, `cross_val_score`, `RandomSearchCV()`, `ColumnTransformer()`,
- Sklearn: estimators and transformers: `X` and `y`, `.fit(X,y)`, `.transform(X)`, `.fit_transform(X)`, `.predict(X)`; `LinearRegression()`, `LogisticRegression()`, `DummyClassifier()`, `StandardScaler()`, `MinMaxScaler()`, `MaxAbsScaler`, `KNNImputer()`, `SimpleImputer()`, `SGDClassifier()`, `LinearSVC()`, `SVC()`, `DecisionTreeClassifier()`, `KNeighborsClassifier()`, `OneHotEncoder()`, `OrdinalEncoder()`, `RandomForestClassifier()`, `GradientBoostingClassifier()`, `VotingClassifier()`
- Sklearn evaluation (accuracy, precision, recall, `f1_score`): `ConfusionMatrixDisplay()`, `classification_report()`, `mean_squared_error()`, `r2_score()`, `RocCurveDisplay()`, `PrecisionRecallDisplay()`
- Sklearn text preprocessing: `CountVectorizer()`, `TfidfVectorizer()`