

# Machine Learning with scikit-learn

## What Is Machine Learning?

- Difference between "Deep Learning" and other ML techniques
- Overview of techniques used in Machine Learning
- Classification vs. Regression vs. Clustering
- Dimensionality Reduction
- Feature Engineering
- Feature Selection
- Categorical vs. Ordinal vs. Continuous variables
- One-hot encoding
- Hyperparameters
- Grid Search
- Metrics

**Open Notebook** ([WhatIsML.ipynb](#))

## Exploring a Data Set

- Looking for anomalies and data integrity problems
- Cleaning data
- Massaging data format to be model-ready
- Choosing features and a target
- Train/test split

**Open Notebook** ([Exploring.ipynb](#))

## Classification

- Choosing a model
- Feature importances
- Cut points in a decision tree
- Comparing multiple classifiers

**Open Notebook** ([Classification.ipynb](#))

## Regression

- Sample data sets in scikit-learn
- Linear regressors
- Probabilistic regressors
- Other regressors

**Open Notebook** ([Regression.ipynb](#))

## Clustering

- Overview of (some) clustering algorithms
- Kmeans clustering
- Agglomerative clustering
- Density based clustering: DBSan and HDBScan
- n\_clusters, labels, and predictions
- Visualizing results

**Open Notebook** ([Clustering.ipynb](#))

## Hyperparameters

- Understanding hyperparameters
- Manual search of parameter space
- GridsearchCV
- Attributes of grid search and wrapped model

**Open Notebook** ([Hyperparameters.ipynb](#))

## Feature Engineering and Feature Selection

- Principal Component Analysis (PCA)
- Non-Negative Matrix Factorization (NMF)
- Latent Dirichlet Allocation (LDA)
- Independent component analysis (ICA)
- SelectKBest
- Dimensionality expansion
- Polynomial Features
- One-Hot Encoding
- Scaling with StandardScaler, RobustScaler, MinMaxScaler, Normalizer, and others
- Binning values with quantiles or binarize

**Open Notebook** ([Features.ipynb](#))

## Pipelines

- Feature Selection and Engineering
- Grid search
- Model

**Open Notebook** ([Pipelines.ipynb](#))

## Robust Train/Test Splits

- cross\_val\_score
- ShuffleSplit
- KFold, RepeatedKFold, LeaveOneOut, LeavePOut, StratifiedKFold

**Open Notebook** ([TrainTest.ipynb](#))

