

1
2
3 Αναγνώριση Προτύπων {
4

5
6 [Εισαγωγή στην Python]
7

8
9 < Υπ. Διδ. Στέφανος Παπαδόπουλος
10 stefpapad@iti.gr >
11

12 }
13
14

Γιατί Python;

Multi-purpose:

- Web και software development
- Μαθηματικά
- Ανάλυση μεγάλων δεδομένων
- Αναγνώριση προτύπων και μηχανική μάθηση

Procedural, object-oriented, functional programming

Μεγάλη ποικιλία εργαλείων και βιβλιοθηκών

Εύκολο συντακτικό

Ενεργή κοινότητα

Χρήση της Python και του Jupyter

Τρεις τρόποι:

1. Εγκατάσταση της Python 3 (<https://www.python.org/downloads>)
Επιλογή IDE (π.χ Vscode ή Pycharm)
Χρήση virtual environment (venv)
Εγκατάσταση του Jupyter μέσω → `pip3 install jupyter`
^ Δεν προτείνεται για το μάθημα!
2. Εγκατάσταση του Anaconda (<https://docs.anaconda.com/anaconda/install>)
(περιλαμβάνει το Jupyter κ.α)
Δημιουργούμε ένα conda environment
→ Run “Anaconda prompt” (in windows, terminal in linux)
→ `conda create --name env_name python=3` (μονο την πρώτη φορά)
→ `conda activate env_name`
→ `conda install jupyter` (μόνο την πρώτη φορά)
→ `jupyter notebook`
Tip: Μπορεί να συμβουλευτείτε το “CONDA CHEAT SHEET”
3. Χρήση του Google Colab (<https://colab.research.google.com>)

Διάλεξη 1 - 11/10/22

- Σύνταξη
- Μεταβλητές
- Τύποι δεδομένων
- Δομές δεδομένων
- Looping
- Συναρτήσεις
- Πράξεις πινάκων με την NumPy
- Δημιουργία γραφημάτων με την Matplotlib

Colab: https://colab.research.google.com/drive/1SKwlaWEj-dkc-L58AaJatXZ_6IXZ8Z8y

Διάλεξη 2 - 18/10/22

- Οπτικοποίηση 3D δεδομένων
- Αντικείμενα και κλάσεις
- Αναγνώση, εξερεύνηση, επεξεργασία και οπτικοποίηση δεδομένων με την Pandas

Colab: https://colab.research.google.com/drive/1sbxaAoufp5MpE2ogCdPK4oNNmXdcS_tv

Διάλεξη 3 - 1/11/22

- Εισαγωγή στη μηχανική μάθηση
- Μηχανική μάθηση με την Scikit-learn

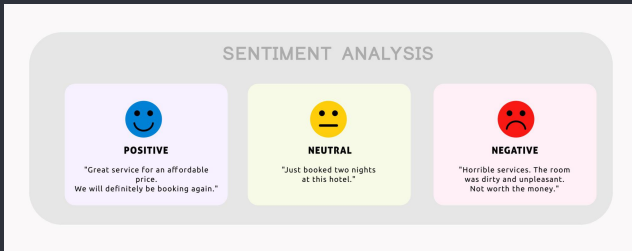
“Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed” - Arthur Samuel

1. Ορισμός προβλήματος

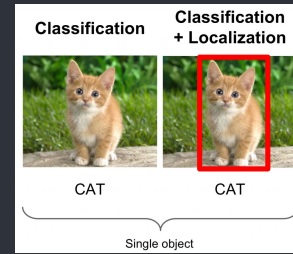
Input → Machine learning (ML) model → Output
Features → ML model → Target



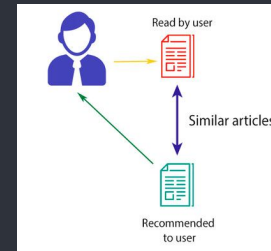
Πρόβλεψη χρονοσειρών



Ανάλυση συναισθήματος σε κείμενα



Εντοπισμός αντικειμένων σε εικόνες



Σύστημα συστάσεων

2. Είδη προβλημάτων

- **Supervised learning** (labeled data)
 - **Classification**
 - Regression
- Unsupervised learning
 - Clustering
 - Dimensionality reduction
 - Association rules
- Semi-supervised learning
- Self-supervised learning
- Reinforcement learning

3. Συλλογή δεδομένων

- Manual annotation
- Web scraping
- Synthetic data generation
- Public datasets, e.g:
 - <https://www.kaggle.com/datasets>
 - <https://datasetsearch.research.google.com/>
 - <https://github.com/awesomedata/awesome-public-datasets>

Τα 5 'V' των δεδομένων:

- Volume
- Variety
- Velocity
- Value
- Veracity

4. Προετοιμασία και εξερεύνηση δεδομένων

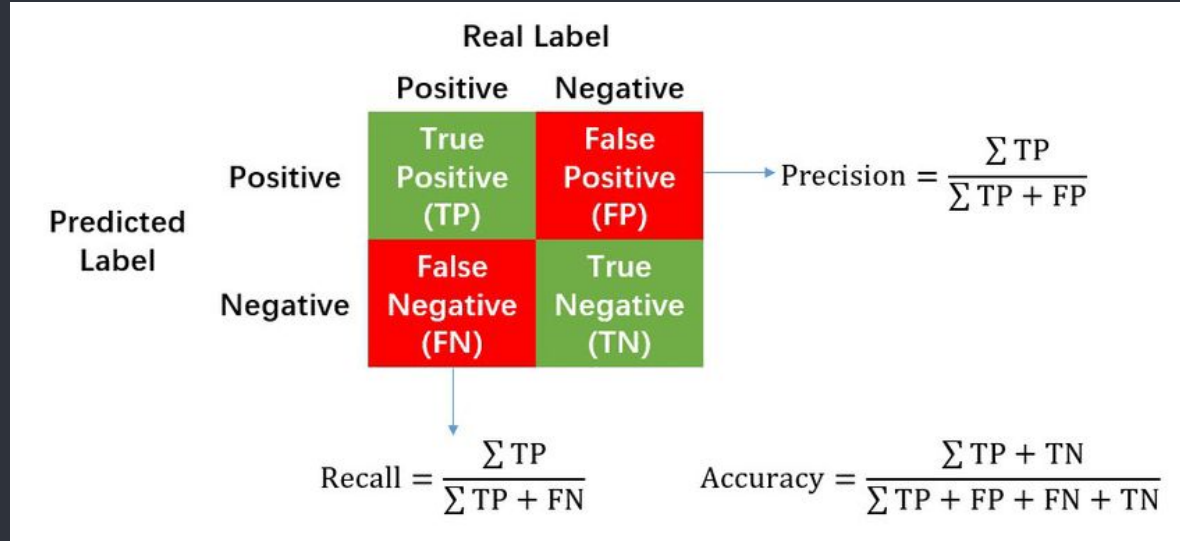
- Dealing with missing values
- Outlier detection
- Data normalisation
- Data cleaning
- Numerical representation
- Dealing with class imbalance
- Visualisation

5. Εκπαίδευση και αξιολόγηση

- Μέθοδος αξιολόγησης
 - Training / validation / testing sets split
 - Cross-validation
- Επιλογή κατάλληλης μετρικής αξιολόγησης για το πρόβλημα
- Επιλογή κατάλληλου ML μοντέλου για το πρόβλημα
- Εκπαίδευση και αξιολόγηση

5. Εκπαίδευση και αξιολόγηση

Key metrics for classification



Recall: How many relevant items are retrieved?

Precision: How many retrieved items are relevant?

Machine learning workflow

- Ορίζουμε το πρόβλημα που θέλουμε να λύσουμε
 - Αναγνωρίζουμε το είδος του προβλήματος:
 - * Supervised ή unsupervised
 - * Classification, Regression, Clustering ή άλλο
 - Συλλέγουμε τα δεδομένα
 - Καθαρισμός / προ-επεξεργασία των δεδομένων
 - Εξερεύνηση δεδομένων και οπτικοποίηση
 - Χωρίζουμε τα δεδομένα σε training, validation, testing sets
 - Επιλογή κατάλληλης μετρικής αξιολόγησης
 - Επιλογή κατάλληλου μοντέλου και εκπαίδευση
 - Αξιολόγηση μοντέλου
- <https://colab.research.google.com/drive/12gLj8MLeeu75hjavaKTGTwWn4JTt73P-?usp=sharing>

Διάλεξη 4 - 7/11/22

Προχωρημένα θέματα μηχανικής μάθησης με την
sklearn

- Sentiment analysis and text representation
- Validation and cross-validation
- Hyper-parameter tuning

<https://colab.research.google.com/drive/1n78xQc03KZ3inxN1d0K0JG5msXG1GLnR?usp=sharing>

Διάλεξη 5 - 16/12/2022

Εισαγωγή στα νευρωνικά δίκτυα και την βαθιά μηχανική μάθηση με το Pytorch:

- Tensors
- Data και data loaders
- Model class
- Optimization
- Training and evaluation loops

<https://colab.research.google.com/drive/1hET5StICzUSxnR2naOZfrdNZ-lsfYjc?usp=sharing>