

# 1. Cvičení

---

# Agenda

- přehled nástrojů používaných pro strojové učení
  - instalace potřebných nástrojů
  - ukázka práce s Jupyter Notebook
  - Python tutoriál
  - NumPy tutoriál
  - ukázka řešení lineární regrese
  - úkoly
-

# Co budeme ke strojovému čení potřebovat a používat

- Python
  - optimalizovaný pro práci s daty
  - velké množství knihoven + kompletní ekosystém pro strojové učení
- knihovny
  - NumPy - práce s multidimenzionálními poli, základ pro ostatní knihovny a standard v ML
  - SciPy - “vědecká” knihovna - lineární algebra, optimalizace, apod.
  - Scikit-learn - základní knihovna pro strojové učení (obsahuje optimalizovanou implementaci většiny používaných algoritmů)
  - Pandas - práce s daty (načítání CSV, vizualizace apod.)
  - Matplotlib - vizualizace dat (grafy, obrázky, ...)
- TensorFlow + Keras
  - frameworky pro Deep Learning (hluboké neuronové sítě)
- Jupyter Notebook
  - webová aplikace pro vytváření dokumentů obsahujících python kód, vizualizace a doprovodný text

# Hardware

- čím výkonnější počítač, tím méně budete čekat :)
  - moje pracovní stanice: i7-6700K, 16 GB RAM, Nvidia GTX 1060 6GB
- pro většinu úloh, co si budeme ukazovat, však stačí běžný notebook
- pro Deep Learning (hluboké neuronové sítě) GPU Nvidia
  - nebo Amazon EC2 Deep Learning AMI
  - nebo Google Colab s GPU
  - pro absolvování předmětu to však není nutné

# Instalace nástrojů

- Python 3.8
  - TensorFlow aktuálně nepodporuje nejnovější Python 3.9
  - <https://www.python.org/downloads/release/python-387/>
    - pro windows zvolit windows installer
    - zvolit Customize installation
      - vše zaškrtnout
      - Install for all users
      - Add Python to environment variables
- volitelně instalace virtualenv pro izolované python prostředí
  - instalace: `pip install virtualenv`
  - vytvoření prostředí nazvaného ml: `python -m virtualenv ml`
  - spuštění prostředí:
    - windows: `.\ml\Scripts\activate`
    - linux: `env/bin/activate`
- knihovny
  - instalace pomocí nástroje PIP
  - `pip install jupyter matplotlib numpy pandas scipy scikit-learn`
    - nebo `python -m pip ...` (pokud by přímo pip nefungovalo)
- Jupyter Notebook
  - spuštění: `jupyter notebook` (v adresáři, kde budete chtít mít pracovat)
  - otevře se `http://localhost:8888/tree` (nebo otevřít v prohlížeči, pokud by se nestalo samo)

# Jupyter Notebook

- proč používat
  - sešit, který má text, zdrojové kódy, grafy, ...
  - proměnné jsou uloženy v relaci, dokud se proces nevypne
  - přehledný výpis přímo pod buňku s kódem, rovnou zobrazené grafy (matplotlib) a tabulky (Pandas)
  - formátování pomocí markdownu
  - nezávislost buněk na sobě
  - => velmi vhodné pro rychlé prototypování a vizualizaci
- hostované varianty
  - Google Colab
  - další viz <https://www.kdnuggets.com/2020/03/4-best-jupyter-notebook-environments-deep-learning.html>

# Jupyter Notebook - ukázka

- spustit příkazem `jupyter notebook`
- otevřít `http://localhost:8888`
- vytvoření sešitu
- pojmenování sešitu
- módy
  - command vs edit
  - buňka code vs markdown
  - klávesové zkratky (A, B, DD, shift+enter, ctrl+enter, H), help -> user interface tour
- hello world
  - kód
  - graf
  - tabulka s daty

# Python tutorial

- <https://docs.python.org/3/tutorial/>



# NumPy tutorial

- otevřít sešit 01\_numpy.ipynb
- nebo [https://numpy.org/doc/stable/user/absolute\\_beginners.html](https://numpy.org/doc/stable/user/absolute_beginners.html)

# Ukázka řešení lineární regrese pomocí strojového učení

- otevřít 01\_linearni\_regrese.ipynb

# Úkoly

1. Zprovoznit si Python, Jupyter Notebook apod. na svém počítači
2. Projít si Python tutoriál:
  - a. <https://docs.python.org/3/tutorial/>
3. Projít si NumPy tutoriály:
  - a. soubor 01\_numpy.ipynb
  - b. [https://numpy.org/doc/stable/user/absolute\\_beginners.html](https://numpy.org/doc/stable/user/absolute_beginners.html)
  - c. <https://numpy.org/doc/stable/user/quickstart.html>