**Content**

[1. Terminology 2](#_Toc16096734)

[1. Gradient Descent 2](#_Toc16096735)

[1.1.1. What is the Gradient? 2](#_Toc16096736)

[1.1.2. Back Propagation 2](#_Toc16096737)

[1.1.3. Learning Rate 3](#_Toc16096738)

[1.1.4. The Complete Mathematical Notation of a Fully Connected Neural Network 3](#_Toc16096739)

[2. Entropy and Information Gain 4](#_Toc16096740)

[2. Technologies 5](#_Toc16096741)

[1. Jupyter Notebooks 5](#_Toc16096742)

[2.1.1. %matplotlib inline and other ipython magic functions 5](#_Toc16096743)

[2. Scikit-Learn 5](#_Toc16096744)

[2.2.1. Pandas 6](#_Toc16096745)

[2.2.2. NumPy 8](#_Toc16096746)

[3. One-class classification (OOC) 10](#_Toc16096747)

[1. PU learning 10](#_Toc16096748)

1. Terminology

* Overfitting - the splitting process results in fully grown trees until the stopping criteria is reached. But, the fully grown tree is likely to overfit data, leading to poor accuracy on unseen data.
* Variance
* Pruning
* Greedy algorithm – always makes the choice that seems to be the best at the moment. This means that it makes a locally-optimal choice in the hope this choice will lead to a globally-optimal solution.
* Correlation
* Epoch – number of times you give data to the neural network. Less epochs underfit, more epochs overfits the data.
* Batch Size – the number of samples that will be passed through to the network at one time. 1000 Images ; 10 batch size -> 100 batch will be needed for 1 epoch. Larger batches = faster training.
  1. Gradient Descent

Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point. If, instead, one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then known as gradient ascent.

Chain rule – derivative of nested functions

Calculate the back propagation for a fully connected neural network.

* + 1. What is the Gradient?

Gradient shows us the maximum of the function.

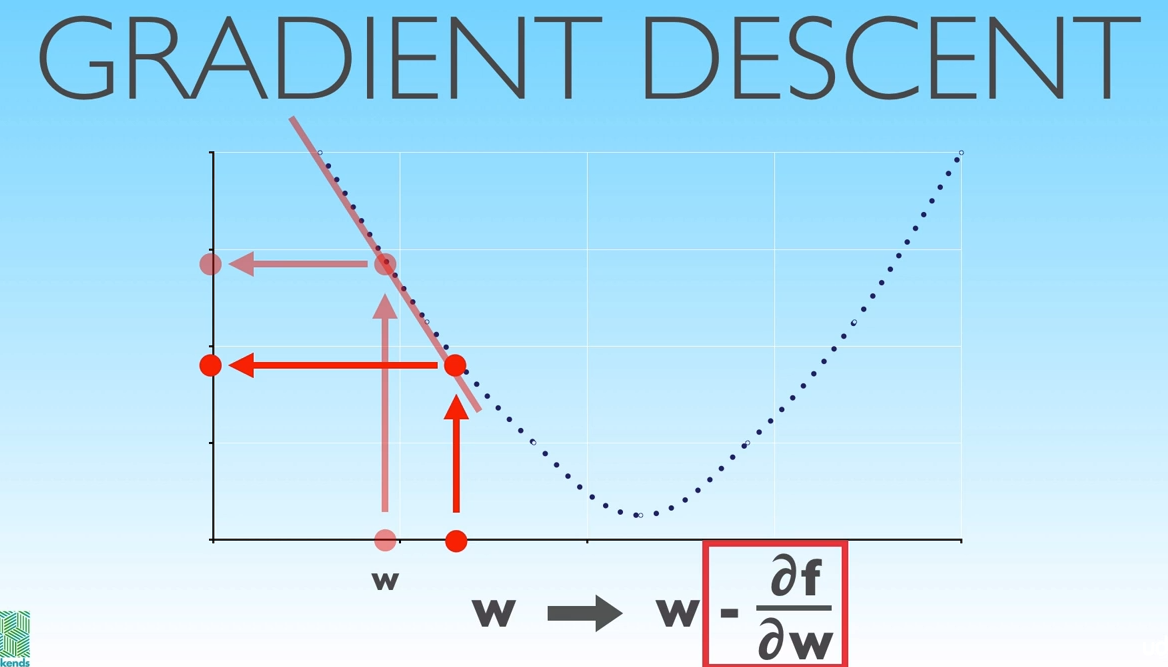
Derivatives: slope, rate of change

Gradient: extends derivative to multivariate functions.

* + 1. Back Propagation

Csokkeno fuggveny - derivacio negative

Novekvo fuggveny – derivalt pozitiv



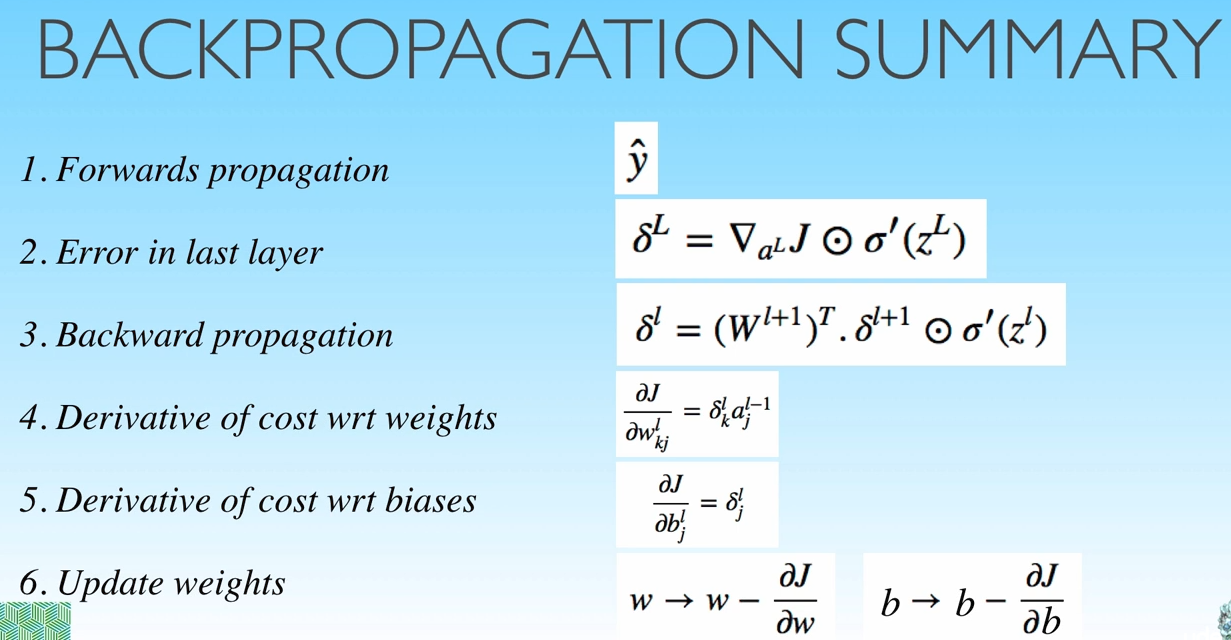
For a neural network we define cost function, that depends on the values of the parameters. We find the values of the parameters by minimalizing the cost by gradient descent. All we are really doing is taking the cost function, calculating its partial derivatives with respect to each parameter and then using the update rule (x := x – df/dx) we decrease the cost by updating the paramters we do this by **subtracting the value of the negative gradient from each of the parameters.**

* + 1. Learning Rate

If the function is very flat we move very slow towards the minimum (using the update rule) but if the function is very steep we can jump over the minimum.



* + 1. The Complete Mathematical Notation of a Fully Connected Neural Network

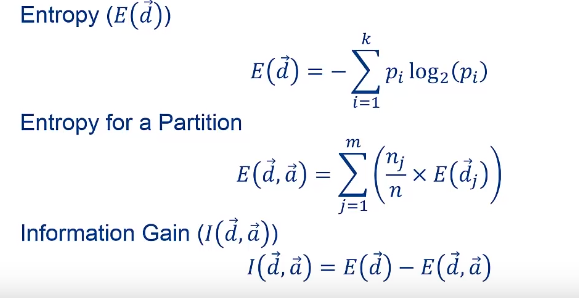


* 1. Entropy and Information Gain

*Entropy is a measure of disorder in a dataset.*

*Information gain a measure of the decrease in disorder achieved by partitioning the original dataset based on some additional attributes.*

We will use the concept of entropy, borrowed from physics, to build a classification method.



Entropy: why ‘-‘ ? Just because we want in to measure in positive. More disorder will be a larger positive value than a negative. Logaritmic function in interval (0 – 1) will be negative.

1. Technologies
   1. Jupyter Notebooks

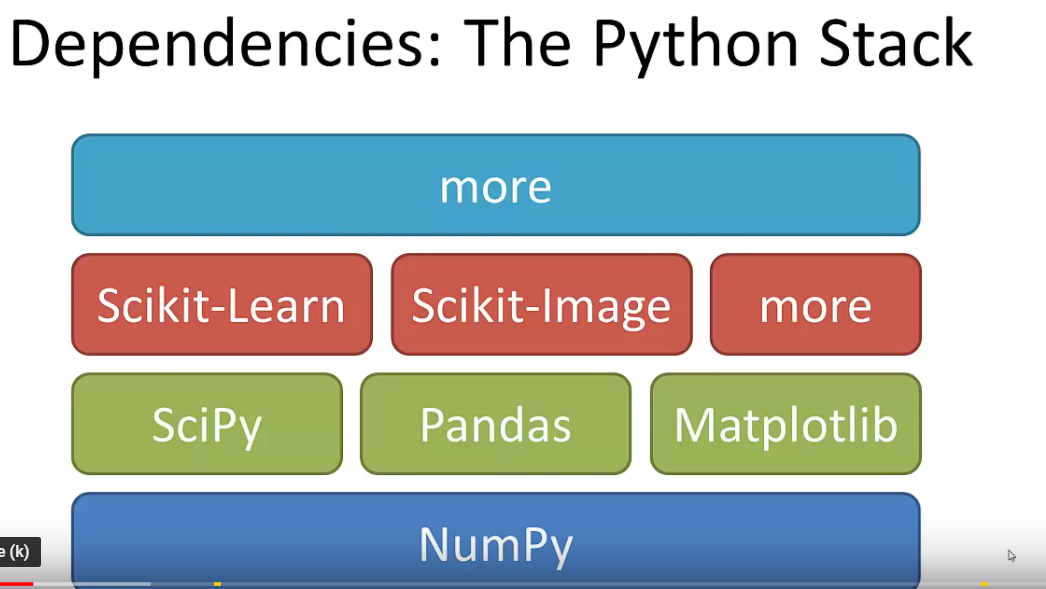
Jupyter Notebook (formerly Ipython (Interactive Python) Notebooks) is a web-based interactive computational environment for creating Jupyter notebooks documents. The "notebook" term can colloquially make reference to many different entities, mainly the Jupyter web application, Jupyter Python web server, or Jupyter document format depending on context. A Jupyter Notebook document is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots and rich media, usually ending with the ".ipynb" extension.

* + 1. %matplotlib inline and other ipython magic functions

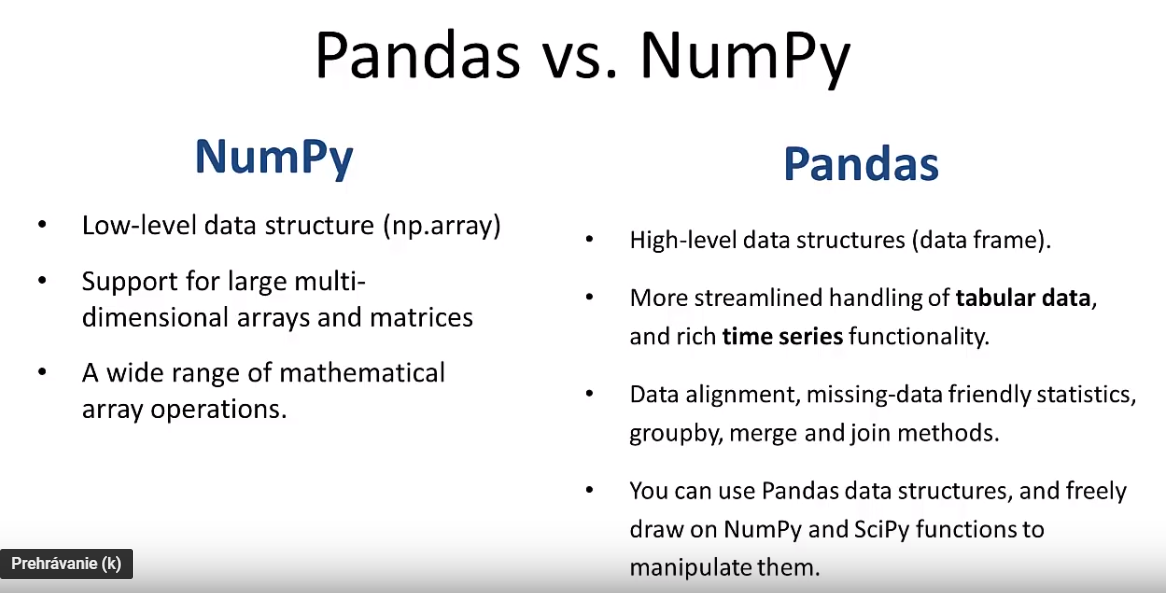
With this backend, the output of plotting commands is displayed inline within frontends like the Jupyter notebook, directly below the code cell that produced it. The resulting plots will then also be stored in the notebook document.

* 1. Scikit-Learn

A free software machine learning library for the Python programming language.



* + 1. Pandas



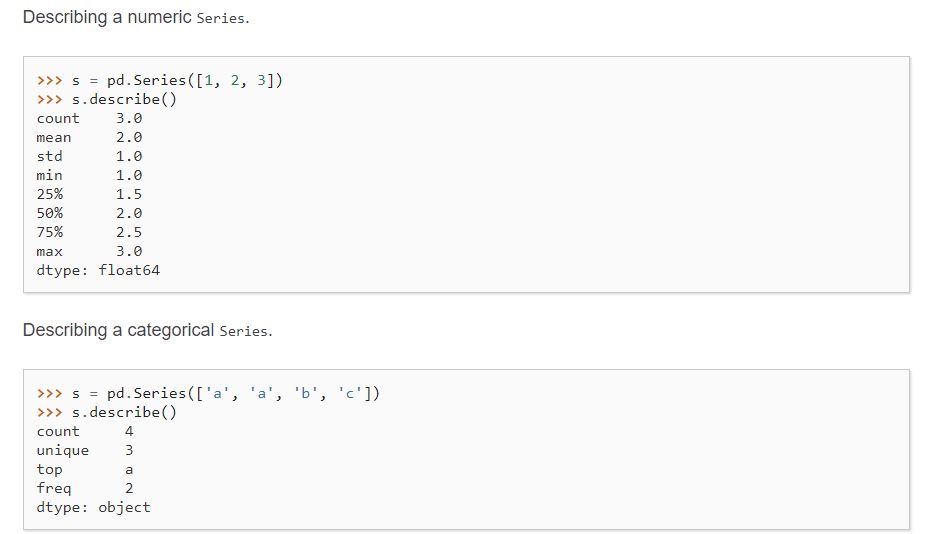
Pandas is an open source Python library providing high-performance, easy-to-use data structures and data analysis tools, runs on top of NumPy.

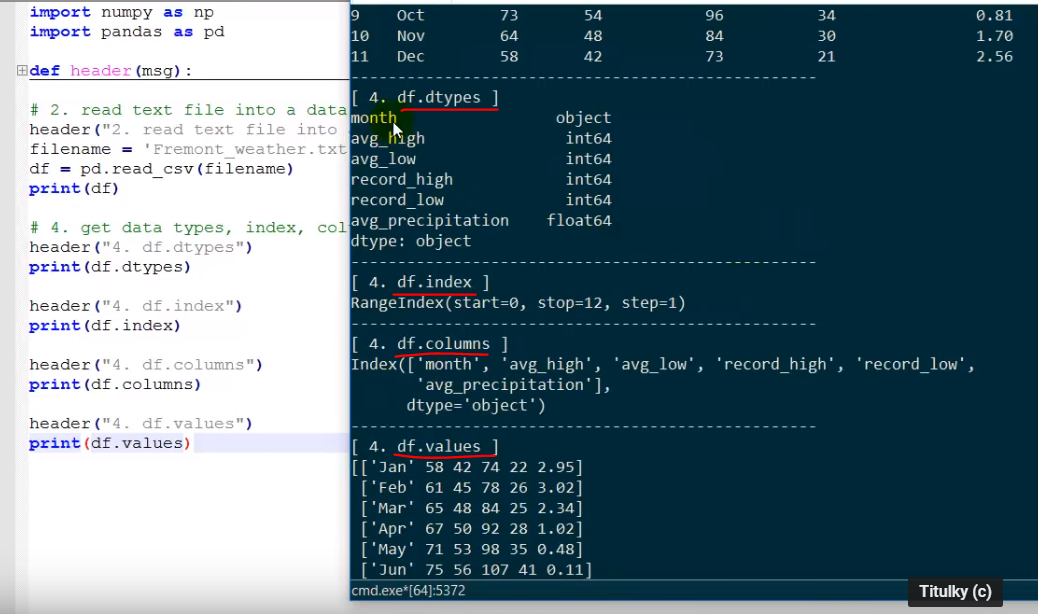
The data structure in Pandas is called DataFrame (df)

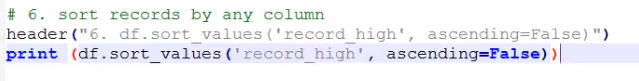
* df= pd.DataFrame()
* df = pd.read\_csv(filename)

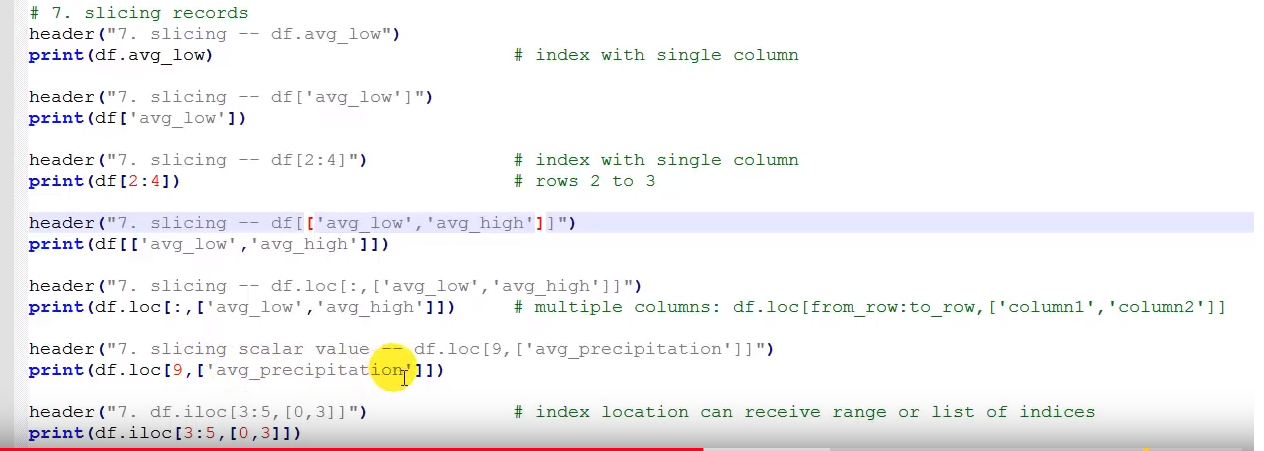
With 1 line of code we read in an entire csv file into a datafarame

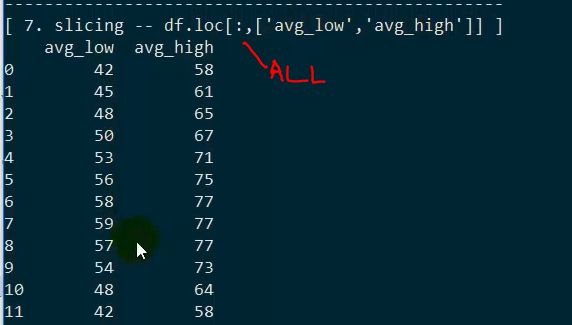
* df.describe()

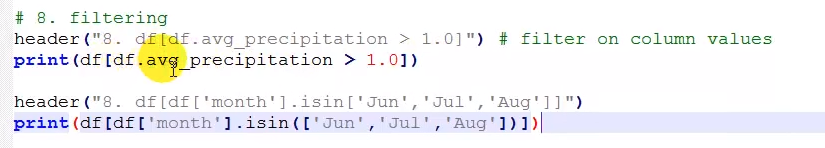




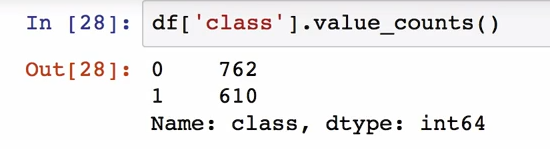




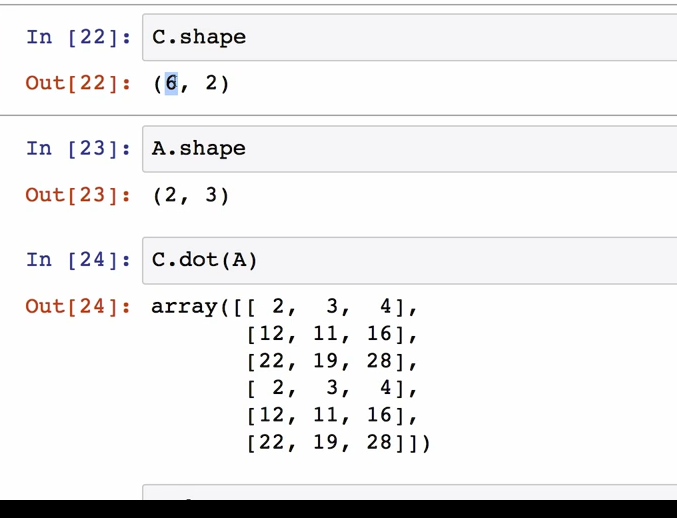




* df.head()
* df.tail(3)



* + 1. NumPy
* Core lib for scientific computing in Python
* It provides a high-performance multidimensional array object, and tools for working with these arrays



Multiplying matrixes – the last dimension of the first and the first dimension of the second matrixes must be equal.

1. One-class classification (OOC)

Several approaches are have been proposed to solve one-class classification (OCC). The approaches can be distinguished into three main categories:

1. density estimation,
2. boundary methods,
3. and reconstruction methods.
   1. PU learning