

BDI Python Code Clinic

19/05/2021

**More Machine Learning in Python  
with scikit-learn  
Session 3**

Irina Chelysheva

# Before we start...

List of the Python SciPy libraries required for this tutorial:

scipy

numpy

matplotlib

pandas

sklearn

+my Python scripts from

[https://github.com/Chelysheva/ML\\_Python\\_course](https://github.com/Chelysheva/ML_Python_course)

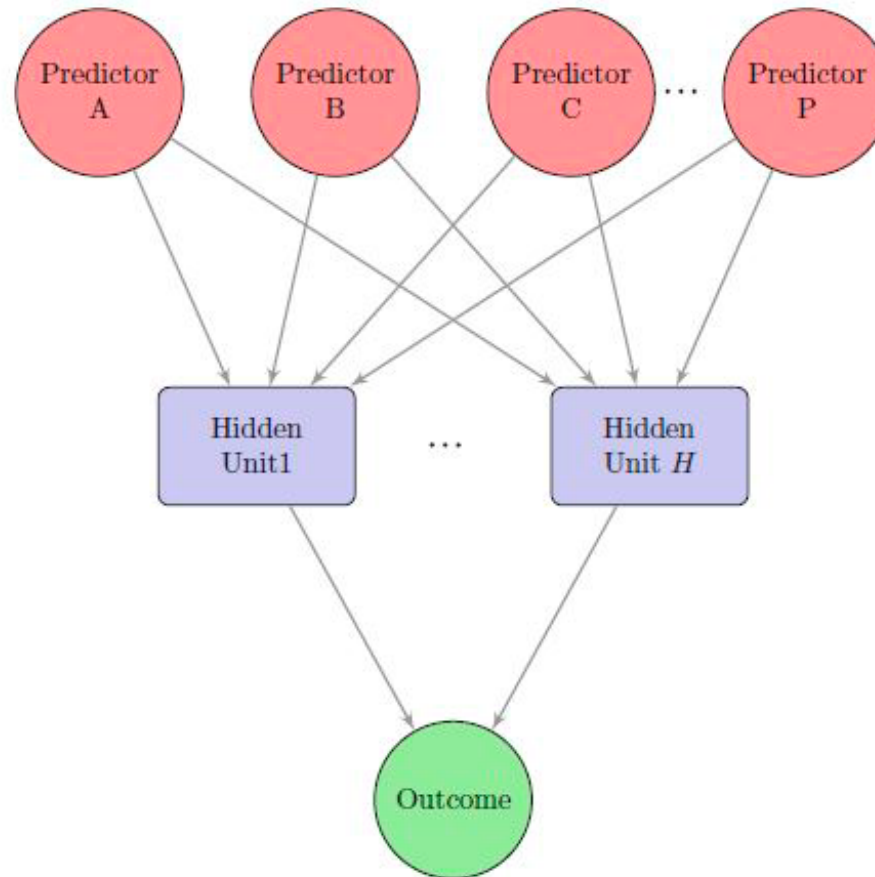
# Neural networks

- (Artificial) neural networks are powerful nonlinear regression techniques inspired by the workings of the human brain
- The main idea is that information is propagated from input nodes (features) to output nodes (predictions), passing through one or more hidden layers of nodes in between.
- But the more hidden layers, the more difficult - and computationally demanding - to train and tune!

2 types:

- With **feed-forward neural networks**, information flows in one direction: forward! It goes from the input nodes through the hidden nodes to the output nodes.
- **Recurrent neural networks** allow for loops and cycles, channeling the information flow in various directions.

# Single-Layer Perceptron

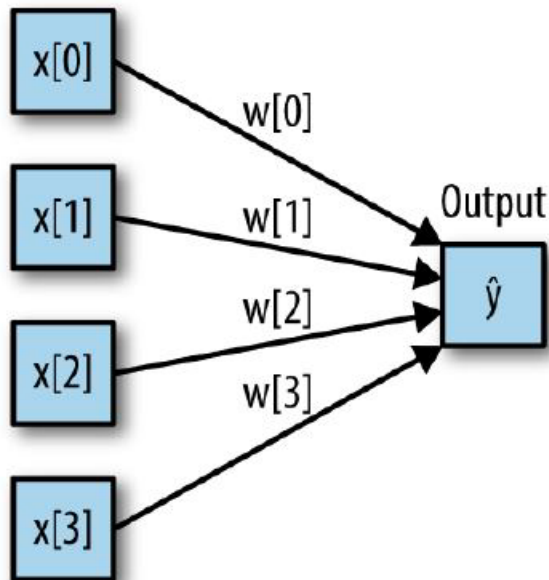


Can perform well for both classification and regression tasks!

# Good for non-linear relationships within dataset

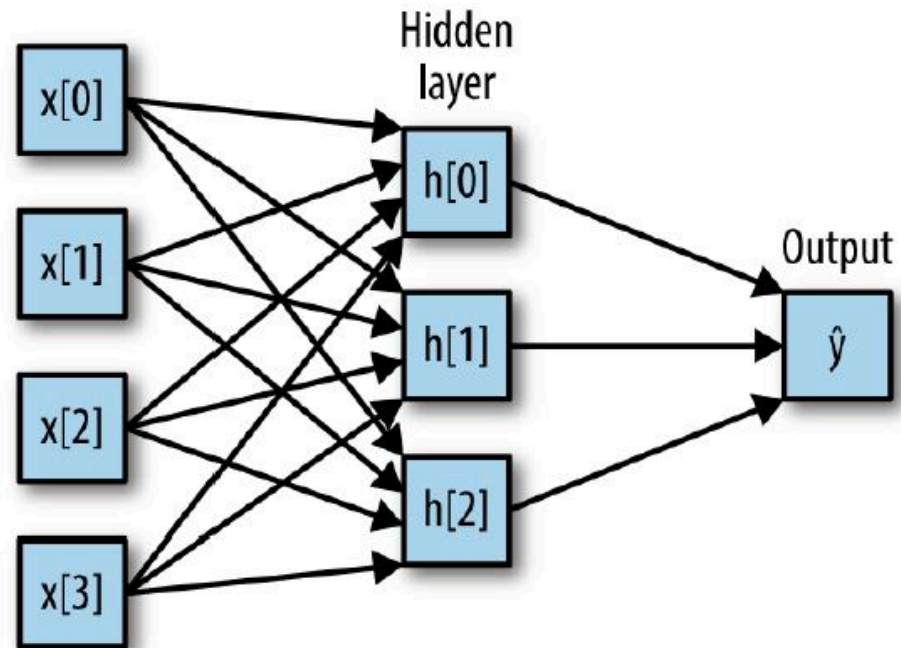
Linear model

Inputs



Neural network

Inputs



Input features and prediction outputs are shown as nodes, the weights (coefficients) are lines between the nodes.

# Principal Component Analysis

- Principal Component Analysis (PCA) finds linear combination of features, called Principal Components (PC), that capture most of the variation. Technically, it conducts a singular value decomposition of the data to project it to a lower-dimensional space.
- First component: The linear combination of the features that captures the most variation of all possible linear combinations.
- Subsequent components capture the most of the remaining variation, while being uncorrelated to the previous PCs.
- Can be used to reduce the number of dimensions before applying ML.

# Today's plan

- Datasets:
  - Breast cancer dataset from sklearn
  - Labeled Faces in the Wild data set from sklearn
- Classification problems
- Apply scaling
- Use neural network algorithm: MLPClassifier
- Perform PCA
- Run model on original data and on the PCA transformed – dimensionally reduced datasets
- Compare the performance – speed and accuracy

# Overview of ML algorithms

