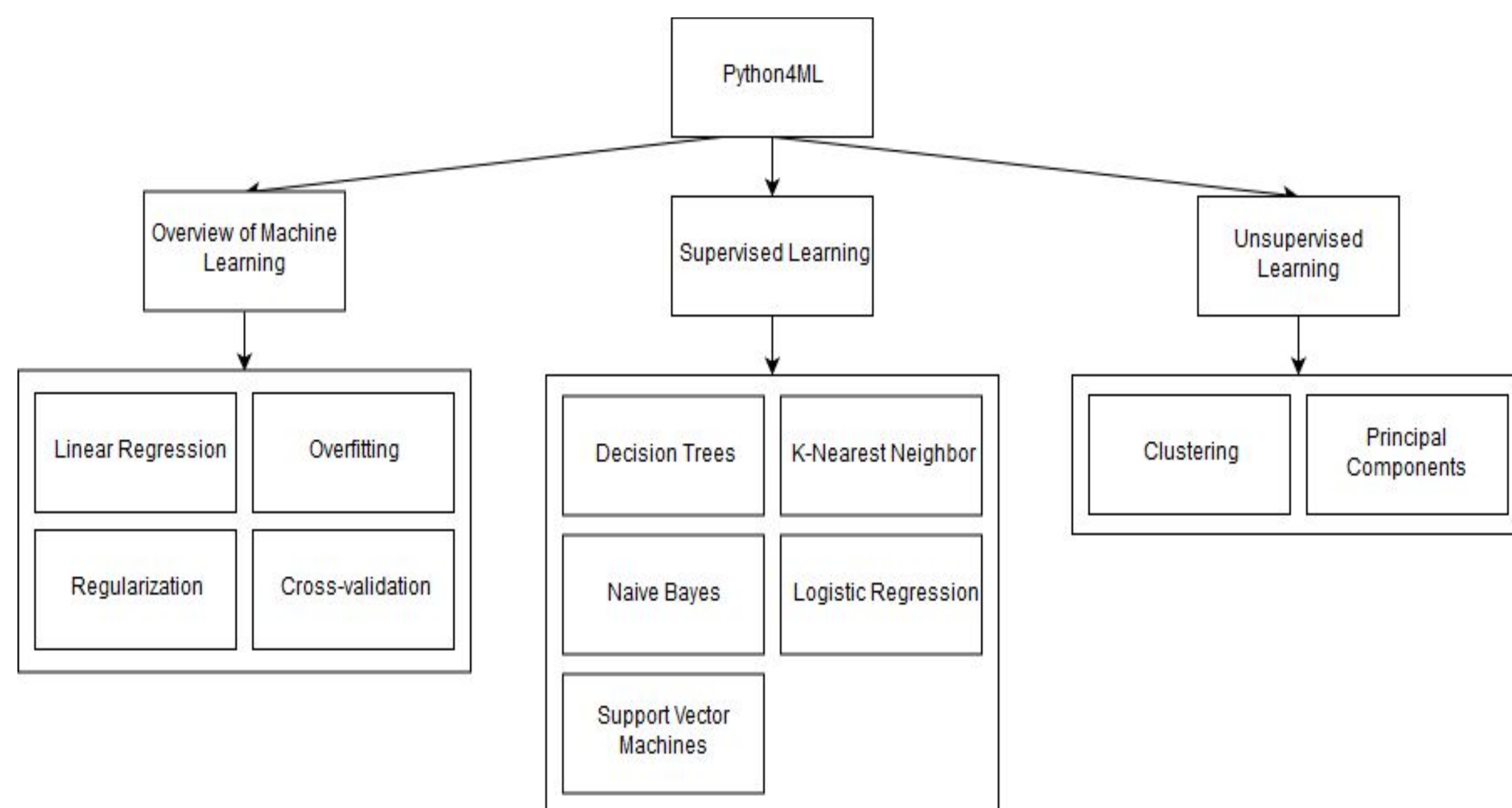


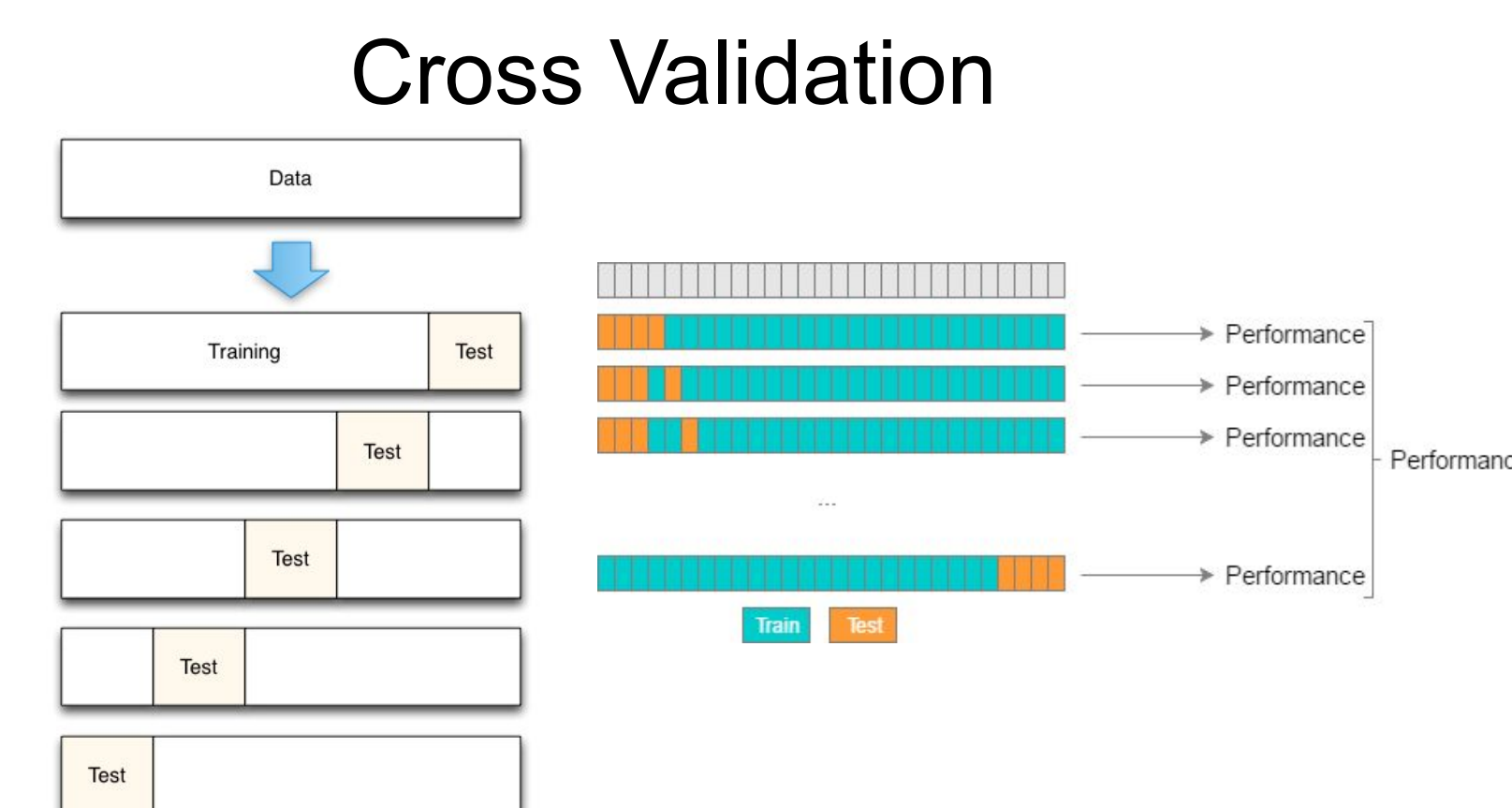
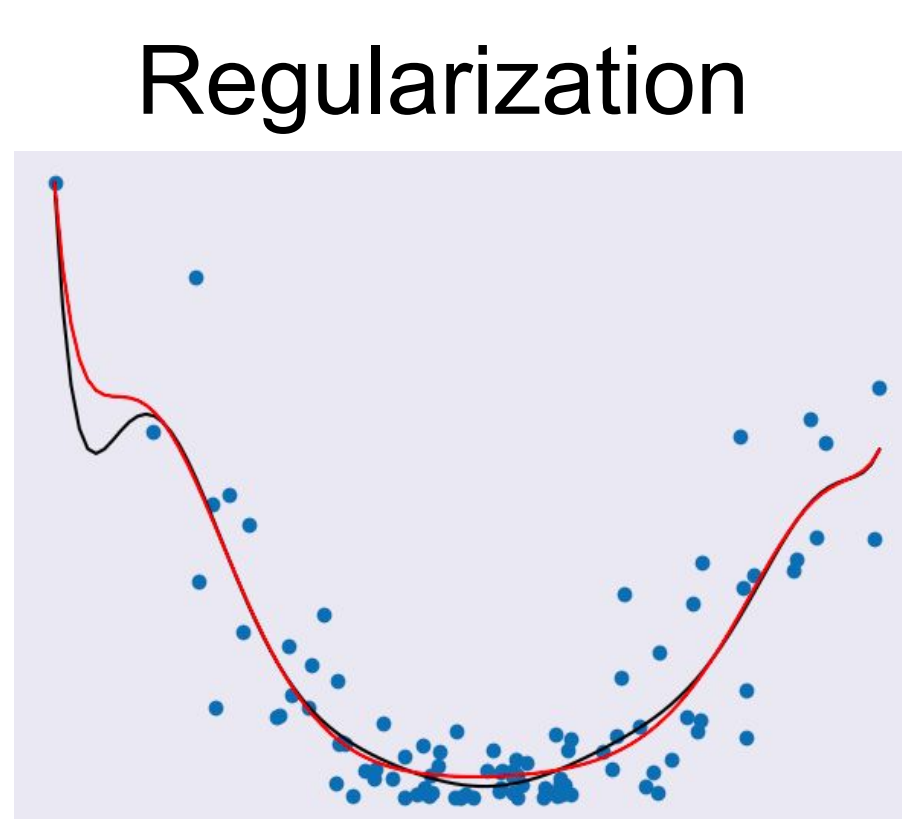
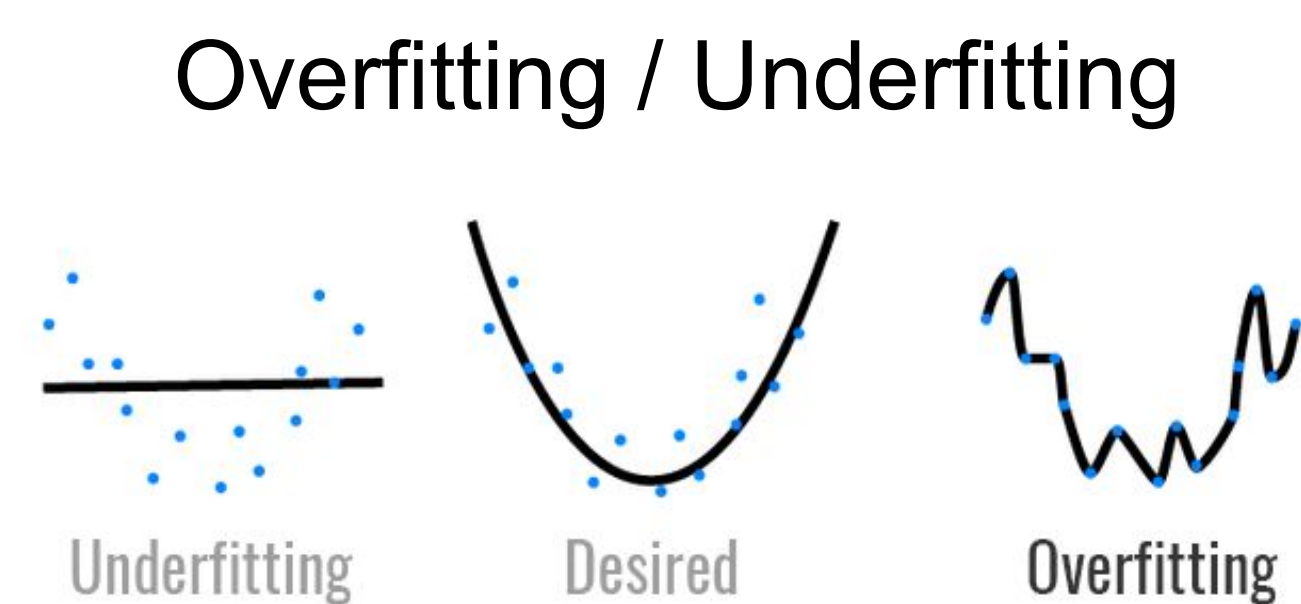
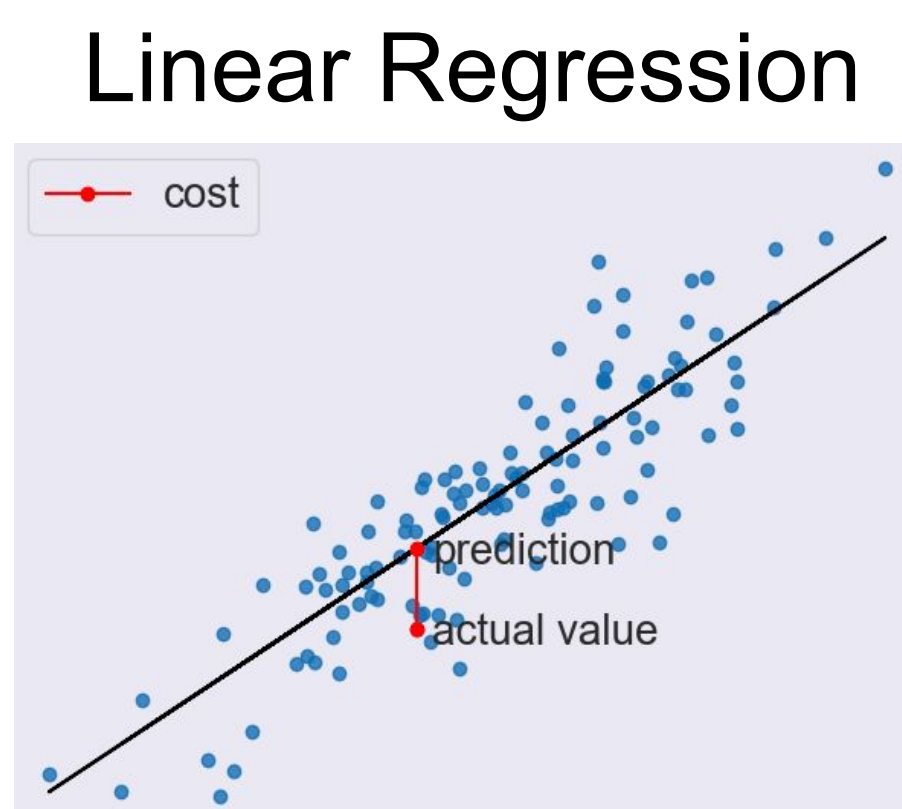
Python4ML

A Machine Learning Course for Everyone

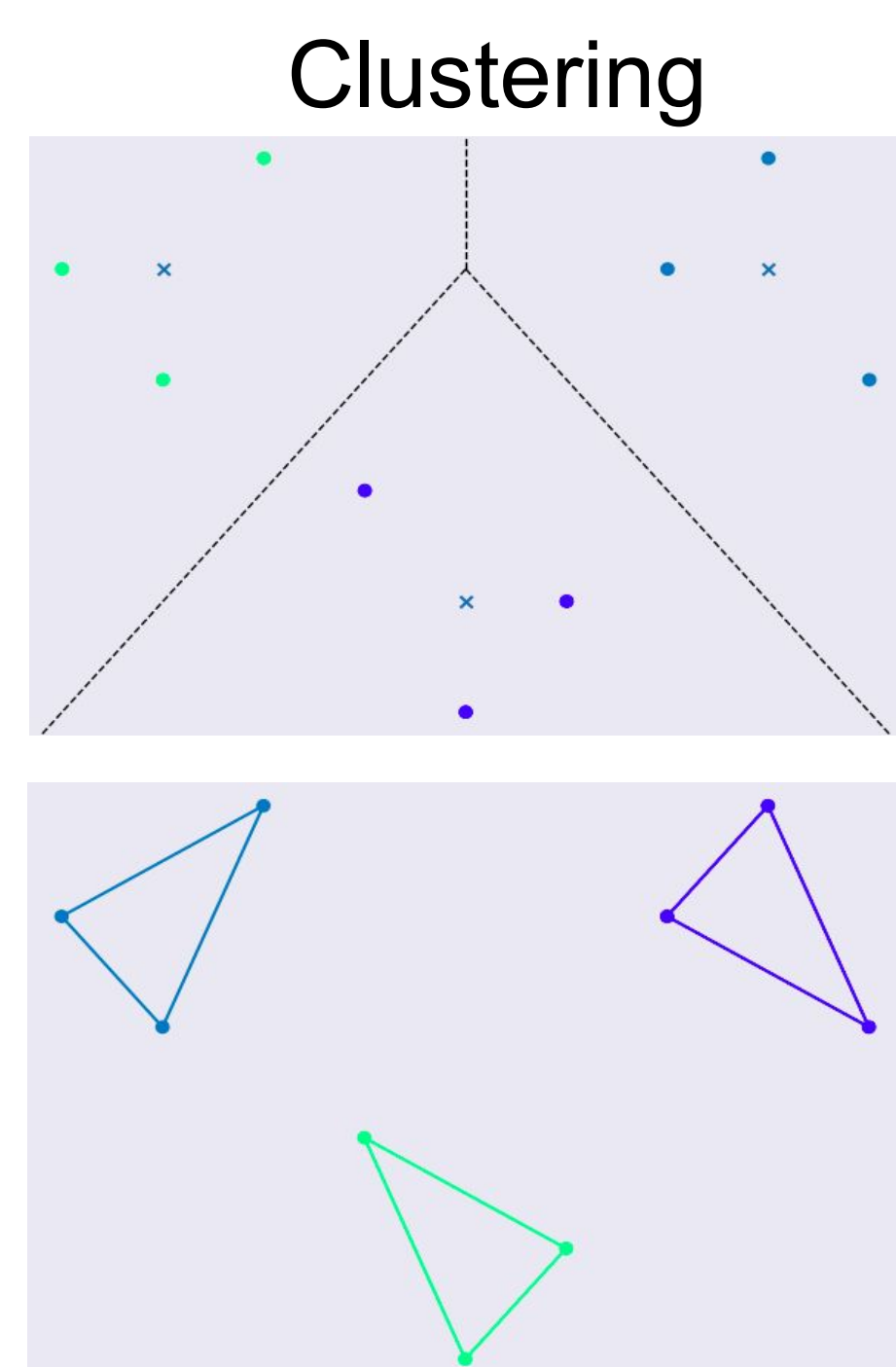
Topics



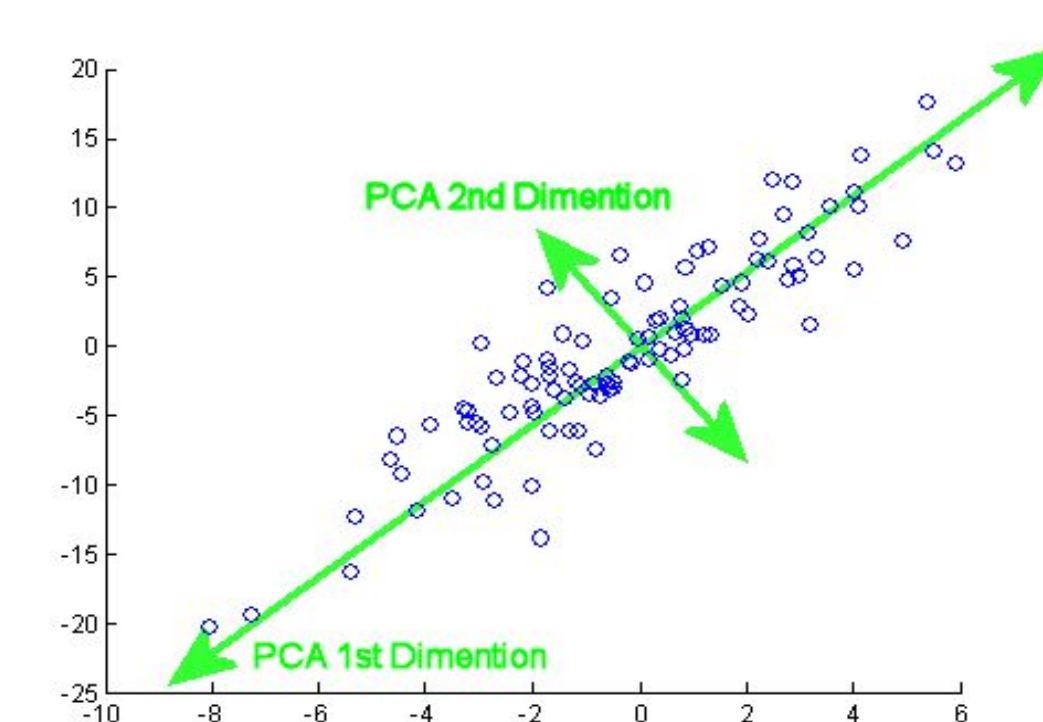
Overview



Unsupervised Learning



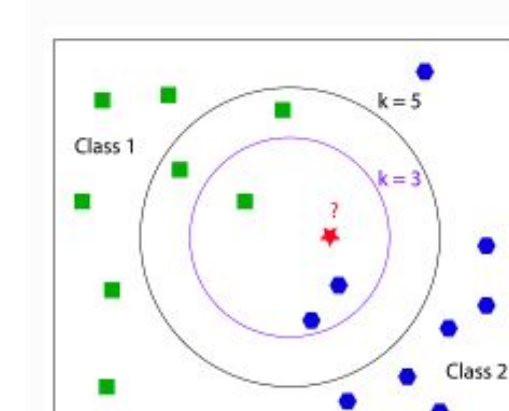
Principal Component Analysis



K-Nearest Neighbors

k-Nearest Neighbors

K-Nearest Neighbors (KNN) is a basic classifier for machine learning. So we are trying to identify what class an object is in. To do this we look at the closest points (neighbors) to the object and the class with the majority of neighbors will be the class that we identify the object to be in. The k is the number of nearest neighbors to the object. So if k = 1 then the class the object would be in is the class of the closest neighbor. Let's look at an example.



Ref: <https://caxdocs.org>

WHAT?

Python4ML is an open-source course for machine learning using the Python programming language.

WHO?

This course is being developed with Virginia Tech's Open Source for Science organization, led by our client Amirsina Torfi.

WHY?

The course is aimed at those with little knowledge of machine learning. We want to facilitate education in an open-source context, bringing important topics together in a high-level overview of ML.

HOW?

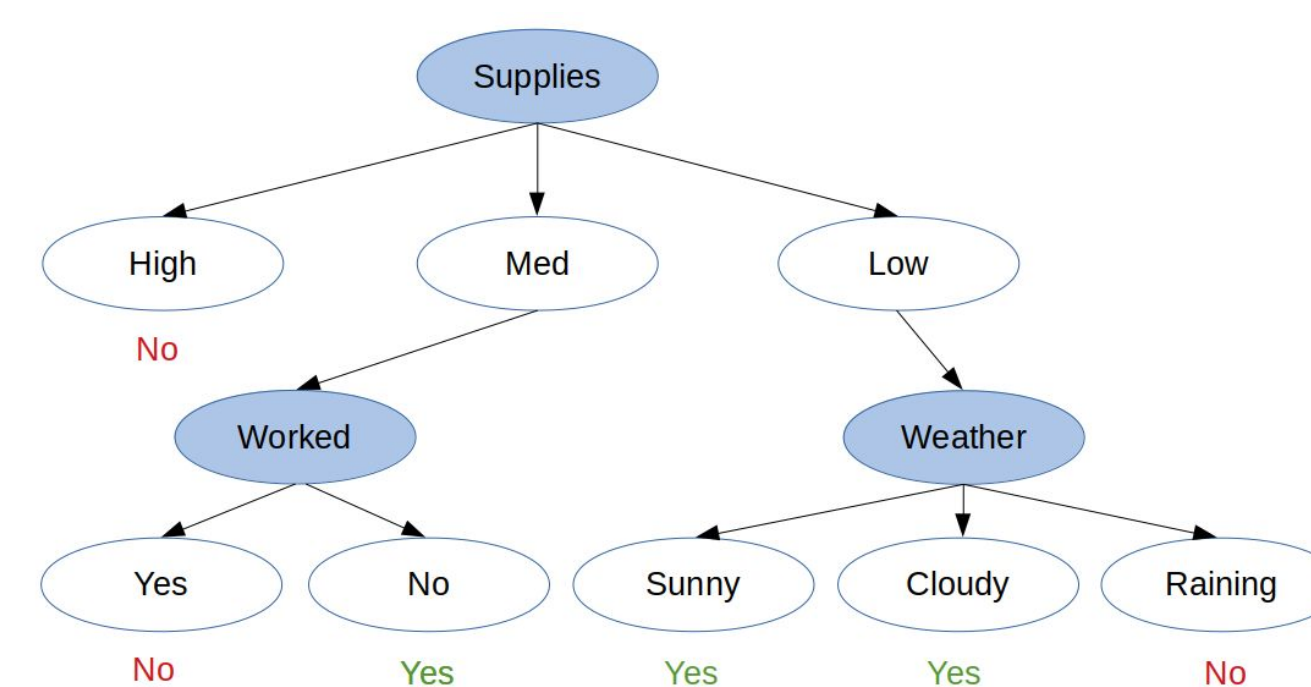
The course is made up of reStructuredText documents and example programs written in Python, using libraries such as scikit learn and matplotlib.

WHERE?

The course is available on GitHub. It can be found by scanning the QR code on this poster.

Supervised Learning

Decision Trees



Logistic Regression

When to use it

Logistic regression is great for situations where you need to classify between two categories. Some good examples are accepted and rejected applicants and victory or defeat in a competition. Here is an example table of data that would be a good candidate for logistic regression.

Studying	Hours	Focused	Success
1	False	False	False
3	True	True	True
0.5	True	False	False
2	False	True	True

Notice that the student's success is determined by the inputs and the value is binary, so logistic regression will work well for this scenario.

Naive Bayes

What is it?

Naive Bayes is a classification technique that uses probabilities we already know to determine how to classify input. These probabilities are related to existing classes and what features they have. In the example above, we choose the class that most resembles our input as its classification. This technique is based around using Bayes' Theorem. If you're unfamiliar with what Bayes' Theorem is, don't worry! We will explain it in the next section.

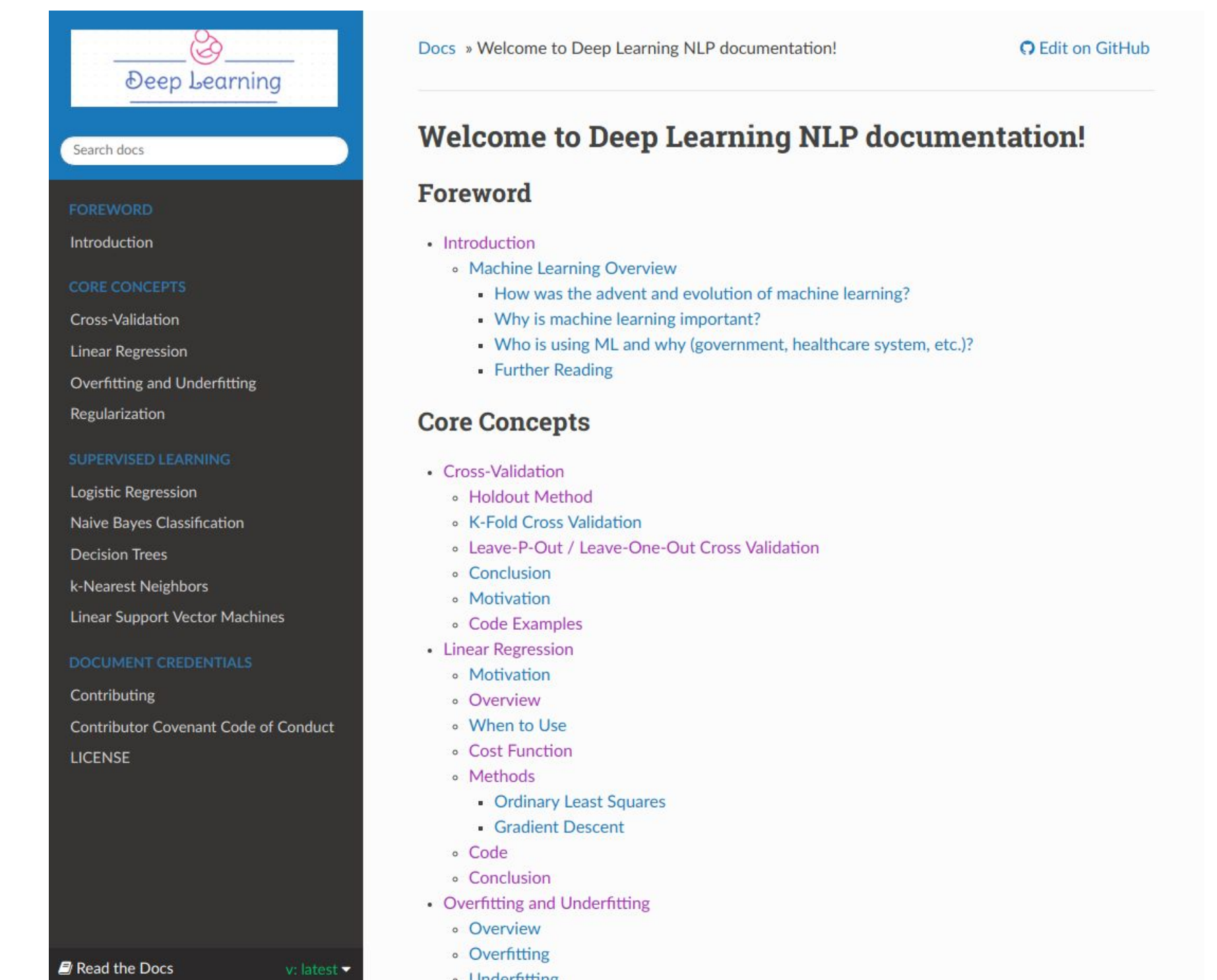
Bayes' Theorem

Bayes' Theorem [Equation 1] is a very useful result that shows up in probability theory and other disciplines.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Equation 1. Bayes' Theorem

Live Course Site



Technologies



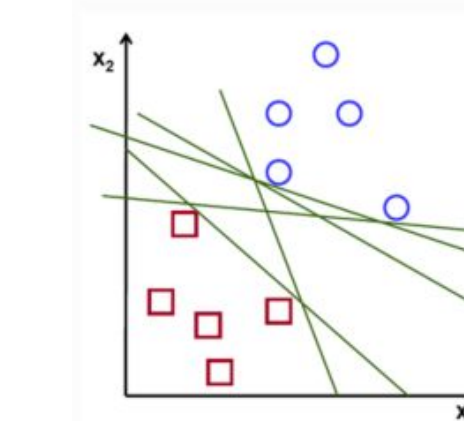
Support Vector Machines

Hyperplane

A hyperplane depends on the space it is in, but it divides the space into two disconnected parts. For example, 1-dimensional space would just be a point, 2-d space a line, 3-d space a plane, and so on.

How do we find the best hyperplane/line?

You might be wondering that there could be multiple lines that split the data well. In fact, there is an infinite amount of lines that can divide two classes. As you can see in the graph below, every line splits the squares and the circles, so which one do we choose?



Ref: <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fc47>

Learn More

Team: James Hopkins, Brendan Sherman, Zachery Smith, and Eric Wynn

Client: Amirsina Torfi, Head of Open Source for Science @ VT

4/30/2019
CS4624
(Multimedia, Hypertext, and Information Access)
Instructor: Edward A. Fox



Scan me