The Vapnik-Chervonenkis theory

Abdelwahid Benslimane
wahid.benslimane@gmail.com

Abdelwahid Benslimane
wahid.benslimane@gmail.com

The Vapnik-Chervonenkis theory is a fundamental theory of machine learning. It was developed by Vladimir Vapnik and Alexey Chervonenkis and contains important concepts such as the Vapnik-Chervonenkis dimension among other things.

In this document, I propose to introduce this theory in a simple way, and staying on the surface. Internet is full of documents allowing to deepen all the notions explained here.

Abdelwahid Benslimane
wahid.benslimane@gmail.com

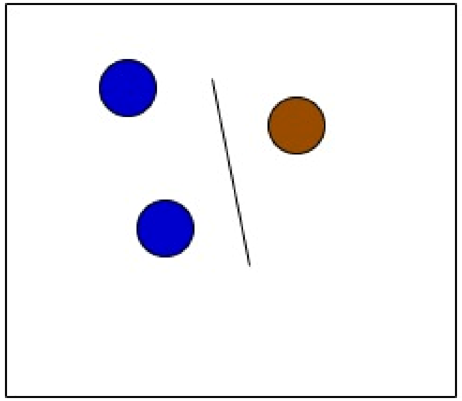
- The Vapnik-Chervonenkis dimension (VC dimension) is a measure of the separating power (complexity) of a set of functions $\mathcal{F}$. It is an integer number attached to this set $\mathcal{F}$

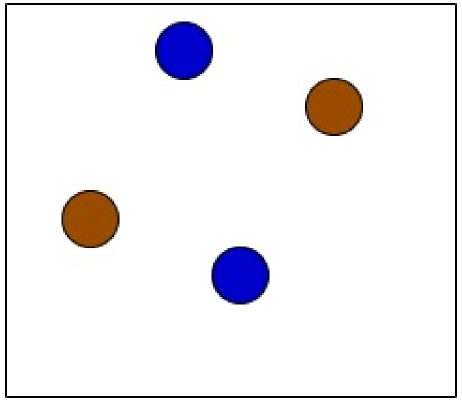
- Before illustrating a little better what the VC dimension is, let us start with a quick definition. If we take the example of a binary classification, for a sample of n points $(x_1, \dots, x_n)$ of R^p, there are 2^n different ways to separate it into two subsamples. We say that a set $\mathcal{F}$ of functions $f(X, w)$ shatters the sample if the 2^n separations can be made by different $f(X, w)$ from the set $\mathcal{F}$

Abdelwahid Benslimane
wahid.benslimane@gmail.com

A set of functions of $R^p \rightarrow R$ has a VC dimension h if :

- There exists a set of h points of R^p that can be shattered, whatever the labelling of the points
- No set of $h+1$ points can be shattered by this set of functions





Example

In 2-D, linear functions (straight lines) can shatter 3 points, but not 4 (in the diagram on the right, no straight line can separate the blue points from the red points).

Abdelwahid Benslimane
wahid.benslimane@gmail.com

A few examples:

- The VC dimension of the set of hyperplanes of R^p is p+1
- The VC dimension of the set of functions $f(X, w) = \mathrm{sign}\big(\sin(w.\mathrm{x})\big)$, with a free parameter w, is infinite!

Abdelwahid Benslimane
wahid.benslimane@gmail.com

- The learning process can generalize well if and only if the set the model belongs to has a finite VC dimension h. A finite VC dimension not only guarantees the generalization, but it is THE ONLY WAY that allows generalization to occur

- The difference between the training error/empirical risk R_{emp} and test error/theorical risk R depends on the ratio between the VC dimension, h, and the size of the of the training set, n. With probability 1- α:

$$R < R_{emp} + \sqrt{\frac{h.\left(ln\left(\frac{2n}{h}\right) + 1\right) - ln\left(\frac{\alpha}{4}\right)}{n}}$$

(Vapnik-Chervonenkis inequality)

Abdelwahid Benslimane
wahid.benslimane@gmail.com

- The Vapnik-Chervonenkis inequality allows us to give an upper bound to the theoretical risk. Moreover, it does not involve p (dimension of the data space) but the VC dimension h

- If two sets of models explain the data with equal quality, then the set with the lower VC dimension should be preferred.

- If two models explain the data with equal quality, then the one coming from a set with a lower VC dimension has a better generalization performance

Instead of observing differences between models, it is better to control them...