VC维

VC维（Vapnik-Chervonenkis 维）是一种衡量给定函数集（假设集）学习能力的指标，用于计算机学习理论中。它是由Vladimir Vapnik和Alexey Chervonenkis提出的。

给定一个假设集 H 和一个集合 S 包含 n 个点，如果存在2^n种不同的方式通过假设集 H 将 S 中的点划分为两类。即对于 S 的每一种二分法，都存在一个假设在 H 中可以实现这种二分，那么我们说 H 可以打散 S 。

上面是通过对S讨论的，即存在一个包含n个点的S集合，我们到底需要多大的一个假设集H来彻底二分S呢，答案是很明显的，就是需要有一个H假设集可以找到2^n种不同的方式将S二分，那么为什么是2^n呢，那不就是n个点之间的两两组合嘛。

VC维定义为假设集 H 可以打散的最大集合的大小。形式上，如果存在大小为 d 的最大集合被 H 打散，但任何大小为 d+1 的集合都不能被 H 完全打散，那么假设集 H 的VC维是 d 。

上面讨论的是H假设集，假如给定一个假设集，这个假设集就是确定的，那么它固有的性质就包括了去打散别的集合，但由于H集合已经被确定了，所以它就一定存在一个最大的能打散的集合S，假如这个集合S的大小是d，那再大一点的集合S也没法被H完全打散了，那这个时候我们找到的H的VC维就是d了。

用公式表示，如果 H 的VC维是 d ，则有：

1. 对于任何大小为 d 的集合 S , H 可以打散 S
2. 对于任何大小大于 d 的集合 S', H 不能打散 S'

VC维是理解机器学习模型的复杂性和泛化能力的一个重要概念。一个模型的**VC维越高**，意味着它在训练数据上的**拟合能力越强**，但同时可能也意味着它对新数据的**泛化能力较差**（过拟合的风险）。

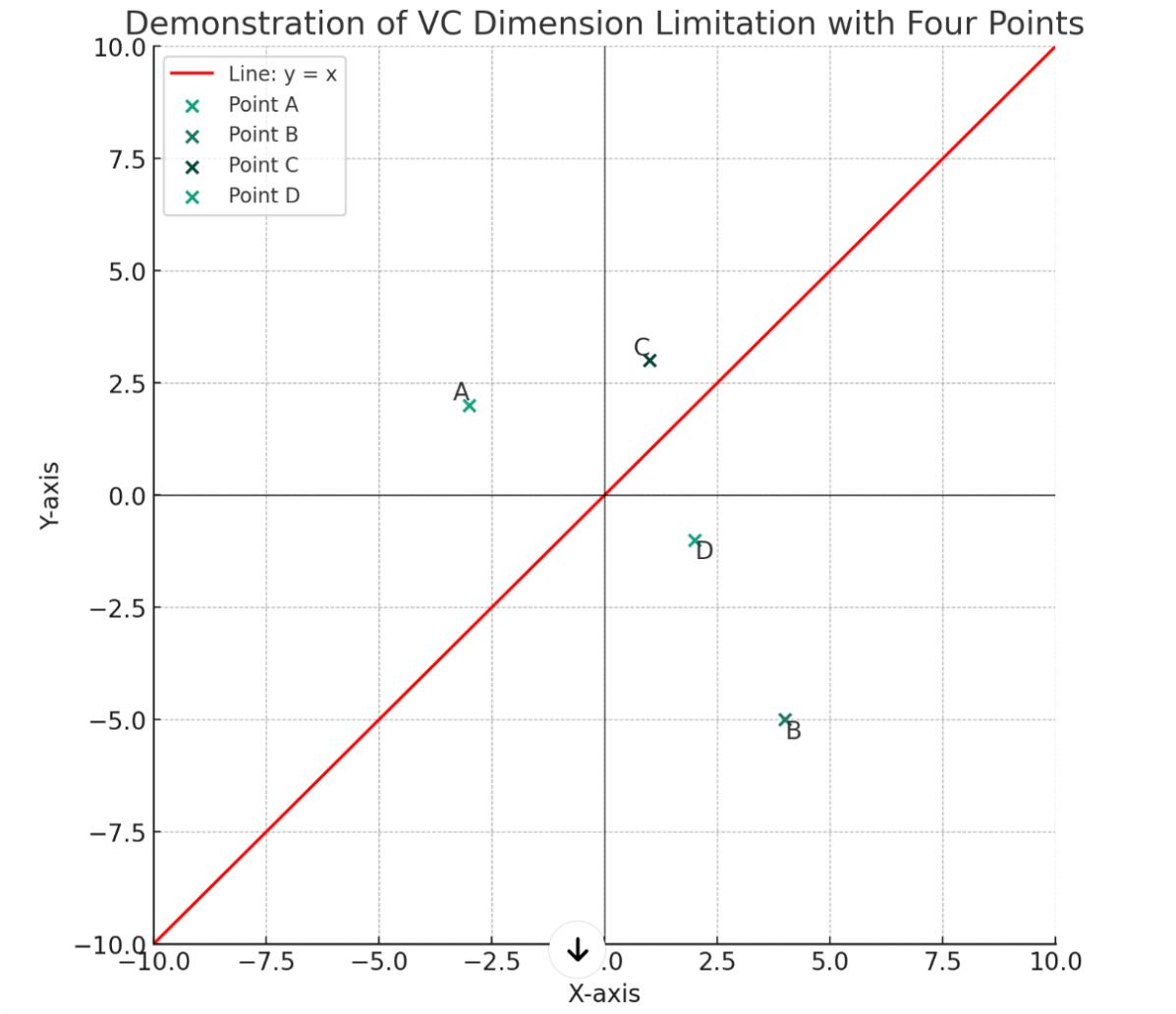
那就说明假设集H的拟合能力很强，毕竟它对于大量数据S的集合都可以被打散，那就说明H对于已有的训练数据S的拟合已经做到很好了，但是就是因为划分的这么细致，很有可能在新数据加入进来的时候会过拟合产生一些误差。

举例，二维平面上的一条直线的vc维为多少?

在二维平面上，一条直线的VC（Vapnik-Chervonenkis）维是3。对于一条直线，在二维空间中，它能够完美地将任意三点分开（假设这三点不共线），但无法对任意四点做到这一点。因此，它的VC维是3。

注意，这里的三个点都是各自有标签的，这条直线需要把一个标签的数据都划分到一类，另一种标签的点划分到另一类，这里并不是那么简单的普通划分

假如有四个点，但是四个点的标签分布



注意这里的C无论怎么分都没法单独分出来，就是没有办法被打散