<https://www.cnblogs.com/xyou/p/9197523.html>

机器学习中，需要进行特征数字化的操作。例：

性别特征：["男"，"女"]

祖国特征：["中国"，"美国，"法国"]

运动特征：["足球"，"篮球"，"羽毛球"，"乒乓球"]

某个人的特征是["男","中国","乒乓球"]，该如何编码？

One-hot编码：一位有效编码，用N位状态寄存器对N个状态进行编码，每个状态都有它独立的寄存器为，并且任意时候只有一位有效。

对于上述的例子：

性别特征有两个：

男 => 10

女 => 01

祖国特征有3个：

中国 => 100

美国 => 010

法国 => 001

运动特征有4个：

足球 => 1000

篮球 => 0100

羽毛球 => 0010

乒乓球 => 0001

["男","中国","乒乓球"]编码为：[1 0, 1 0 0, 0 0 0 1]