「one-hot编码」

· 概念：也称一位有效编码，主要是采用N位状态寄存器来对N个状态进行编码，每个状态都由独立的寄存位寄存，并在任意时刻只有一位有效
one-hot是分类变量作为二进制向量的表示。

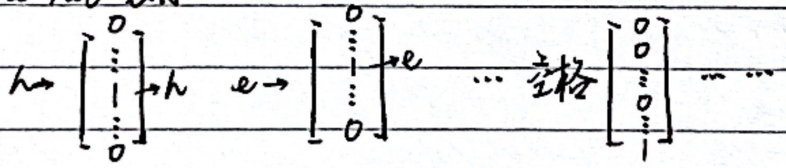
· 例子

1. → 编码对象：['中国','日本','美国','中国']
→ 确定分类变量：中、美、日 三类
→ 特征整数编码并按大小排列：中-0，美-1，日-2

⇒ one-hot表示
[[100],[001],[010],[010]]

2. → 编码对象：hello world
→ 分类变量和整数编码：a-0，b-1，c-2，d-3，…，z-25 空格-26

⇒ one-hot表示



· 优点：one-hot是将类别变量转换为机器学习算法易于利用的一种形式
· 缺陷：要求每个类别之间相互独立，如果之间存在某种连续型的关系，或许使用分布式(distributed representation)更合适