L1范数与L2范数区别：

平方L2范数在数学和计算上都比L2范数本身更方便。例如，平方L2范数对x中每个元素的导数只取决于对应的元素，而L2范数对每个元素的导数却和整个向量相关。但是在很多情况下，平方L2范数也可能不受欢迎，因为它在原点附近增长得十分缓慢。在某些机器学习应用中，区分恰好是零的元素和非零但值很小的元素是很重要的。在这些情况下，我们转而使用在各个位置斜率相同，同时保持简单的数学形式的函数：L1范数。

行列式等于矩阵特征值的乘积。行列式的绝对值可以用来衡量矩阵参与矩阵乘法后空间扩大或者缩小了多少。如果行列式是 0，那么空间至少沿着某一维完全收缩了，使其失去了所有的体积。如果行列式是 1，那么这个转换保持空间体积不变。

协方差与独立性：两个变量如果协方差为零，它们之间一定没有线性关系。独立性是比零协方差的要求更强，因为独立性还排除了非线性的关系。两个变量相互依赖但是具有零协方差是可能的。

在具有相同方差的所有可能的概率分布中，正态分布在实数上具有最大的不确定性。因此，我们可以认为正态分布是对模型加入的先验知识量最少的分布。

信息论的基本想法是一个不太可能的时间居然发生了， 要比一个非常可能的时间发生能提供更多的信息。

KL散度衡量对于同一个随机变量的两个单独的概率分布P（x）和Q（x）之间的差异。非负，不对称。

香农熵： 也记做H（P）

KL散度：

交叉熵：