内积矩阵与协方差矩阵

一、内积矩阵

假设维空间单个样本,用列向量表示，即

由个样本组成的样本矩阵表示为(因为编程时一般习惯array的第一个维度为样本数), 即

这里有如下两种观点，从行的角度，每一行代表了1个样本，其有n个特征，总共有m个这样的样本。从列的角度，每一列代表了1个特征的m份抽样，总共有n个这样的特征。从这两个角度出发，分别引出了样本的内积矩阵和协方差矩阵。

下面，先介绍样本的内积矩阵。内积矩阵是从行的角度，研究两两样本的相似性。由于有m个样本，不难得知这样的相似度结果共有个。如果将其用矩阵表示，那应该是维。

如何表征样本的相似性呢？我们知道单个样本是由n个特征表示的n维列向量，而向量的内积正是刻画两个向量相似度的一种手段。

也就是说，最终的结果应该应该是下面这个样子。

即样本的内积矩阵可以表示为，

内积矩阵的每一个元素表示了样本和样本的相似度(内积)。

二、协方差矩阵

接下来，从列的角度(特征)研究协方差矩阵。用表示样本矩阵的第列。此时样本矩阵表示如下，

不失一般性，假设样本的每一个特征都已经被中心化，即样本矩阵的每一列中的元素都减去这一列元素的均值。以第一列为例子，

其中

根据数理统计可知，两个**随机变量**(在这里随机变量就是样本的特征)的协方差为：

由于，我们已经将特征中心化(即),

对其进行次抽样，可以得到**样本**的协方差

下面将对样本的个特征，分析其两两的协方差。不难得知，样本的协方差矩阵应该为维。

为了方便，我们将样本矩阵表示如下：

仿照内积矩阵的推导，可以得知样本的协方差矩阵为

也就是说，

三、总结

总结一下，样本矩阵有两种理解角度。从行的角度，也就是单个样本的角度，可以研究样本的相似性，推导出样本的内积矩阵。

从列的角度，也就是单个特征的角度，可以研究特征之间的相关性，推导出样本的协方差矩阵。

注意，不要死记硬背，具体使用公式时可以从维度一致的角度推导。