**第十六章读书报告**

09118223 吴亦珂

读书进度：读到16.1

1. **问题列表**

（我提出）再由协方差的性质下面那一行的推导如何推出的，感觉没怎么看懂？

讨论结果：主要是为基本单位向量，并且第i个分量为1，其余为0。所以说中间的协方差矩阵与相乘之后只剩下。而只剩下。所以最后相乘的结果是。

（我提出）书上前面说可以由线性表示，但是上说可以由线性组合表示，这个应该怎么推导？为什么可以用线性组合表示就可以说相关系数的平方和为1？

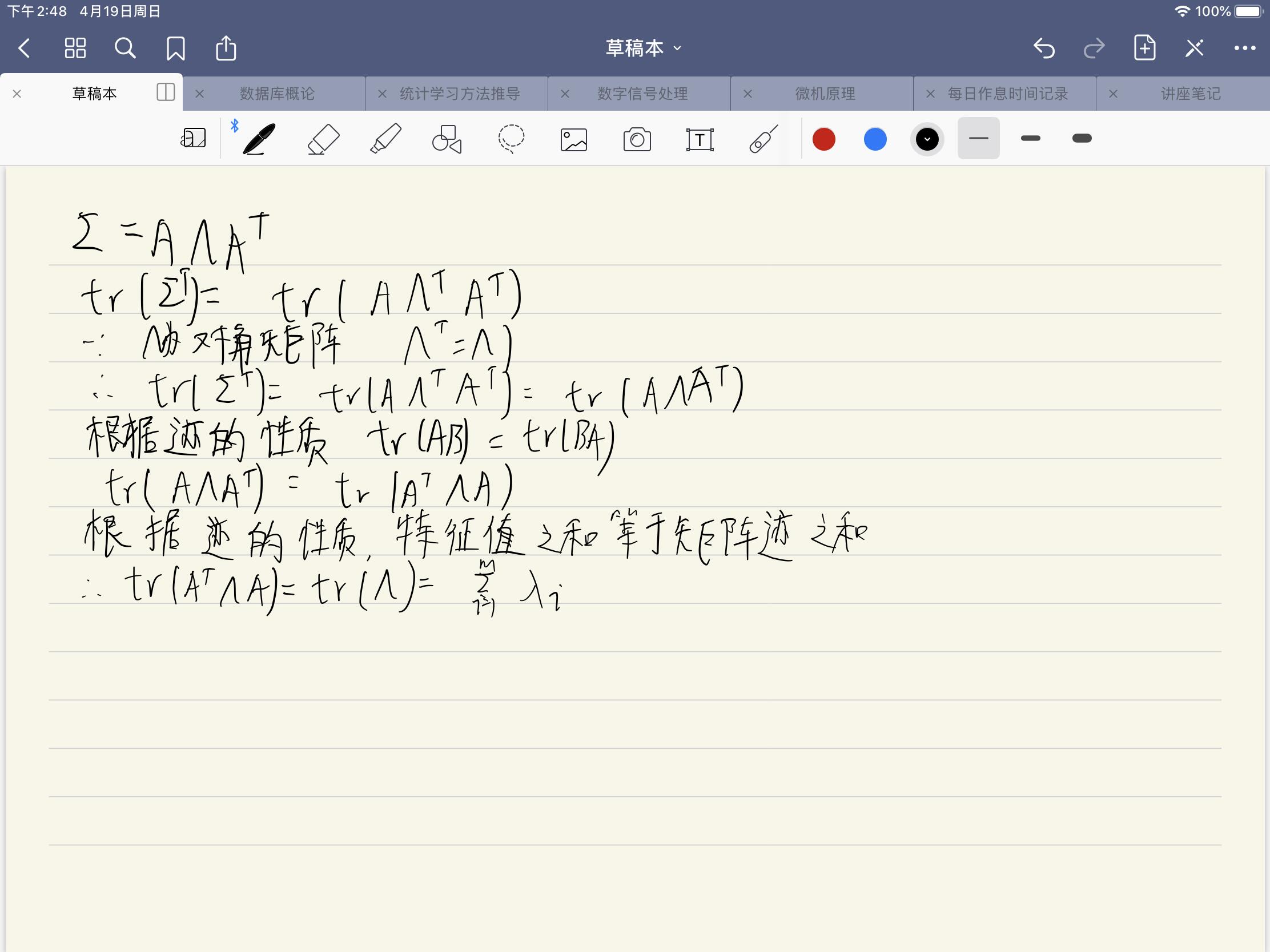
讨论结果：若将用X矩阵表示，用Y矩阵表示，前面系数写成矩阵A的形式，则可知Y=AX，所以，又可以知道。因此可以由线性组合表示。因为二者可以线性表示，所以两者的相关系数为1.又因为，，……互不线性相关，所以可以表示为和的形式。因此可知16.22成立。

（别人提出）为什么主成分分析时需要对数据进行规范化，使其方差为1？

讨论结果：因为如果不进行规范化，不同变量的量纲可能相差较大。如果对人的体育进行预估。跳远以米为单位，数值一般为各位数，而跳绳个数可能有几百个。而进行主成分分析时，要选取方差最大。那么易知，跳绳个数数值很大，因而方差较大，对结果影响较大。所以，这就违背了算法的思想。因而要先进行规范化才好操作。同时，这么做也使得在进行算法处理时，运算速度更快，方便处理。

（别人提出）证明x的方差和等于y的方差和时用了矩阵的迹的什么性质？

讨论结果：



（别人提出）主成分分析和SVM有什么关联吗？是否可以从奇异值分解这个角度考虑主成分分析是如何推导的？  
主成分分析是找出方差最大的方向，将数据映射到新的空间当中，使得原有数据的维度降低。SVM是找出一个最优平面将两者分开。SVM的核函数实际上是将原始数据映射到了一个高维空间中，使其变得线性可分。PCA其实是可以由特征值分解实现的。

1. **下周读书计划**

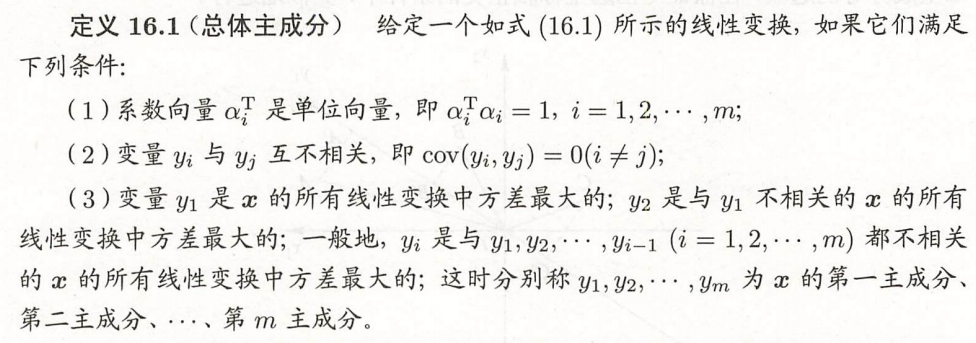
下周计划完成第16章内容。后面设计样本主成分分析的内容，应该仔细阅读，相加推导。

1. **读书收获**

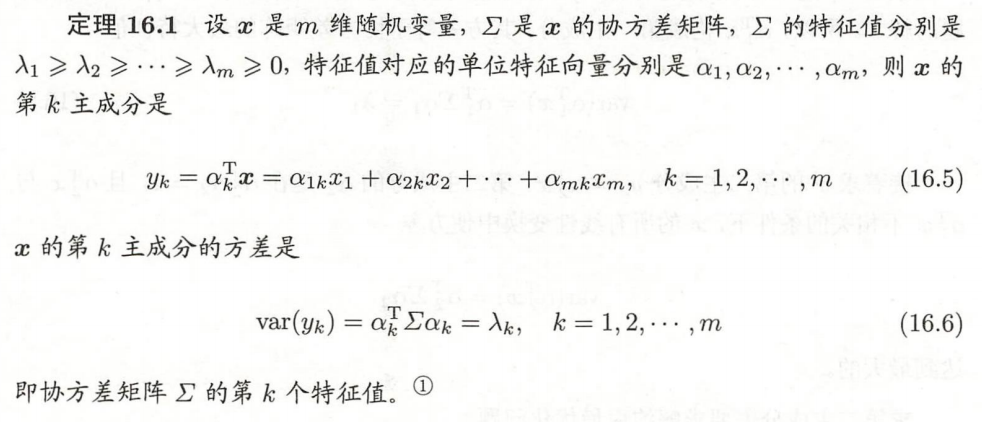
16.1主要介绍了主成分分析的基本思想、定义、性质以及计算的方法。

1. 基本想法

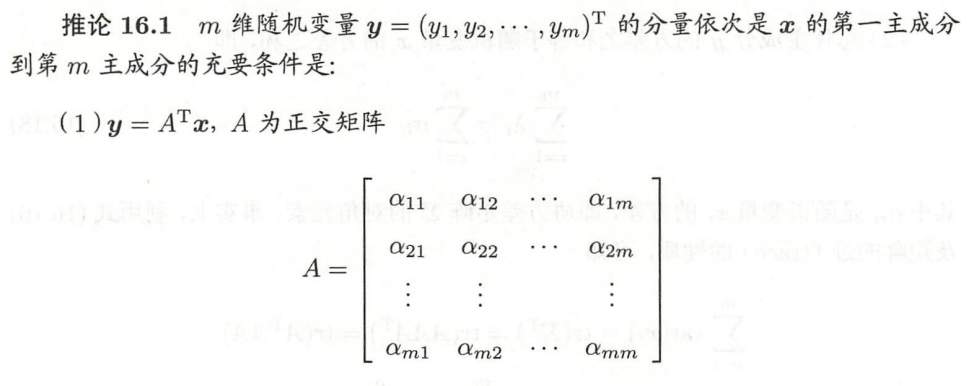
考虑到原有数据各个变量之间总是有相互的关系，所以将数据投影到新的坐标轴的新的坐标系中。使得数据的方差尽可能大，直观来说，就是使数据尽可能拉得更开。同时各个变量线性无关

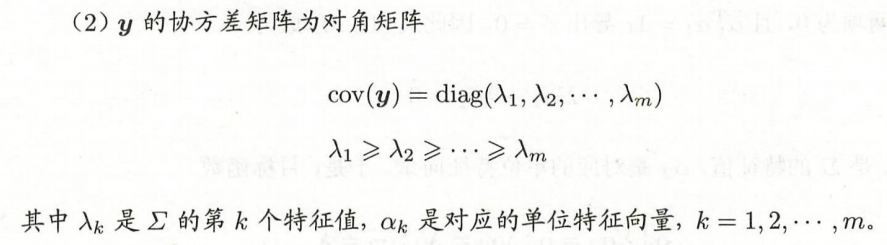
1. 定义
2. 性质

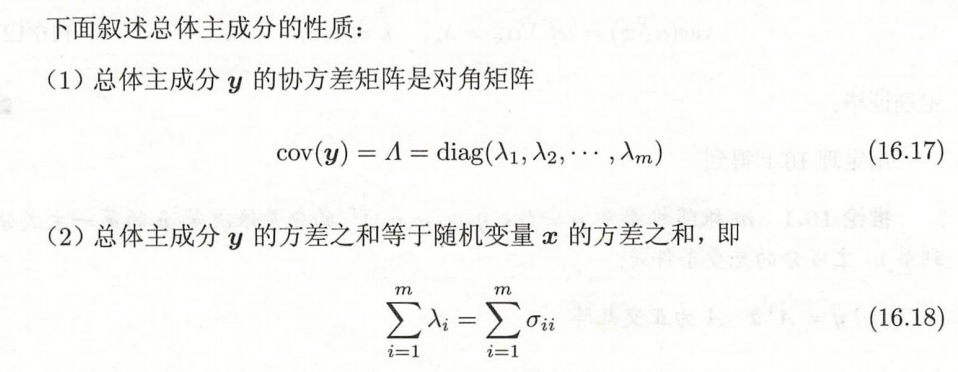
由PCA的定义，可以得出其性质以及计算PCA的方法

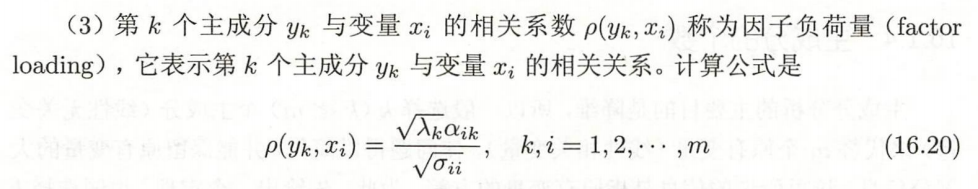


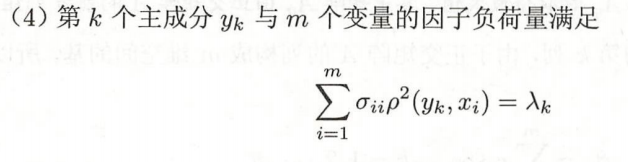
推论：

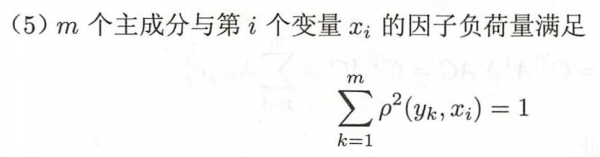




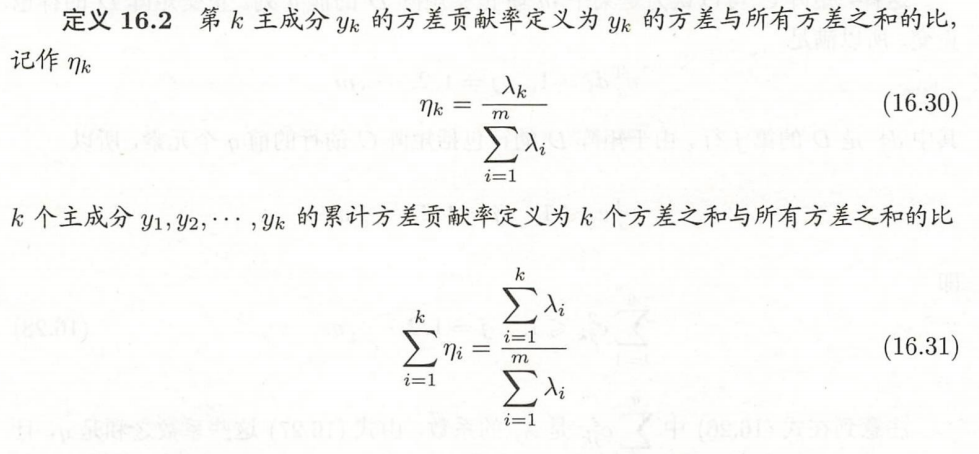
具体性质：







1. 选择k个主成分的方法:



1. 由于量纲不同，所以在进行PCA时应该先进行正则化，使其方便处理。处理完后，对于规范化变量的总体主成分同样具有以下性质：

