读书报告

1. （必填）读书计划

1、本周完成的内容章节：《统计学习方法》第十六章

2、下周计划：《统计学习方法》第十七章

1. 读书摘要
   * 1. 主成分分析是一种常用的无监督学习方法，这一方法利用正交变换把由线性相关变量表示的观测数据转换为少数几个由线性无关变量表示的数据，线性无关的变量称为主成分。主成分的个数通常小于原始变量的个数，所以主成分分析属于降维方法。
     2. 主成分分析中，首先对给定数据进行规范化，使得数据每一变量的平均值为0 ，方差为1 。之后对数据进行正交变换，原来由线性相关变量表示的数据，通过正交变换变成由若干个线性无关的新变量表示的数据。新变量是可能的正交变换中变量的方差的和〈信息保存〉最大的，方差表示在新变量上信息的大小。将新变量依次称为第一主成分、第二主成分等。
     3. 主成分分析的主要目的是降维，所以一般选择k(k<<m)个主成分（线性无关变量）来代替m个原有变量（线性相关变量），使问题得以化简，并能保留原有变量的大部分信息。这里所说的信息是指原有变量的方差。对任意正整数q，1<=q<=m,考虑正交线性变换，其中y是q维向量， 是qxm矩阵，令y的协方差矩阵为，则的迹在B=时取得最大值，其中矩阵由正交矩阵A的前q列组成。
     4. 假设x为m维随机变量，其均值为，协方差矩阵为，考虑由m维随机变量x到m维随机变量y的线性变化，，其中=(，，，…，)。

如果该线性变换满足以下条件，则称之为总体主成分：

1. =1,i=1,2,3,…,m
2. = 0 (i不等于j)
3. 变量是x的所有线性变换中方差最大的；是与不相关的x的所有线性变换中方差最大的；一般地，是与，，…，都不相关的x的所有线性变换中方差最大的；这时分别称，，…，为x的第一主成分、第二主成分、…、第m主成分