# PCA神经网络

1 零均值化

零均值化就是求每一列的平均值，然后该列上的所有数都减去这个均值。

2 求协方差矩阵

M行n列，每行代表一个样本，则协方差是n行n列，

[1 2

2 4]

代表第一个特征和第一个特征之间的协方差为1 , 代表第一个特征和第二个特征以及反过来，他们之间的协方差为2，

3 求特征值，特征矩阵

特征值代表矩阵运动的速度，特征矩阵代表变化之后的矩阵，此处特征值代表方差信息量，几乎没有波动的数据，特征值就会比较小

4 保留主要的成分

取前k个数据，k要取多少，则按公式计算比例

Sigma(k,lambda)/sigma(n,lambda)>=percentage

Percentage取99%

5 PCA主要用于特征比较多的情况，如果没有多少特征就不用

6