

PCA算法

xhj

November 7, 2018

1 最近重构性

最近重构性要求样本点到降维后的超平面的距离之和最短，以此来确定该超平面。

2 最大可分性

2.1 基于特征值和特征向量求主成分

1、中心化（均值化）。目的是为了更方便之后的求解。因为需要求解协方差矩阵，协方差的公式为 $cov(x, y) = E\{(x - Ex)(y - Ey)\}$ 。利用中心化之后的矩阵就可以直接求协方差。

2、求协方差矩阵。协方差矩阵的维数由数据的特征的数目决定。

$$cov = \begin{pmatrix} cov(x_0, x_0) & cov(x_0, x_1) & \dots & cov(x_0, x_n) \\ cov(x_1, x_0) & cov(x_1, x_1) & \dots & cov(x_1, x_n) \\ \dots & \dots & \dots & \dots \\ cov(x_n, x_0) & cov(x_n, x_1) & \dots & cov(x_n, x_n) \end{pmatrix} \quad (1)$$

通过求解协方差矩阵，可以观察样本中各个特征之间的相关度。通过协方差也可以求解相关系数 ρ ， $\rho = \frac{cov(x, y)}{\sqrt{D(x)D(y)}}$ 。

3、求解协方差的特征值和特征向量。