

机器学习第五次作业

2020年11月

1 第一题

令 \mathbf{X} 是一个均值中心化的数据矩阵，即 \mathbf{X} 的样本均值 $\bar{\mathbf{x}} = \frac{1}{n} \sum_i \mathbf{x}_i = 0$ 。

- (a) 证明数据到一个任意向量 \mathbf{u} 一维投影的方差等于 $\mathbf{u}^\top \mathbf{C} \mathbf{u}$ ，其中 $\mathbf{C} = \frac{1}{n} \sum_i \mathbf{x}_i \mathbf{x}_i^\top$ 是样本协方差矩阵。
- (b) 证明 $k=1$ 的PCA将数据投影到最大方差的方向。

2 第二题

拉普拉斯特征映射算法目的是找到一种低维表示，用来最好地保持通过权重矩阵 \mathbf{W} 衡量的近邻关系，该算法工作原理如下：

- (1) 找到每个数据点的 t 个最近邻点。
- (2) 构造 \mathbf{W} ，1个 $m \times m$ 的稀疏对称矩阵，其中如果 (x_i, x_j) 是近邻，则 $\mathbf{W}_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / \sigma^2)$ ，否则等于0。
- (3) 构造对角矩阵 \mathbf{D} 使得 $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$ 。
- (4) 通过最小化近邻点之间的加权距离，找到 k 维表示：

$$\mathbf{Y} = \underset{\mathbf{Y}'}{\operatorname{argmin}} \sum_{i,j} \mathbf{W}_{ij} \|\mathbf{y}'_i - \mathbf{y}'_j\|_2^2 \quad (1)$$

下面我们假定 $k = 1$ ，我们想要寻找一个一维表示 \mathbf{y} 。证明式子 (1) 等价于 $\mathbf{y} = \underset{\mathbf{y}'}{\operatorname{argmin}} \mathbf{y}'^\top \mathbf{L} \mathbf{y}'$ ，其中 $\mathbf{L} = \mathbf{D} - \mathbf{W}$ 是图的拉普拉斯矩阵。