

模式识别导论第五次作业——简述题

杨登天-202028015926089-微电子研究所

1. 在短文本分析和图像匹配应用中，使用的典型特征提取方法分别是什么？

解答：短文本分析的典型特征提取方法是 TF-IDF 方法和 Word2Vec 方法。

图像匹配应用中，典型特征提取方法是 SIFT、CNN 和视觉词袋。

解答完毕

2. 简述 PCA 和 LDA 的异同

解答：

相同点：

- 1) 都利用样本的特征将高维的样本处理到低维，缓解维数灾难，减少计算量。
- 2) 都降低了问题复杂度。
- 3) 都减少了数据噪声的影响。
- 4) 都涉及有效特征的选取和无效特征的剔除。
- 5) 都减少了特征之间可能存在的相关性，降低了分类器学习难度。

不同点：

- 1) PCA 针对无类别标签的样本，属于非监督学习方法，而 LDA 针对有类别的样本，属于监督学习方法。
- 2) PCA 的目的只是将数据以最大程度保持样本差异信息的条件下处理到低维，并没有专用目的，因此 PCA 在结合其他算法下可以被广泛应用于数据分析包括聚类、分类、可视化，数据去噪和人脸识别，而 LDA 以最大化数据的线性可分性将样本变换到低维空间后需要对样本分类，因此其直接目的是样本的分类。
- 3) PCA 寻找数据方差最大的投影方向，而 LDA 方法寻找 Fisher 分类准则最大的投影方向。

解答完毕

3. 给定均值为 d 维0向量的样本集合 $X = [x_1, x_2 \dots x_N] \in R^{d \times N}$ ，其中 d 是样本特征的个数， N 是样本个数，即， $\sum_{i=1}^N x_i = 0$ ， x_i 是第 i 个样本。假定特征变换后数据的维度为 k ，分别写出使用PCA和KPCA对样本集合进行特征变换的流程。

解答：

问题一、使用 PCA 对样本集合进行特征变换的流程

步骤一、计算样本均值

$$\mu = \sum_{i=1}^N x_i = 0$$

步骤二、计算样本的离散度矩阵

$$S = \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T = \sum_{i=1}^N x_i x_i^T$$

步骤三、对离散度矩阵 S 进行特征值分解，取前 k 个最大的特征值对应的特征向量组成线性变换矩阵 W ，其中 W 的第 i 列对应第 i 个特征向量。

步骤四、对输入样本进行降维

$$y_i = W^T(x_i - \mu) = W^T x_i$$

问题二、使用 KPCA 对样本集合进行特征变换的流程

步骤一、确定核函数，多用 RBM（径向基函数），计算核矩阵 K

$$K = \begin{pmatrix} \kappa(x_1, x_1) & \kappa(x_1, x_2) & \dots & \kappa(x_1, x_N) \\ \kappa(x_2, x_1) & \kappa(x_2, x_2) & \dots & \kappa(x_2, x_N) \\ \dots & \dots & \dots & \dots \\ \kappa(x_N, x_1) & \kappa(x_N, x_2) & \dots & \kappa(x_N, x_N) \end{pmatrix}$$

步骤二、求解该核矩阵的特征值和特征向量，并按照特征值从大到小排列特征向量的矩阵 V ，其中每个特征向量表达为 v_i 。

步骤三、分别提取前 k 个主成分投影值。以下假定映射 $\phi(x)$ 可以使得 x 映射到高维，并且在核方法下无需知道该映射的具体内容。再假定 w 为特征矩阵 V 经过映射以后的还原到非线性变换空间内的特征向量，但事实上该特征向量无需求解。

$$w\phi(x) = \sum_{i=1}^N v_i \phi(x_i) \phi(x) = \sum_{i=1}^N v_i \kappa(x_i, x)$$

因此对于任意一个样本，可以通过上式提取前 k 个主成分投影值。

解答完毕

4. 根据特征选择与分类器的结合程度，特征选择方法可以分为哪三类？各有什么特性？

解答：根据特征选择与分类器的结合程度，可以分为过滤式特征选择、包裹式特征选择和嵌入式特征选择。

其中过滤式特征选择——先对数据集进行特征选择，然后再训练学习器，特征选择过程与后续学习器无直接关联；“选择”与“学习”独立

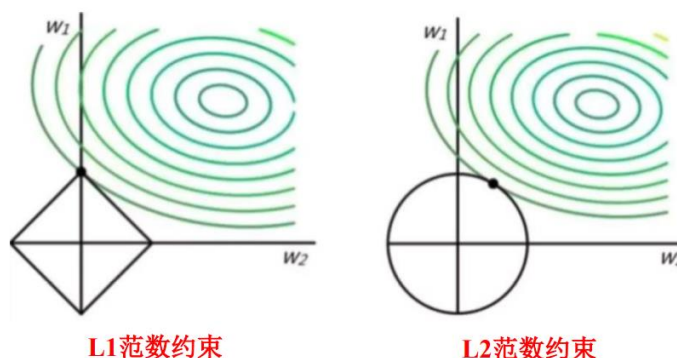
包裹式特征选择——以分类器性能作为特征选择的准则，为给定分类器“量身定制”特征子集；“选择”依赖“学习”

嵌入式特征选择——同时进行分类器学习和特征选择，在训练分类器时自动的进行特征选择。“选择”与“学习”同时进行

解答完毕

5. L1 范数和 L2 范数哪个适合用于做特征选择？简述原因。

解答：L1 范数更适合用于特征选择，有如下理由



- 1) 从 L1 范数的角度出发，L1 范数是一种稀疏约束，损失函数更容易收敛到某个或者某些特征为 0 的解，而特征为 0 则表明该特征相对于其他特征而言并不重要，因此起到特征选择作用。
- 2) 从 L2 范数的角度出发，从上图也能看出，相较于 L1 范数，损失函数更不容易收敛

到某个或者某些特征为 0 的解, 但是通过控制 L2 范数在目标函数中的比例可以实现特征的衰减, 从而使得解尽快收敛。但很难达到特征选取的目的。

解答完毕

6. 简述最优特征选择方法的基本思想。

解答:

参照樊彬老师提供的课件, 其基本思想可以表达为: 将所有可能的特征选择组合以树的形式表示, 采用分支定界方法对树进行搜索, 使得搜索过程尽早达到最优解, 而不必搜索整个树。

解答完毕

7. 处理模式识别高维数据的两种基本方法是什么? 它们有什么不同之处?

解答: 两种基本方法是特征变换和特征选择。特征变换和特征选择的不同之处

解答完毕