```
此类学习技术在训练阶段仅仅把样本保存起来,训练时间开销为零,待收到测试样本后再进行处理
                              懒惰学习
                                       代表算法: k近邻学习, 懒惰决策树
                                       在训练阶段就对样本进行学习处理的方法
                    方法分类
                              注意:一些学习方法既可以使用懒惰学习,也可以使用急切学习
                                       是一种常用的监督学习方法,也是"懒惰学习的代表",其关键是1.取合适的k值,2.合适的距离计算公式。
                                               1.给定测试样本,基于某种距离度量找出训练集中与其靠近的k个训练样本
                                               2.基于这k个"邻居"的信息进行预测
                                       工作机制
                      k近邻学习(kNN)
                                                在分类任务中可使用"投票法",即选择这k个样本中出现最多的类别标记作为预测结果,可使用加权投票
                                                在回归任务中可使用"平均法",即将这k个样本的实值输出标记的平均值作为预测结果,可使用加权平均
                                       注意:若k近邻学习退化为最近邻学习器的时候,在符合假设的情况下,其泛化错误不超过贝叶斯最优分类器的两倍
                                         在高维情况下样本稀疏很难实现"密采样",也就不能满足k近邻学习的假设,且高维空间还具有距离计算困难等问题。此类问题被称为"维数灾难",所以需要降维,又称"维数约减"
                                   背景
                                          很多时候,虽数据样本是高维的,但与学习任务密切相关的也许仅是某个低维分布,即高维空间中的一个低维"嵌入"
                                   是什么
                                                 一种经典的降维方法,要求原始空间中样本之间的距离在低维空间中得以保持
                                                目标:获得样本在d'为空间(d'《d)的表示,且任意两个样本在d'为空间中的欧式距离等于其在原始空间中的距离
                                                         1.通过降维前后保持不变的距离矩阵D求取内积矩阵B
                                                         2.对矩阵B做特征值分解,得到特征值组成的对角矩阵以及特征向量矩阵V
                                   多维缩放 (MDS)
                                                工作机制
                                                         3.取对角矩阵d'个最大特征值与相应特征向量组成Z变换矩阵(正交矩阵)
                                                                                                    \mathbf{Z} = 	ilde{\mathbf{\Lambda}}^{1/2} 	ilde{\mathbf{V}}^{\mathrm{T}} \in \mathbb{R}^{d' 	imes m}
                                                         4.得到向量在d'维中的表达
                                                       1.新空间中的属性是原空间中属性的线性组合,基于线性变化来惊醒将为的方法称为线性降维方法,它们都符合工作机制中4的表达式
                        低维嵌入
                                                注意
                                                      2.各种线性降维方法的不同之处是对低维子空间的性质有不同的要求,相当于对W施加了不同的约束
                                                       3.降维效果的评估是通过比较学习器的性能
                                                                             1.最大重构性: 样本点到这个超平面的距离都足够近
                                                  一种最常用的线性降维方法,具有两种性质:
                                                                             2.最大可分性: 样本点在这个超平面上的投影能尽可能分开
                                                          1.输入样本集,低维空间维数d'
                                                          2.对所有样本进行中心化
                                                 工作机制
                                                          3.对协方差矩阵做特征值分解
                                                          4.取最大的d'个特征值所对应的特征向量w1,w2,...wd'
                                   主成分分析(PCA)
                                                          5.输出投影矩阵W
                                                             1.人为选取
                                                             2.交叉验证来选取较好的d'值
                                                 如何选取维数d'
                                                             3.从重构的角度设置一个重构阈值,例如t = 95%, t为差异性大小
                                                           1.增加了采样密度
                                                 降维的影响
                                                           2.在一定程度上具有降噪效果
降维与度量学习
                                       多现实任务中,可能需要非线性映射才能找到恰当的低维嵌入,KPCA是一种常用的非线性降维方法,其基于核技巧对线性降维方法进行"核化"
                                               1.通过核技巧将样本映射至高维特征空间
                                               2.再在特征空间中实施PCA
                                       对一个新样本,为获得投影后的坐标,KPCA需对所有样本求和,计算开销较大
                                  流形学习(manifold learning)是一类借鉴了拓扑流形概念的降维方法。"流形"是在局部与欧氏空间同胚的空间,即它在局部具有欧氏空间的性质,能用欧式距离来进行距离运算
                                                   出发点是测地线距离即两点之间的本真距离
                                                           1.输入样本集D,近邻参数k,低维空间维数d'
                                                           2.对于每个样本x(i)确定其k个近邻,将x(i)与k近邻点之间的距离设置为欧氏距离,与其他点的距离设置为无穷大
                                                   工作机制
                                   等度量映射 (Isomap)
                                                           3.调用最短路径计算任意两样本点之间的距离dist
                                                           4.将dist作为MDS算法的输入,得到其样本集D在低维空间的投影作为输出
                                                         1.对于新样本,将其映射到低维空间的一般方法是将训练样本的高维空间坐标作为输入,低维空间坐标作为输出,训练一个回归学习器来对新样本的低维空间坐标进行预测
                        流形学习
                                                   注意
                                                                               1.指定近邻点个数(缺点是样本可能较稀疏)
                                                         2.对近邻图的构建通常有两种做法
                                                                               2.指定距离阈值,小于阈值的点被认为是近邻点(缺点是可能会造成"短路"或"断路"问题)
                                                  试图保持领域内样本之间的线性关系
                                                          1..输入样本集D,近邻参数k,低维空间维数d'
                                                          2.LLE纤维每个样本x(i)找到其近邻下标几何Q(i)
                                   局部线性嵌入(LLE)
                                                          工作机制
                                                          5.对M进行特征值分解,取最小的d'个特征值对应的特征向量组成的矩阵为变换矩阵
                                                           6.输出样本集D在低维空间的投影Z
                                每个空间对应了在样本属性上定义的一个距离度量,而寻找合适的空间,实质就是在寻找一个合适的距离度量,度量学习的动机就是直接"学习"出一个合适的距离度量
                                                                                                \mathrm{dist}^2_{\mathrm{wed}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = ||\boldsymbol{x}_i - \boldsymbol{x}_j||_2^2 = w_1 \cdot dist_{ij,1}^2 + w_2 \cdot dist_{ij,2}^2 + \ldots + w_d \cdot dist_{ij,d}^2
                                         1.通过假定不同属性的重要性不同,引入属性权重w,以确定一个便于学习的距离度量表达形式 =(x_i-x_j)^{\mathrm{T}}\mathbf{W}(x_i-x_j),
                                工作机制
                                        2.因为现实生活中两个属性之间往往相关,则需将正交的权重矩阵W替换为一个普通的(半)正定对称矩阵M(学习而得)
                      度量学习
                                                            \operatorname{dist}^2_{\mathrm{mah}}(oldsymbol{x}_i,oldsymbol{x}_j) = (oldsymbol{x}_i - oldsymbol{x}_j)^{\mathrm{T}} \mathbf{M} (oldsymbol{x}_i - oldsymbol{x}_j) = \|oldsymbol{x}_i - oldsymbol{x}_j\|_{\mathbf{M}}^2 从而完成映射
                                        3.根据矩阵M,得到马氏距离
                                                            假定我们希望提高近邻分类器的性能,则可将M直接嵌入到近邻分类器的评价指标中去,通过优化该性能指标相应地求得M
                                M的学习
                                        对M的学习需要设置一个目标
                                                            1.我们可以把错误率作为度量学习的优化目标(近邻成分分析(NCA))
                                                            2.还可在度量学习中引入领域知识,如样本相似则可定义"必连"约束集合,不相似即可定义"勿连",以希望相似的样本之间距离较小,不相似的样本之间距离较大来对M进行学习
                          1.直接对矩阵对象(例如一幅图像)进行降维操作会比将其拉伸为向量(例如把图像逐行拼接成一个向量)在进行降维操作有更好的性能(2DPCA)
                    Tips
                          2.常见的流形学习还有拉普拉斯特征映射(LE),局部切空间对齐(LTSA)。流形学习欲有效进行邻域保持则需样本密采样,而这恰是高维情形下的重大障碍,因此其在实践中的性能往往没有预期好
```