

1、传统特征机器学习（比如方向梯度直方图 HOG）跟深度学习有什么不同？深度学习的超越传统特征机器学习的原因是？

(1)、特征构造不同：传统的特征学习依靠与人工构造特征，其中 HOG, LBP 等解决了如何描述数据问题。而深度学习时通过学习自动生成合适的特征，深度学习模拟人类视觉特征，将特征划分为低层特征和高级特征，其中低级特征一般是在网络前几层学习到的（纹理特征），在网络深层一般学习到更高级的特征，例如头部，手臂等。高级特征几乎不会出现在传统的特征提取当中。

(2)、特征映射不同：传统的机器学习把特征映射到目标空间，SVM, adaboost 等都是解决如何把输入特征映射到目标空间然后完成分类，回归任务。而深度学习中最后的分类被全连接层代替。

(3)、处理数据量不同：传统的机器学习方法可能更适合小数据集，深度学习更适合处理大数据问题。

(4)、使用的设备不同：深度学习依赖于更加高端的设备（GPU），传统的机器学习可能仅需 CPU 就可以实现。

(5)、时间不同：深度学习训练时间长，传统特征学习训练时间短。

原因：

(1)、深度学习在模型训练的同时，对特征抽取进行集成。传统的机器学习依赖于特定领域的专家进行人工特征提取；深度学习不需要对任何特征有先验知识，将特征抽取和分类放在一起完成。

(2)、科技发展，有了 GPU, TPU 等更高级的硬件设备。深度学习融合了这些高端设备。

2、 $1 \times 1$  卷积核的作用，池化层的作用 27

$1 \times 1$  卷积核的作用：

$1 \times 1$  卷积可以控制输入特征图的深度，可以减小或增大它，或者在不改变深度时仅添加非线性。

融合不同特征图通道间同一位置的信息并且减小通道数。输出特征图的通道数与卷积核的个数有关，当卷积核的个数小于输入特征图的通道数时，可以起到减小通道数的作用，同时将输入特征图的多通道特征通过单个卷积核进行卷积操作融合。

融合不同特征图通道间同一位置的信息并且增加非线性。当输入特征图和输出的特征图的通道数完全相同时， $1 \times 1$  卷积核可以为输入数据的特征图增加非线性。

池化层的作用： 37

池化层是将初级的视觉特征筛选并结合成更高级、抽象的视觉特征的过程。通过采样实现，经过池化层后，特征图数量不变，尺寸变小。因此池化主要功能是：

- 保留主要特征的同时减少参数和计算量。经过 MaxPool 可以减小卷积核的尺寸，同时又可以保留相应特征，所以主要用来降维
- 在一定程度上能防止过拟合，由于这一层没有参数，不需要学习。
- 特征不变性，这种不变性包括 translation(平移), rotation(旋转), scale(尺度). 这就使网络的鲁棒性增强了，有一定抗扰动的作用

### 3、SVM 转化为对偶问题的优点

- 对偶问题将原始问题中的不等式约束转为了对偶问题中的等式约束。
- 改变了问题的复杂度，由求特征向量  $w$  转化为求比例系数  $\alpha$ ，在原始问题下，求解的复杂度与样本的维度有关，即  $w$  的维度。在对偶问题下，只与样本数量有关（对应为  $m$ ）。
  - ◆ SVM 原始问题模型严重依赖于数据集的维度  $d$ ，如果维度  $d$  太高就会严重提升运算时间。
  - ◆ 对偶问题事实上把 SVM 从依赖  $d$  个维度转变到依赖  $m$  个数据点，考虑到在最后计算时只有支持向量才有意义，所以这个计算量实际上比  $m$  小很多。
- 求解更高效，因为只用求解  $\alpha$  系数，而  $\alpha$  系数只有在支持向量才非 0，其它全部为 0。
- 方便核函数的引入，进而推广到非线性分类问题。

### 4、核函数的作用

当样本在原始空间线性不可分时，可将样本从原始空间映射到一个更高维的特征空间，使得样本在这个特征空间内线性可分。在求解对偶问题时仅需计算特征向量的内积。

(1)、引入了核函数，把高维向量的内积转变成了求低维向量的内积问题。即在特征空间的内积等于它们在原始样本空间中通过核函数  $K$  计算的结果。

(2)、核函数是一种表征映射、实现内积逻辑关系且降低计算复杂度的一类特殊函数，定义为  $K(x,y)=\langle \phi(x),\phi(y) \rangle$ ，一方面数据变成了高维空间中线性可分的数据；另一方面不需要求解具体的映射函数，只需要给定具体的核函数即可。

### 5、特征的相似度定义，性质

特征相似度是将样本之间的相似性进行数值化表示。

性质：

非负性： $d(x,y)$  的取值范围为  $[0,1]$  之间，仅当  $d(x,x)=1$

对称性： $d(x,y)=d(y,x)$

### 6、预剪枝与后剪枝优缺点

预剪枝优点和缺点

- 优点 1：预剪枝让决策树的很多分支没有展开，降低了过拟合风险
- 优点 2：显著减少训练时间和测试时间开销
- 缺点 1：欠拟合风险。有些分支的当前划分虽然不能提升泛化性能，但在其基础上进行的后续划分却有可能导致性能显著提高。预剪枝基于“贪心”本质禁止这些分支展开，带来了欠拟合风险

后剪枝的优缺点

- 优点 1：后剪枝比预剪枝保留了更多的分支，欠拟合风险小，泛化性能往往优于预剪枝决策树
- 缺点 1：训练时间开销大：后剪枝过程是在生成完全决策树之后进行的，需要自底向上对所有非叶结点逐一考察

### 7、密度直接可达，密度可达，密度相连定义

- **直接密度可达**: 对于样本集合D, 如果样本点q 在p 的Eps邻域内, 并且p 为核心点, 那么点q 从点p 直接密度可达(密度直达)。
- **密度可达**: 对于样本集合D, 给定一串样本点 $p_1, p_2, \dots, p_n$ ,  $p=p_1, q=p_n$ , 假定对象 $p_i$  从  $p_{i-1}$  直接密度可达, 那么点q 从点p 密度可达。

**密度相连**: 对于样本集合D 中的任意一点O, 如果存在点p 到点o 密度可达, 并且点q 到点o 密度可达, 那么点q 到点 p 密度相连。

## 8、DBSCAN 相对于 k-means 的优缺点

6.5..

DBSCAN 的主要优点:

- 可以对任意形状的稠密数据集进行聚类, 相对的, K-Means 之类的聚类算法一般只适用于凸数据集。
- 可以在聚类时发现异常点, 对数据集中的异常点不敏感。
- 聚类结果没有偏倚, 相对的, K-Means 之类的聚类算法初始值对聚类结果有很大影响。

DBSCAN 的主要缺点:

- 如果样本集的密度不均匀、聚类间距差相差很大时, 聚类质量较差, 这时用 DBSCAN 聚类一般不适合。
- 如果样本集较大时, 聚类收敛时间较长。
- 调参相对于传统的 K-Means 之类的聚类算法稍复杂, 主要需要对距离阈值  $\epsilon$ , 邻域样本数阈值 MinPts 联合调参, 不同的参数组合对最后的聚类效果有较大影响。

## 9、LDA 和 PCA 区别

LDA (线性判别式分析 Linear Discriminant Analysis) 属于机器学习中的监督学习算法, 常用来做特征提取、数据降维和任务分类。LDA 算法的目标是使降维后的数据类内方差最小, 类间方差最大 (即使数据在低维度上进行投影, 投影后希望每一类别数据的投影点尽可能的接近, 而不同类别的数据中心之间的距离尽可能的大。)

相同点:

- 1) 两者均可以对数据进行降维。
- 2) 两者在降维时均使用了矩阵特征分解的思想。
- 3) 两者都假设数据符合高斯分布。

不同点:

- 1) LDA 是有监督的降维方法, 而 PCA 是无监督的降维方法
- 2) LDA 降维最多降到类别数  $k-1$  的维数, 而 PCA 没有这个限制。
- 3) LDA 除了可以用于降维, 还可以用于分类。
- 4) LDA 选择分类性能最好的投影方向, 而 PCA 选择样本点投影具有最大方差的方向。