2 线性判别器

Fisher: ，点到

解得，判别：，表示分界点

感知器：

优化：，如果线性可分，一定能有限步收敛，

线性回归：，，显著性检验

MSE：，最小二乘求解析解

Logistic回归：

多分类：c个线性函数，，把i类y错分到j

3 学习机器性能评估

测试是真实可靠估计：测试和训练独立、测试和未来iid、测试数目足够大

自举法/B.632：训练集有放回采样，训练正确率\*0.632，测试\*0.368

灵敏度，特异度，发现率，准确率

ROC曲线：横坐标FP（1-特异度），纵坐标TP（灵敏度），AUC=0.5无效

4 人工神经网络 三要素：网络结构、激活函数、学习算法

BP可能问题：收敛慢、曲线震荡、过拟合、泛化能力差解决：换激活函数、BN、目标调整、数据增强、调网络结构、初始化、学习率、momentum、weight decay、提前终止

MLP：一个隐层（3层）任意凸形成区域划分；两个隐层（4层）任意形状

5 支持向量机和统计学习理论

最优超平面，结果：

对偶：

线性不可分：

对偶：

Mercer：，当

SVM：，SV网络：x先和每个支持向量过核函数，再通过线性网络，权值为；复杂度：取决于样本数目

统计学习理论：，分类：，拟合：，概率密度估计：

实际求解：

生长函数结构：要么，要么

SRM原则：选函数集的一系列子集，复杂度递增，选一个适当的子集，让风险上界最小

支持向量回归：，对偶：

6 其它非线性方法

KNN无限样本的错误率：，P为无穷样本的错误率

KNN的问题：计算存储量、票数接近时风险大、样本不能无穷多

**分支定界算法** 先分好类，算好中心点和半径,新样本x，节点则x的近邻不可能在中；不是x的最近邻；最后逐个排除

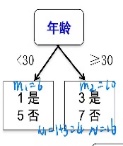
**剪辑近邻法** 用参考集对考试集分类，去掉考试集分错的，再用考试集剩下的分类

**压缩近邻法**开始一个样本，每次一个样本用分类，分错就放进，最后用分类

**Prototype（原型）近邻法** 先对每类KMeans，中心C个点为代表，用中心C点一近邻

**可拒绝近邻法** K近邻票数才同意，否则拒绝

3 距离计算,闵可夫斯基距离 余弦距离：

ID3：信息增益, ，

基尼不纯度：

误差不纯度：

**C4.**5 ，连续取值有可以多个分枝

**CART算法** 对连续取值二分，建回归树，选x的阈值和yhat让最小

**避免过学习** 预剪枝；后剪枝：剪枝后比剪枝前验证集显著提升，可以剪枝

**bagging** 并行；随机森林 样本同上，每次对特征进行采样，最后决策树投票

**boosting** 训练新学习器拟合残；AdaBoost 样本有同样权重，训练分类器计算错误率按分错的样本权重求和，分类器权重，下一次样本权重，归一化，输出**Gradient Boosting**初始化，每次算，用训练一个新学习器，计算更新权重，更新模型。

**XGBoost** 加正则化项，L在处泰勒展开，变成一个关于的二次函数

**非线性回归**：引入 dummy variable、交叉项

**最近邻回归** k个近邻取值加权求和，权重：1/k，1/d，

**投票**：简单多数；加权投票（与准确率或方差有关）；排序输出（Borda）；概率输出

7贝叶斯决策理论，后验概率，似然比

**最小错误率贝叶斯决策** 条件：类别数一定；和已知

决策方法：，或用贝叶斯公式，。错误率计算： **最小风险贝叶斯决策**

**正态分布下的贝叶斯决策：**

8 概率密度函数的估计

**概念** 参数估计：已知概率密度函数的形式，估计参数

**最大似然估计** 似然函数：

**贝叶斯估计** 期望风险：，最小化R即最小化**，**如果损失函数为平方损失时，

计算步骤：确定的先验分布；求样本的联合分布；求的后验概率分布；求的贝叶斯估计量

正态贝叶斯：，先验，已知

**贝叶斯学习：**在线学习**，**方法：

**直方图方法**（下面开始非参数估计）

，N是样本总数，V是某个仓的体积，k是落入这个仓中的样本数

**kN近邻估计** 找k近邻的大小，

**Parzen窗法** 窗宽应该随着样本数目的增大减小,窗宽越大越平滑

9 贝叶斯网络和隐马尔可夫模型

**马尔科夫链**：

**HMM建模** 整体概率

**前向算法** ，t时刻隐状态为j的概率，复杂度

**后向算法**，状态为j时后面观测值出现的概率

**Viterbi算法** ,终止结果，复杂度

**HMM训练问题；贝叶斯网络**：有向无环图

估计概率：E：

M：;估计E：E：

M：

**朴素贝叶斯分类器** 假设：各个分量条件于独立。；DAG条件独立：

拉普拉斯平滑:，C为类别,S为x取值数

**DAG上的贝叶斯推理** Z子节点未知时XY独立，剩下Z已知时XY独立

**贝叶斯网络的学习，参数学习**：已知模型结构，但是不知道模型的参数值

最大化（有缺失EM和无缺失参数）

10 特征选择

学习算法的方差和偏差分解

是真实值，y是观测值，是数据有限时的估计函数，当数据趋于无穷时，趋于f，

分析高维数据的方法：SVM；特征选择；正则化方法

**特征选择的好处**：有利于可视化和理解；降低数据采样和数据存储的开销；降低模型使用时的计算开销；避免维数灾难

**最优子集**：遍历所有特征组合，找最好的；**过滤法**：计算输入和输出之间的相关性，选相关性最好的特征；不能在测试集上选特征

Fisher：；相关系数：；互信息：；基于熵的判据：计算依照某一维特征分类的熵；KL散度：；统计检验：t检验

**包裹法**：需要有验证集，不能在测试集上选特征

前向选择：复杂度，p为模型个数，维度小于p时也适用

后向选择维度必须大于等于p； 混合算法：先加特征，再减，再加

**特征选择的遗传算法**:随机初始化染色体：p选k个，有种；计算每条染色体的适应度；按照某种概率模型进行采样，进行繁殖，生成下一代种群；迭代，直到适应度到阈值

**模型选择**AIC=2d-2ln(L)；BIC=ln(n)d-2ln(L)，d：模型中拟合参数的数量；L：似然函数的极大值；n：样本数

**随机置换法**：随机打乱样本的标记，通过同样的学习过程，统计出没有分类信息情况下识别正确率分布，得到真实值的p值，如果，则显著

**嵌入法**：正则化方法,L2比L1更稀疏

11 特征提取和数据可视化

**PCA**第二大推导：对a2求导后左右同乘**,** ，方差为特征值

奇异值分解计算PCA：对X作UV分解：，选V的前k行作为特征向量组成矩阵；KPCA：把改成

**KL展开**：用无穷个的正交归一化向量系作基，展开x，用有限项来逼近x

**KL变换**：最小化MSE预测误差，变换矩阵也是的前d个特征向量；当求取特征值的矩阵为协方差矩阵时，KL变换等价于PCA

**监督KL变换**：计算类内离散度矩阵,P为先验概率；对作KL变换，新特征，方差为，计算，其中为类间离散度矩阵，对J(yi)排序，选前d个分量，uj组成变换阵

**流形学习**：认为高维冗余数据是d维数据在D维空间中扭曲的结果。

LLE：认为局部关系在低维空间中也满足，先在高维空间中找X的一个近邻X’，并学习利用X’重建X的权重W；在低维空间中学习固定W，学习重建误差最小的Y；

Isomap：近用欧式，远用相邻欧氏距离相加

**高维数据的低维显示**PCA MDS：把样本之间的距离关系在低维中表示出来

t-SNE：利用概率分布度量样本之间的距离，把高维空间中的欧氏距离转化为条件概率密度函数，，；低维概率建模：，优化目标最小，梯度下降求解

困惑度：数据密集的地方上述方差应该小，分布的信息熵随方差的增大而增大，故可以通过困惑度反推方差是否应该增大或减小。

12 聚类分析

**动态聚类** 选距离作为相似性度量；定义准则函数；初始化分类方法和迭代算法

K-Means，，类的初始划分：经验；随机分k类，选各类中心；密度法：选密度最大第一个中心点，该点一定距离外密度最大的第二个；先k-1聚类，最后一个中心点找离剩下k-1个最远的

缺点：要求类别数已知；最小方差划分，不一定符合内部规律；对类别的硬划分；和初始点有关；和距离度量方式有关

基于核的相似性度量：定义x到之间的距离，准则函数，相当于中心点，相当于距离

**模糊K均值方法** 损失函数；求解：分别对m和求偏导，迭代；去模糊化：

**GMM** 假设数据为多个高斯分布的和

E：求后验期望：

M：对数似然最大化：

解得：样本类别划分：

**一致聚类**整体一致性矩阵，新的距离度量。衡量聚类一致性：，用CDF曲线下的面积衡量一致性，画出，找到肘点。

**SOM**权值学习：

**聚类结果评估**基于互信息的分数：，U为真实标签，V为聚类结果，，值越大越好

轮廓系数：，为样本i和同类的平均距离，b(i)为样本和距离它最近的类平均距离，在-1到1，取值越大越好

13 深度学习

卷积层参数个数计算：N\*(k\*k\*d+1)，N是卷积核个数

输出图的大小计算：，n是图像尺寸，k是卷积核尺寸

**反馈神经网络**LSTM：cell state：前向通道；遗忘门：决定前向通道记得多少；输入门：处理输入数据：输出门：综合前面几个输出

**常用训练技巧** 训练顺序：整体梯度下降；随机；批量**；**优化器：SGD；momentum：归一化：BN，归一化+线性变换，加快网络收敛，但是不要抹掉之前的学习结果Dropout

14 表示学习：提取表示特征供后续适用、利用无标签样本学习、便于降维可视化、异质信号融合、解决语义相似度问题，在嵌入空间计算相似度

**AE训练**：把每两层的节点作为一个RBM网络，用CD算法，编码器权值对称的到解码器，最后再整体进行finetune

**自然语言处理** 不用onehot表示：没有包含语义信息；词向量维度过高；信息过于稀疏；GPT-3：网络结构更好；基于上下文的表示方法（不同上下文语义不同）；模型参数增加

**网络图的表示学习** 节点表示：Deepwalk；图表示：两个结构相似的图应该有相似的表示社团检测：利用节点表示的距离矩阵进行聚类

15 深度生成模型

RBM：只有一个隐层，隐层单元之间没有连接；给定可见单元，隐层单元条件独立；CD算法：用h和x互相生成无穷多次，就可以认为是平衡时的采样

DBN：类似MLP结构的RBM，sigmoid激活；把最后一个隐层变成玻尔兹曼机，从而解除依赖关系；训练：CD，从x层开始

**VAE** 目标函数：，用拟合，；再参数化技巧：直接输出分布不能BP，所以在正态分布中采样，平移，再BP；有显式表示和可解释性，但相对对抗难度更大,容易分布拟合误差高,效果差

**GAN** 目标函数：

优：不用对隐变量做推断；理论上任何可微函数都可D/G；G参数更新来自D不是样本

缺：可解释性不足；难以同步G和D的训练，模型收敛不稳定，容易collapse

16 迁移学习、半监督学习、小样本学习

**基于样本的迁移学习** 找其它数据集类似的样本；**基于网络的迁移学习** 预训练

**半监督学习** 假设：样本距离相近，类标相同；分界面不会穿过高样本密度区域；同一个低维流形中的样本有相同的标签

自训练算法：用监督标非监督，反复；基于图的方法：画出一张图，K近邻有权重

半监督SVM：分界面要让没有类标的样本产生的分类面间隔最大化

**小样本学习** 估计误差：由于样本量不足造成的参数误差；近似误差：原空间本身的问题导致误差，无法通过数据改变

基于模型的方法：基于先验知识定义模型，减少模型的搜索范围，比如对数据降维再训练；基于数据的方法：基于先验扩充数据集：数据增强；基于算法的方法：找到一条更好的搜索路径：在大数据集上预训练，在小数据集上做微调