

2019 级导航工程《机器学习》课后习题（全）

第1章 《机器学习概论》习题

一、名词解释

特征空间 *

特征向量 *

假设空间 *

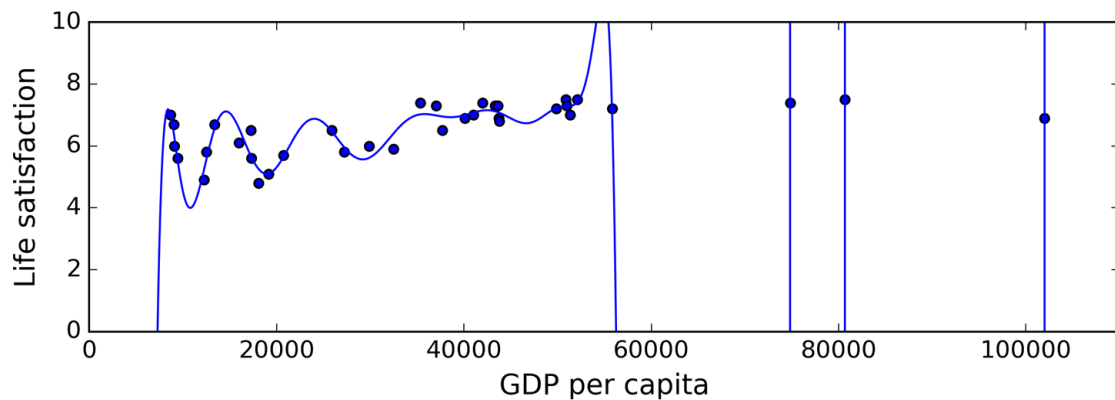
奥卡姆剃刀

* 二、如下图是通过 GDP 来对幸福度进行回归，蓝色点为已知样本，蓝色线表示回归曲线，请回答

(1) 该模型是“欠拟合”还是“过拟合”？

(2) 分别解释“欠拟合”、“过拟合”的定义，以及模型在欠拟合、或过拟合下对未知样本进行预测，会导致什么结果？

(3) 简单说明如何解决过拟合和欠拟合问题。



三、请说明监督学习和非监督学习的不同，并判断“聚类分析和降维算法”属于监督学习还是非监督学习，说明为什么。

四、论述回归和分类的差异

注：带*的为本年度作业题，需提交纸质作业计算期末成绩

第2章 《模型评估与选择》习题

一、名词解释

VC 维、经验误差*、泛化误差*、训练集*、测试集*、验证集*、查准率*、查全率*、偏差和方差*、平衡点、假正例率*、真正例率*、分层采样*、交叉验证法*、自助采样*

二、如数据集包含 1000 个样本，其中 500 个正例、500 个反例，将其划分为包含 70% 样本的训练集和 30% 样本的测试集，请给出“留出法”评估中训练集和测试集所包含的正例和反例个数，并简要说明留出法评估需要遵循哪些原则

三、请根据下表的学习器对 10 个样本的预测结果，给出前五个为正例时的查准率和查全率，绘制出 10 个样本预测结果的 P-R 曲线，并分析平衡点与学习器性能的关系。

编号	Label	预测指标
1	正	0.98
2	正	0.95
3	正	0.91
4	负	0.87
5	正	0.85
6	负	0.70
7	负	0.55
8	负	0.22
9	负	0.13
10	负	0.07

*四、现在用已经训练好的学习器对 8 个测试样本（4 个正例，4 个反例）进行预测，假设预测结果为：

(s1, 0.77, +)、(s2, 0.65, -)、(s3, 0.56, +)、(s4, 0.56, -),

(s5, 0.50, +)、(s6, 0.30, +)、(s7, 0.20, -)、(s8, 0.15, -)

请画出 8 个测试样本的 ROC 曲线，并给出 AUC 的值。

五、请简要说明泛化误差可拆解为哪些因素？解释什么叫偏差-方差窘境？

注：带*的为本年度作业题，需提交纸质作业计算期末成绩

第3章 《贝叶斯分类器》习题

一、名称解释

后验概率*

损失函数*

朴素贝叶斯分类器*

贝叶斯估计

最大似然估计 (MLE)

二、贝叶斯估计和最大似然估计是估计类条件概率的两种策略，请简述它们的差异？

三、基于最小风险的 Bayes 决策里面条件风险和期望风险各是什么含义？简单阐述最小风险的 Bayes 决策和最小错误 Bayes 决策的差异？什么情况下两者是等价的？

四、请简要说说 EM 算法的思想。在贝叶斯分类中什么情况需要使用 EM 方法？

五、若已知先验概率 $p(\omega_1), p(\omega_2)$ ，类条件概率密度 $p(x|\omega_1), p(x|\omega_2)$ ，写出至少一种贝叶斯判别函数及其决策规则。

*六、假定对某一类人群进行疾病筛查，正常人群为 ω_1 ，患者为 ω_2 ，设正常和患者的先验概率为：

$P(\omega_1)=0.9, P(\omega_2)=0.1$ ；现有一位被检查者，其观测值为 x ，从类条件概率密度曲线上查得 $P(x|\omega_1)=0.2, P(x|\omega_2)=0.4$ ，同时已知风险损失函数为

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 0 & 6 \\ 1 & 0 \end{bmatrix},$$

试对该被检查者用以下两种方式进行分类：

用最小错误率的贝叶斯决策，并写出判别函数和决策方程；

用最小风险的贝叶斯决策，并写出判别函数和决策方程。

注：带*的为本年度作业题，需提交纸质作业计算期末成绩

第4章 《线性模型》习题

一、名词解释

线性回归*

广义线性模型*

逻辑斯谛回归 (Logistic Regression) *

线性判别分析 (LDA) *

支持向量 (SV) *

加权空间

梯度下降法*

非线性分割

二、对于线性回归，假设函数表示为 $h_w(x_0, x_1, \dots, x_n) = w_0x_0 + w_1x_1 + \dots + w_nx_n + w_{n+1}$ ，其中 $w_i (i = 0, 1, \dots, n+1)$ 为模型参数。 $x_i (i = 0, 1, \dots, n)$ 为每个样本的 $n+1$ 个特征值，共有 m 个样本。则对于该假设函数，损失函数可表示为：

$$J(w_0, w_1, \dots, w_n, w_{n+1}) = \frac{1}{2m} \sum_{j=1}^m (h_w(x_0^j, x_1^j, \dots, x_n^j) - y^j)^2$$

请给出使用梯度下降法优化该损失函数的算法过程。

*三、假设三个 $\omega_1, \omega_2, \omega_3$ 判别函数为：

$$\begin{cases} g_1(x) = -x_1 - x_2 + 5 \\ g_2(x) = -x_1 + 3 \\ g_3(x) = -x_1 + x_2 \end{cases}$$

请画图给出一对多情况下的判别边界，以及三个类别的判别区域，并判断样本(1,3)及(4,5)的类别，如果不能确定其类别，请说明原因

四、计算题：假设现有二分类数据集 X ，根据类别可将其划分为两个子集 $X_0, X_1 \in X$ ，请分别写出 LDA (Fisher 线性判别分析) 求投影方向 w 的计算步骤和公式。

五、简述逻辑斯谛回归 (Logistic Regression) 中损失函数设计的思想，比较其与线性回归基于均方误差最小化设计的损失函数有什么不同

注：带*的为本年度作业题，需提交纸质作业计算期末成绩

第5章 《决策树》习题

一、名词解释

熵和条件熵*

信息增益*

信息增益率

基尼指数*

决策树剪枝

叶结点

二、简要说明一个决策树的生成过程？在决策树中为什么要进行特征选择？

三、请简述 ID3、C4.5、CART（分类）算法在最优属性划分方面的主要差异。

四、简要说明分类树和回归树的区别和联系。

*五、计算题

对下表分别用 ID3 和 CART 算法生成决策树（要求写出详细的计算步骤）。

ID	年龄	有工作	有自己的房子	信贷情况	类别
1	青年	否	否	一般	否
2	青年	否	否	好	否
3	青年	是	否	好	是
4	青年	是	是	好	是
5	青年	是	是	非常好	是
6	青年	是	是	一般	是
7	青年	否	否	一般	否
8	中年	否	否	一般	否
9	中年	否	否	好	否
10	中年	是	是	好	是
11	中年	否	是	非常好	是
12	中年	否	是	非常好	是
13	中年	是	是	好	是
14	老年	否	是	非常好	是
15	老年	否	是	好	是
16	老年	是	否	好	是
17	老年	是	否	非常好	是
18	老年	否	否	一般	否

注：带*的为本年度作业题，需提交纸质作业计算期末成绩

第6章 《集成学习和随机森林》习题

一、名词解释

同质集成

异质集成

基学习器*

平均法 (averaging) *

投票法 (Voting) *

堆叠法 (Stacking)

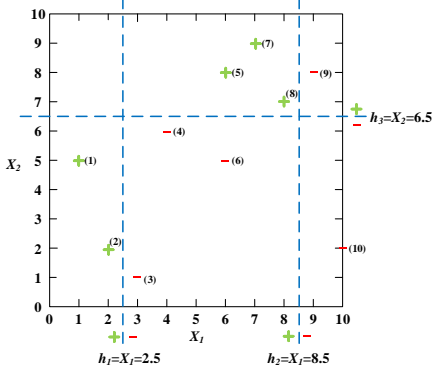
二、集成学习把多个学习器结合起来，如何能获得比最好的单一学习器更好的性能？请从准确性和多样性的角度进行分析。

三、在训练阶段，堆叠法中次级训练集是利用初级学习器产生的，若直接用初级学习器的训练集来产生次级训练集，存在过拟合风险，请简要阐述可采用什么策略避免堆叠法中的过拟合风险？

四*、给定如图表所示的训练样本，弱学习器采用平行于坐标轴的直线 h_1 、 h_2 和 h_3 ，请用 Adaboost 算法的实现强分类过程。

$$h_1 = \begin{cases} 1 & , X_1 < 2.5 \\ -1 & , X_1 > 2.5 \end{cases}, h_2 = \begin{cases} 1 & , X_1 < 8.5 \\ -1 & , X_1 > 8.5 \end{cases}, h_3 = \begin{cases} 1 & , X_2 > 6.5 \\ -1 & , X_2 < 6.5 \end{cases}$$

样本序号	1	2	3	4	5	6	7	8	9	10
样本点 X	(1,5)	(2,2)	(3,1)	(4,6)	(6,8)	(6,5)	(7,9)	(8,7)	(9,8)	(10,2)
类别 Y	1	1	-1	-1	1	-1	1	1	-1	-1



六、请从偏差-方差的角度分析 Boosting 和 Bagging 方法对泛化性能的影响？

注：带*的为本年度作业题，需提交纸质作业计算期末成绩

第 7 章 《维度归约》 习题

一、名词解释

维数灾难*

欧氏距离

马氏距离

测地线(geodesic)距离

向量的内积*

特征值*

特征向量*

流形学习

二、简要说明“特征选择”与“特征提取”的区别和联系

三、线性降维和非线性降维的“线性”和“非线性”分别体现在哪里？

四、降维分析中涉及的投影矩阵通常要求是正交的。请问（1）正交投影矩阵相对于非正交投影矩阵用于降维的优点是什么？（2）为什么在 PCA 中投影矩阵一定是正交的？

五*、参照教材图 6-11，绘制一个数据集示例，让 PCA 和 LDA 找到相同的好方向，另绘制一个数据集，让 PCA 和 LDA 都找不到一个好的投影方向

六*、计算题：用 PCA 方法计算这 5 个二维样本数据 $\begin{pmatrix} -1 & -1 & 0 & 0 & 2 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}$ 在一维主分量上的投影值

注：带*的为本年度作业题，需提交纸质作业计算期末成绩

第 8 章 《聚类》 习题

一、名词解释

Jaccard 系数* 内部指标 外部指标 DB 指数

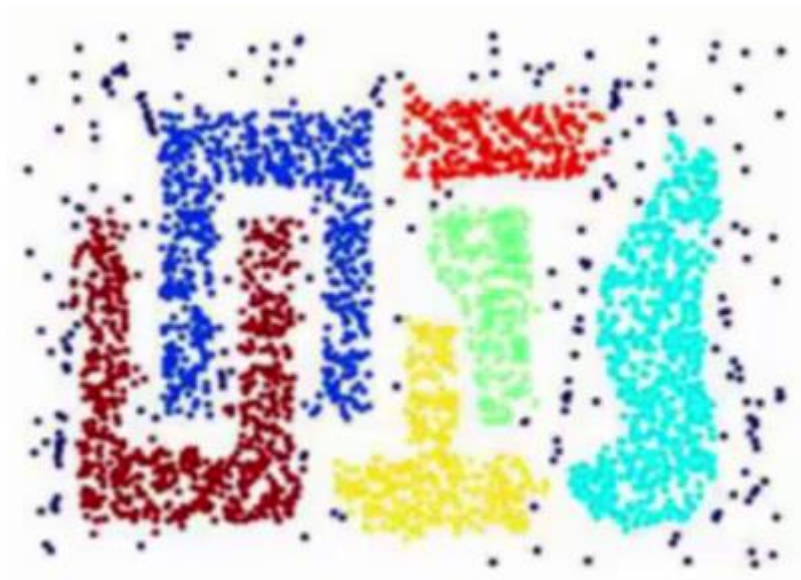
核心对象* 密度直达 密度可达* 密度相连

二、k-means 算法与 kNN 方法的 k 分别指什么？这两个方法的主要区别是什么

三、距离是一种常用的数据相似度的度量函数，请写出四种距离计算的度量函数。

四、在聚类算法中，什么是性能度量？如何衡量聚类性能的优劣？

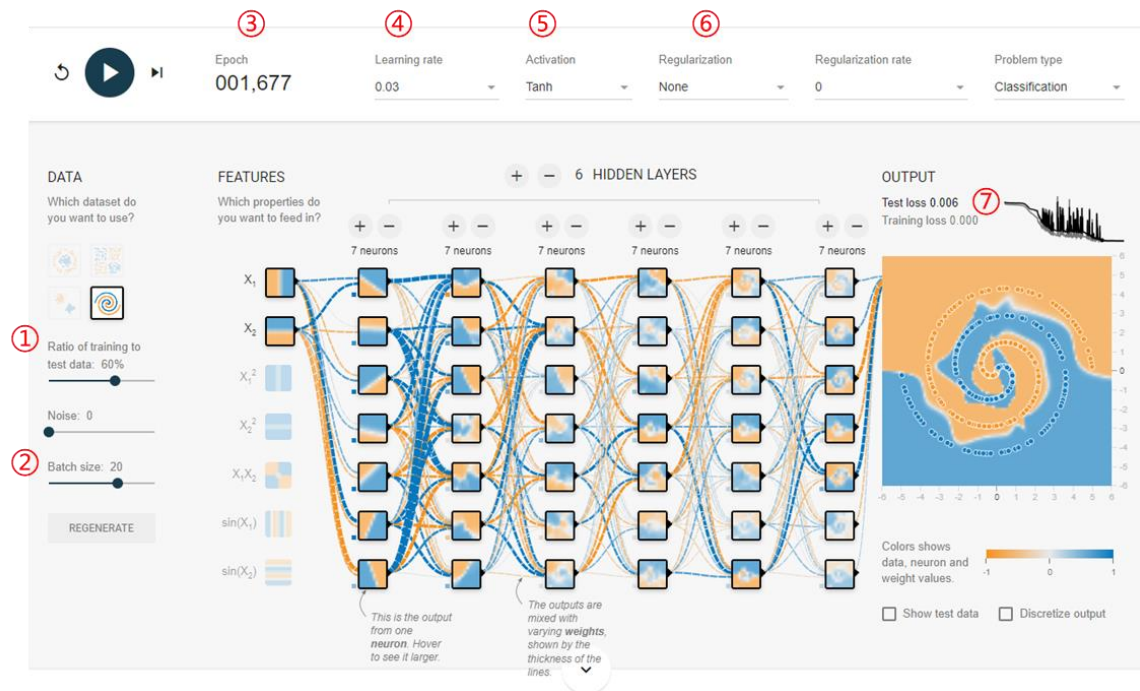
五*、对于下图的数据分类，宜采用什么类聚类算法？其基本原理是什么？该算法中有哪两个重要参数？这两个参数的取值对聚类的结果有什么影响？



六、层次聚类法中如何计算两个类簇的相似度？

第 9 章、第 10 章 《神经网络》、《深度学习》习题

一*、请根据 Tensorflow 提供的一个神经网络学习网页，解释下列名词的含义



- ① Ratio of training to test data （训练/测试集比例）
- ② Batch size （批数据大小）
- ③ Epoch （步数）
- ④ Learning rate （学习率）
- ⑤ Activation （激活函数）
- ⑥ Regularization （正则化）
- ⑦ Test loss 和 Training loss （测试损失和训练损失）

二、试回答感知器里的激活函数为什么要选用非线性函数？

三、试说明 PCA 降维方法与感知器的关联

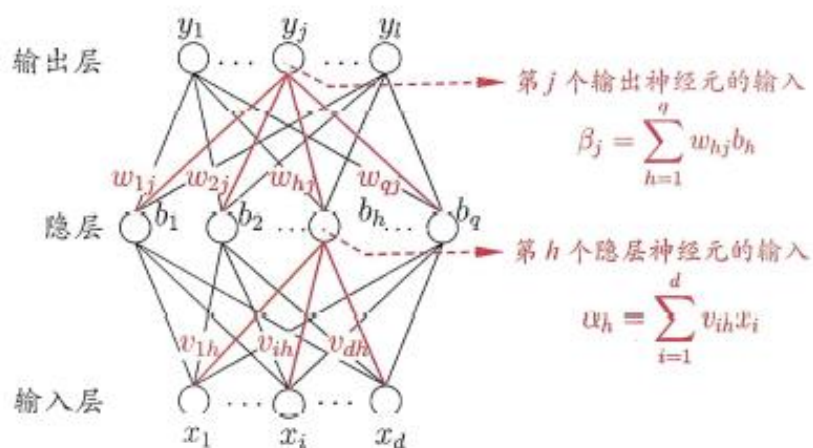
四、试说明 logistic 回归与感知器的关联

五*、根据下图 BP 神经网络，选择 sigmoid 函数作为激活函数，可知 v_{ih} 的梯度下降更新公式如下，试推导一下（sigmoid 函数具有以下性质： $f'(x)=f(x)(1-f(x))$ ）

$$\Delta v_{ih} = \eta e_h x_i$$

$$\text{其中, } e_h = b_h(1 - b_h) \sum_{j=1}^l w_{hj} g_j \quad g_j = \hat{y}_j^k(1 - \hat{y}_j^k)(y_j^k - \hat{y}_j^k)$$

η 是学习率， \hat{y}_j^k 是第 k 个样本在第 j 个输出神经元上的输出值



六*、一个卷积神经网络的输入为 $32 \times 32 \times 3$ 的图像，第一个隐藏层卷积核大小为 5×5 ，共 10 个卷积核。请问这一层总共含有多少参数（含偏置参数）？

七、CNN 中下采样（池化）的目的是什么，常见的池化策略有哪些？

注：带*的为本年度作业题，需提交纸质作业计算期末成绩