

中国科学技术大学信息学院自动化系

2021~2022 学年第 2 学期考试试卷

□A 卷

课程名称: 人工智能导论 课程代码: 210707

开课院系: 信息学院自动化系 考试形式: 闭卷

姓 名: _____ 学 号: _____ 专 业: _____

题 号									总 分
得 分									

一、选择题（共 10 分，每题 2 分）

- 在其他条件不变的前提下，以下哪些做法容易引起机器学习中的过拟合问题。（D）
A. 增加训练集量 B. 减少神经网络隐藏层节点数 C. 删除稀疏的特征
D. SVM 算法中使用高斯核/RBF 核代替线性核
- 池化层在卷积神经网络中扮演了重要的角色，下列关于池化层的论述正确的有(ACD)。
A. 池化操作具有平移不变性 C. 池化操作可以实现数据的降维
B. 池化操作是一种线性变换 D. 池化操作可以扩大感受野
- 在 epsilon-greedy 算法中，epsilon 的值越大，采取随机动作的概率越（B），采用当前 Q 函数最大动作的概率越（）
A.小，小 B.大，小 C.大，大 D.小，大
- 下面哪项操作能实现跟神经网络中 Dropout 的类似效果？（B）
A. Boosting B. Bagging C. Stacking D. Mapping
- 在一个神经网络中，下面哪些方法可以用来处理过拟合？（ABCD）
A. Dropout B. 分批归一化(Batch Normalization) C. 正则化(regularization)
D. 数据增强

二、填空（10 分，每题 2 分）

- 输入图片大小为 200×200，依次经过一层卷积（kernel size 5×5，padding 1，stride 2）之后，输出特征图大小为 99x99。
- 机器学习三要素为 模型、学习准则、优化。
- 留一法是指 交叉验证法的特例，每次留出一个样本当测试集，其他样本训练，循环 N 次（N 为样本数），取平均值作为最终的评估结果。
- Mgu 的全称是 最一般合一，其本质是 在其置换下，所有公式集中谓词公式变为等价了。
- 经验风险是指 模型关于训练样本集的平均损失，

结构风险是指 为了防止过拟合，在经验风险上加上表示模型复杂度的正则化项后的综合风险。

三、判断（在下面的表格中填入√ 或 × ； 共 10 分）

1	2	3	4	5	6	7	8	9	10
F	T	T	T	T	F	T	F	F	T

- 1、面对大规模数据集时，模型训练更容易发生过拟合现象；
- 2、在 A* 算法结束之前，OPEN 表中任何满足 $f(n) < f^*(s)$ 的节点 n ，一定被扩展。
- 3、训练 CNN 时，可以对输入数据进行旋转、平移、缩放等预处理操作来提高模型的泛化能力。
- 4、逻辑回归算法是一种广义的线性回归分析方法，它仅在线性回归算法的基础上，利用 Sigmoid 函数对事件发生的概率进行预测。
- 5、对于大规模数据集而言，数据增强是一种有效提升数据质量的手段。
- 6、信息熵是度量样本集合纯度最常用的一种指标。信息熵的值越大，说明样本集合的纯度越高。
- 7、训练完 SVM 模型后，不是支持向量的那些样本我们可以丢掉，也可以继续分类。
- 8、梯度下降法一定是会收敛在局部最优解。
- 9、多轮博弈后，问题的纳什均衡解会达到方案最优解。
- 10、在各类机器学习算法中，过拟合和欠拟合都是不可能彻底避免的。

四、回答下列问题（共 30 分）

- 1、（10 分）请简述近似线性可分问题下支持向量机算法的解决方案及分类效果。（略）
引入松弛因子 ξ 和惩罚参数 C ，形成软间隔，根据 ξ 和 C 分情况讨论样本分类效果

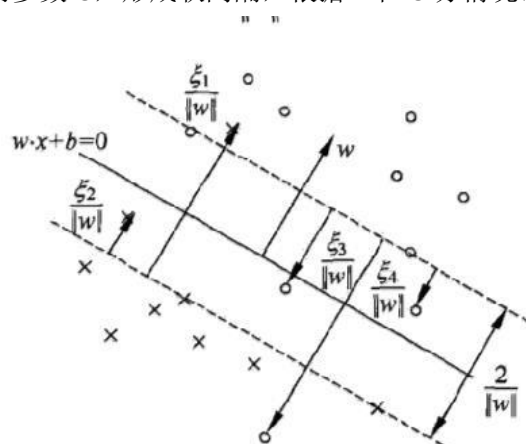


图 7.5 软间隔的支持向量

软间隔的支持向量 x_i 或者在间隔边界上，或者在间隔边界与分离超平面之间，或者在分离超平面误分一侧。若 $\alpha_i^* < C$ ，则 $\xi_i = 0$ ，支持向量 x_i 恰好落在间隔边界上；若 $\alpha_i^* = C$ ， $0 < \xi_i < 1$ ，则分类正确， x_i 在间隔边界与分离超平面之间；若 $\alpha_i^* = C$ ， $\xi_i = 1$ ，则 x_i 在分离超平面上；若 $\alpha_i^* = C$ ， $\xi_i > 1$ ，则 x_i 位于分离超平面误分一侧。

- 2、（10 分）简述 bootstrap 原理、数据处理过程及优缺点。

每次随机从 D 中挑选一个样本，做有放回的抽样放入 D' ，重复 m 次，我们就得到了一个包含 m 个样本的数据集 D' ，显然， D 中有一部分的样本会出现多次，另一部分的样本从不出现。

样本在 m 次采样中始终不被采到的概率是 $(1-1/m)^m$ ，取极限得到约等于 0.368。

优点：在数据集较小，难以有效划分训练测试集时很有用

缺点：改变初始数据的分布，引入估计偏差

3、(10 分) 试说明强化学习中三种策略评估方法的基本思想及各自的优缺点。(略)

动态规划 需要知道模型

蒙特卡罗 需要扩展到终点

时序差分 综合考虑了随机性和及时性

五、用消解原理判定，下列子句集是否可满足，如果不可满足，写出其归结过程：

(1) $\{\neg P(x), P(a) \vee P(f(x))\}$

(2) $\{\neg P(x) \vee P(f(x)), P(a)\}$

(3) $\{\neg P(x), P(f(x))\}$

(4) $\{\neg P(x) \vee \neg P(f(x)), P(f(x))\}$ [10 分]

1) $\{\neg P(x), P(a) \vee P(f(x))\}$ 等价于 $\{\neg P(x), P(a) \vee P(f(y))\}$ ，用代换 $\{a/x\}$ 和 $\{f(y)/x\}$ 作两次消解即可得到空字句，因此 $\{\neg P(x), P(a) \vee P(f(x))\}$ 不可满足。

(2) $\{\neg P(x) \vee P(f(x)), P(a)\}$ 可满足。

(3) $\{\neg P(x), P(f(x))\}$ 等价于 $\{\neg P(x), P(f(y))\}$ ，用代换 $\{f(y)/x\}$ 作消解即可得到空字句，因此 $\{\neg P(x), P(f(x))\}$ 不可满足。

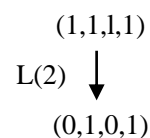
(4) $\{\neg P(x) \vee \neg P(f(x)), P(f(x))\}$ 等价于 $\{\neg P(y) \vee \neg P(f(y)), P(f(x))\}$ 用代换 $\{y/x\}$ 和 $\{f(x)/y\}$ 作两次消解即可得到空字句，因此 $\{\neg P(x) \vee \neg P(f(x)), P(f(x))\}$ 不可满足。

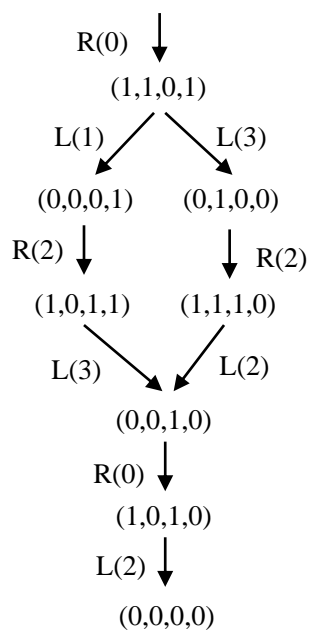
六、有一农夫带一只狐狸、一只小羊和一蓝菜过河（从左岸到右岸）。假设船太小，农夫每次只能带一样东西过河；考虑到安全，无农夫看管时，狐狸和小羊不能在一起，小羊和那蓝菜也不能在一起。请为该问题的解决设计状态空间，画出状态空间图，并用 A* 算法解决。(15 分)

第一步，定义问题的描述形式

用四元组 $S = (f, w, s, v)$ 表示问题状态，其中， f, w, s 和 v 分别表示农夫，狼，羊和青菜是否在左岸，它们都可以取 1 或 0，取 1 表示在左岸，取 0 表示在右岸。

该问题求解过程的状态空间图如下：



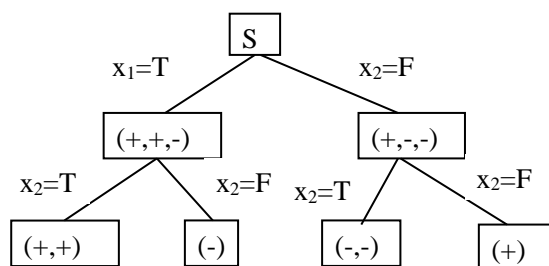


七、设训练例子集如下表所示：

序号	属性		分类
	x ₁	x ₂	
1	T	T	+
2	T	T	+
3	T	F	-
4	F	F	+
5	F	T	-
6	F	T	-

请用 ID3 和 c4.5 算法完成其学习过程。(15 分)

ID3 和 c4.5 都一样的，决策树如下图所示：



扩展 x₂ 后得到的完整决策树