1.1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | X | Y | Function f:X->Y | 数据集细节 |
| a | 病史和一些症状 | 患者是否患有疾病 | 根据临床诊断和经验判断患者是否患有本病 | 目标疾病的主要症状和病史需要尽可能清楚和正确 |
| b | 手写数字 | 识别数字结果 | 通过对数字特征的识别，对数字进行分类并得到结果 | 手写数字需要清晰，并尽可能多地涵盖不同的手写习惯 |
| c | 电子邮件和有关这些电子邮件的一些信息，如发件人、时间等 | 电子邮件是否为垃圾邮件 | 该功能可以根据一些关键词、发件人、标题或其他信息来判断电子邮件是否为垃圾邮件 | 需要更新关键字或发件人设置，以提高功能判断 |
| d | 一周中的价格、温度和天数的数据集 | 电力负荷结果 | 使用线性回归预测结果 | 为了使结果更加准确，需要尽可能多的数据 |
| e | 问题的信息和一些相关数据 | 这个问题的解决方案 | 利用我的经验或研究 | 信息应该与问题相关 |

1.2

|  |  |
| --- | --- |
| a | 学习方法 |
| b | 设计方法 |
| c | 学习方法 |
| d | 设计方法 |
| e | 设计方法 |

1.3

|  |  |  |
| --- | --- | --- |
|  | 学习类型 | 训练数据 |
| a | 监督学习 | 书籍的属性和用户的兴趣 |
| b | 强化学习，无监督学习 | 井字棋规则，重复训练 |
| c | 监督学习/无监督学习 | 电影的属性和不同类型的标签/使用无监督学习，不需要获得不同类型的标签) |
| d | 强化学习，无监督学习 | 如何播放音乐和乐谱(音符，乐谱规则) |
| e | 监督学习 | 最大允许债务与客户之间的关系(数据函数) |

1.4

2/3

1.5

1.6

B

1.7

B

1.8

A

1.9

C

1.10

B

1.11

B

1.12

D

1.13

C

1.14

A

1.15

CD

1.16

模型参数是指在训练过程中可以通过优化算法学习到的、用于决定模型预测值的内部变量。这些参数是模型的一部分，用于进行最终预测，因此它们的值在训练完毕后是固定的，不能再次修改

超参数是指我们在模型开发过程中需要手动调整的参数。这些参数不能够被模型训练过程自动学习到，而是需要手动选择不同的值进行实验找到一个最优的值

1.17

过拟合(or lucky)

解决方案：

1)获取更多数据

2)简化模型（选择更简单的算法、减少使用的参数或特征数量、对模型正则化）

3)减少训练数据中的噪声

1.18

亦称循环估计，是一种统计学上将数据样本切割成较小子集的实用方法（用来比较模型的技术，用于模型选择和超参数优化）

优势：

1)优化样本数据不足

2)小样本数据量的交叉验证可以用来选择不同模型，也可以用来选择同模型的不同超参数