

# 数理统计视角下的机器学习

Chen,Ruixi

2025

## 目录

<b>1 摘要</b>	<b>2</b>
<b>2 前言</b>	<b>2</b>
<b>3 感知机</b>	<b>2</b>
3.1 xxx 方程 . . . . .	2
3.2 xxx 情况下的边界条件和 xx 现象 . . . . .	2
3.3 xx 在 xxx 条件下的 xxx 现象 . . . . .	2
<b>4 支持向量机</b>	<b>3</b>
4.1 A. 在 xx 条件下测量 xxx . . . . .	3
4.1.1 a1. 计算出 xx 的电阻和电感 . . . . .	3
4.1.2 a2.Complete by yourself! . . . . .	3
4.1.3 a3.Complete by yourself! . . . . .	3
4.2 . . . . .	3
4.2.1 行间公式 . . . . .	3
4.2.2 相对于行内公式 . . . . .	3
<b>5 聚类</b>	<b>3</b>
5.1 误差分析 . . . . .	3
5.1.1 实验中的系统误差 . . . . .	3
5.1.2 实验中的偶然误差 . . . . .	3
5.2 实验后的思考 . . . . .	3
<b>6 主成分分析</b>	<b>4</b>
<b>7 朴素贝叶斯</b>	<b>4</b>
<b>8 决策树</b>	<b>4</b>
<b>9 回归</b>	<b>4</b>
<b>10 集成学习</b>	<b>4</b>

## 1 摘要

本文用于记录学习该课程的思考和作为复习资料。

## 2 前言

机器学习就是统计学的一个分支。

机器 (Learner) 通过学习收集到的数据，从而具备了决策能力。这与数理统计的目的是完全相同的，从样本推测总体，并利用统计决策理论进行推断。这与其他演绎式的数学是不相同的，统计的本质是归纳式的。

下面就从数理统计的角度来复习机器学习。

统计的第一步是抽样 (sampling)。样本是相互独立，且具有与总体相同的分布，这就是机器学习的重要前提：独立同分布 (independent and identically distributed, i.i.d.)。用向量来表示样本  $x_i$ ，且可能带有标记  $y_i$ (以监督学习为例)。抽样得到的样本构成训练集  $T = \{(x_i, y_i)\}$ (Training set)。无论是人还是机器，可能做出的决策  $f(x_i)$  都依赖于对  $T$  的分析。所有可能的决策函数  $f(x_i)$  (模型) 构成了假设空间  $F = \{f | y_i \simeq f(x_i)\}$ ，“学习”就是从假设空间中找出最优的决策函数，而判断优劣的依据就是风险函数  $R(f)$ 。这样，机器学习就变成了一个最优化问题：

运用算法找到风险函数的全局最优解，从而确定决策函数（模型）的参数。

总结，机器学习就是这样一个过程：为机器设定一个决策函数（模型），通过求解风险函数的全局最优解来确定最优的决策函数。

好，运用这个思路，我们来逐个复习机器学习中的各种算法。如 [1]

## 3 感知机

### 3.1 xxx 方程

在 xx, xxx, xxxx 条件下，考察条件为 xx 的 xx 的情况，利用 xxxx 定律在无位移的水平方向和有位移的竖直方向分别列出以下方程：

$$\begin{cases} T_2 \cos\alpha_2 - T_1 \cos\alpha_1 = 0 \\ T_2 \sin\alpha_2 - T_1 \sin\alpha_1 = \rho dx \frac{\partial^2 y}{\partial x^2} \end{cases} \quad (1)$$

$$(2)$$

### 3.2 xxx 情况下的边界条件和 xx 现象

xxxx 时发生 xxxx 现象。由 xxx 方程可知，xxx 波形为  $y^+ = f(vt + x)$ ，xxx 波形为  $y^- = f(vt - x)$ 。

### 3.3 xx 在 xxx 条件下的 xxx 现象

Complete by yourself!

## 4 支持向量机

### 4.1 A. 在 xx 条件下测量 xxx

#### 4.1.1 a1. 计算出 xx 的电阻和电感

在 xx 上将 xx 的两端串联 xx 和 xx 相连，将 xx 的两端串联进 xx，分别将 xx 接在  $L_1$ ,  $L_2$ , xx 的两端测量 xx 并记录。

#### 4.1.2 a2.Complete by yourself!

Complete by yourself!

#### 4.1.3 a3.Complete by yourself!

实验得到的数据如下：

线圈名称	$R'(\Omega)$	$V_a(V)$	$V(V)$	$V_r'(V)$	$V_o(V)$
线圈 1(空气芯)	123	456	789	012	345
线圈 2(空气芯)	123	456	789	012	345
线圈 3(铝芯)	123	456	789	012	345
线圈 4(铝芯)	123	456	789	012	345

### 4.2

#### 4.2.1 行间公式

这是一个不确定度计算。

$$U_k = tinv(x, y) \times s_k = xxx$$

#### 4.2.2 相对于行内公式

这是一个不确定度计算：  $U_k = tinv(x, y) \times s_k = xxx$

## 5 聚类

### 5.1 误差分析

#### 5.1.1 实验中的系统误差

来自 xxx 的精度影响。

受空间内 xx 与 xx 的干扰。

#### 5.1.2 实验中的偶然误差

接线时可能有 xxxx 情况，导致 xxx。xx 上的 xx 在某情况下有 xx 的问题存在，经反复调整后得以正常测量。

### 5.2 实验后的思考

可说明自己做本实验的总结、收获和体会，对实验中发现的问题提出自己的建议。

## 6 主成分分析

Change the picture by yourself!



示例图片

## 7 朴素贝叶斯

## 8 决策树

## 9 回归

## 10 集成学习

## 参考文献

- [1] 王合英, 孙文博, 陈宜保, 陈宏, 张慧云, and 张留碗. 自主探究实验对学生综合素质和创新能力的培养. 实验技术与管理, 035(012):24–28, 2018.