

class 13 第一阶段总复习

模型分类

- 描述与理解 (Description)
- 估计与预测 (Estimation) – 水温下降 (数列)、污染物预测 (微分方程, 函数拟合), 房价预测 (线性回归、决策树-机器学习算法)、刹车问题 (函数拟合)、未成年人身高体重
- 评价与决策 (Decision) – 电动车选择 (加权求和, 权重设定AHP,EWM, TOPSIS)、岗位匹配 (线性规划)
- 溯因与解释 (Explanation)

重温水温下降

冰箱温度为8摄氏度, 饮料现在温度是26摄氏度, 已知经过十分钟, 饮料温度降到了20摄氏度, 请预测多久饮料会降低到10摄氏度。

假设: 温度的下降速率与物体与周围温度的差值成正比

符号: T : 物体的温度, t : 时间(min), T_e : 环境温度, T_0 : 初始温度

模型

$$\frac{dT}{dt} = -k(T - T_e) \rightarrow T = T_e + (T_0 - T_e)e^{-kt}$$

已知 $T_e = 8, T_0 = 26, T(10) = 20$, 可以算出 $k =$

重温线性拟合

模数 为了研究车速和刹车距离之间的数量关系, 收集了一组实验数据如下:

车速 /km · h ⁻¹	20	40	60	80	100	120	140
刹车距离 /m	6.5	18.8	33.6	58.1	83.4	118.0	153.5

他想要得到车速与刹车距离之间的关系。 □

拟合的底层原理: 最小二乘法 (least square method)

假设现在有数据点 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 我们假设的函数为 $y = f(x)$, 最小二乘法指的是

$$\min L = \sum_{i=1}^n (y_i - f(x_i))^2$$

如果我们进行线性拟合, 则

$f(x) = ax + b$, 我们希望找到

$$\min L = \sum_{i=1}^n (y_i - ax_i - b)^2$$

```
1 from scipy.optimize import curve_fit
```

重温房价预测

对于一个机器学习项目，大致步骤如下：

- 获取数据
- 对数据进行预处理、探索性数据分析
- 对数据集进行拆分（训练集和测试集）
- 调用机器学习模型（线性回归、决策树、支持向量机SVR、随机森林、朴素贝叶斯、神经网络、K近邻、梯度提升树、XGBoost、LightGBM等）
- 对模型进行评估

```
1 from sklearn.ensemble import RandomForest # 引入模型
```

重温电动汽车品牌选择

评价问题

评价系统的5要素：

- 评价指标
- 评价对象
- 评价者
- 权重
- 综合方式

评价指标：全面、可操作

权重：AHP,熵权法，变异系数法，主成分法PCA

综合方式：加（加权平均）、减（好处-坏处）、乘、除（比率）、距离

重温线性规划

汽车厂生产三种类型的汽车，已知各类型每辆车对钢材、劳动时间的需求，利润及工厂每月的现有量。

	小型	中型	大型	现有量
钢材（吨）	1.5	3	5	600
劳动时间（小时）	280	250	400	60000
利润（万元）	2	3	4	

- 制订月生产计划，使工厂的利润最大。
- 如果生产某一类型汽车，则至少要生产80辆，那么最优的生产计划应作何改变？

规划问题三要素：

- 目标函数
- 约束条件
- 决策变量

设决策变量为小、中、大的车型数量，分别为 x, y, z

使用pulp库

```
1 import pulp as lp
```

辛普斯悖论

医院	存活人数	死亡人数	存活率
A	900	100	90%
B	800	200	80%

分组存活率

医院	病情	入院人数	存活人数	死亡人数	存活率
A	重症	100	30	70	30%
A	非重症	900	870	30	96.7%
B	重症	400	210	190	52.5%
B	非重症	600	590	10	98.3%

建议

<https://reformship.github.io/pages/32himcm.html>

选10篇论文进行阅读，联系overleaf的使用，写一篇overleaf的论文