



回归分析：回归分析基础理论

回归分析简介

1. 回归分析的定义

回归分析是一种统计学上的方法，用于探索变量之间的关系，特别是因变量与一个或多个自变量之间的关系。它可以帮助我们理解当自变量变化时，因变量如何变化。回归分析的目标是建立一个数学模型，描述这种关系，并利用模型进行预测或解释。

2. 回归分析的应用领域

回归分析广泛应用于多个领域，包括但不限于：

- 经济学：预测销售量、分析价格对需求的影响。
- 医学：研究药物剂量与治疗效果之间的关系。
- 社会科学：分析教育水平对收入的影响。
- 工程学：预测设备的故障率与使用时间的关系。
- 环境科学：研究污染物浓度与健康影响之间的关系。

3. 回归分析的类型

3.1 线性回归

线性回归是最基本的回归分析类型，它假设因变量与自变量之间存在线性关系。线性回归模型可以是简单的（只有一个自变量）或多元的（有多个自变量）。

示例：简单线性回归

假设我们有一组数据，表示房屋面积（平方米）与价格（万元）之间的关系，我们想要建立一个模型来预测价格。

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# 示例数据
area = np.array([50, 60, 70, 80, 90, 100, 110, 120, 130,
                 140]).reshape((-1, 1))
price = np.array([30, 35, 40, 45, 50, 55, 60, 65, 70, 75])

# 创建线性回归模型
```

```

model = LinearRegression()

# 训练模型
model.fit(area, price)

# 预测价格
predicted_price = model.predict(np.array([150]).reshape((-1, 1)))

# 打印预测结果
print("预测价格：", predicted_price[0])

# 绘制数据点和回归线
plt.scatter(area, price, color='blue')
plt.plot(area, model.predict(area), color='red')
plt.title('房屋面积与价格关系')
plt.xlabel('面积（平方米）')
plt.ylabel('价格（万元）')
plt.show()

```

3.2 多元线性回归

多元线性回归扩展了简单线性回归，允许模型包含多个自变量。例如，预测房价可能不仅考虑面积，还会考虑位置、房龄等因素。

示例：多元线性回归

假设我们有房屋面积、房龄和位置等级的数据，我们想要预测房价。

```

import pandas as pd
from sklearn.linear_model import LinearRegression

# 示例数据
data = {
    'Area': [50, 60, 70, 80, 90],
    'Age': [10, 15, 20, 25, 30],
    'Location': [3, 4, 5, 4, 3],
    'Price': [30, 35, 40, 45, 50]
}
df = pd.DataFrame(data)

# 创建多元线性回归模型
model = LinearRegression()

# 训练模型

```

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/318002035024006111>