
Q01: 基础问题 1

prompt: 什么是回归分析？请描述其最简单的形式。

model_name: GPT

model_version: 4o

response: 回归分析是一种统计方法，用于建模因变量（目标变量）与一个或多个自变量（预测变量）之间的关系。最简单的回归形式是线性回归，假设因变量与自变量之间存在线性关系，公式为：

$$Y = \beta_1 + \beta_2 X$$

其中，Y 为因变量，X 为自变量， β_1 和 β_2 为模型参数。

Q02: 基础问题 2

prompt: 什么是普通最小二乘法（OLS），其目标是什么？

model_name: GPT

model_version: 4o

response: 普通最小二乘法（OLS）是一种用于估计回归模型参数的统计技术，其目标是通过最小化预测值与实际值之间的误差平方和，找到最佳拟合线。公式为：

$$Q(\beta) = (Y - X\beta)^T(Y - X\beta)$$

通过求解方程 $\frac{\partial Q}{\partial \beta} = 0$ ，可以得到参数的最优估计值。

Q03: 扩展问题 1

prompt: 在回归分析中，什么是偏差-方差权衡？其对模型复杂度的影响是什么？

model_name: GPT

model_version: 4o

response: 偏差-方差权衡是回归模型在预测新数据时的一种性能指标平衡。模型的偏差反映了预测与真实值之间的系统性误差，而方差表示模型对训练数据的敏感程度。较简单的模型通常具有较高的偏差但较低的方差，而复杂模型则相反。最佳模型在偏差和方差之间达到平衡，从而最小化总体误差。

Q04: 扩展问题 2

prompt: 请解释支持向量机（SVM）回归中“软间隔”概念及其引入的原因。

model_name: GPT

model_version: 4o

response: 在SVM回归中，“软间隔”用于处理线性不可分的数据。通过引入松弛变量 ξ_i 和 ξ_i^* ，允许部分数据点在容许误差范围 ϵ 之外，以减少对模型的严格限制。优化目标变为最小化模型复杂度和误差之和，公式为：

$$\min_{|W|} \frac{1}{2} \|W\|^2 + C \sum (\xi_i + \xi_i^*)$$

引入“软间隔”提高了模型的鲁棒性，允许在存在噪声和异常值的情况下进行回归。

Q05: 前沿问题

prompt: 在深度学习回归模型中，为什么需要使用“正则化”技术？请举例说明。

model_name: GPT

model_version: 4o

response: 在深度学习回归模型中，正则化用于减少过拟合，从而提高模型在未见数据上的泛化能力。常用的正则化技术包括：

- **LASSO (L1正则化)** : 最小化误差平方和加上权重的L1范数, 鼓励稀疏模型, 公式为:

$$\min Q(\beta) + \lambda \|\beta\|_1$$
- **Ridge (L2正则化)** : 加入权重的L2范数, 减少模型的复杂度, 公式为: $\min Q(\beta) + \lambda \|\beta\|_2^2$

这些方法帮助模型在高维数据和有限训练样本中表现更好。