

title: [W16--提ch03]高级算法大模型辅助学习 author: - name: 胡昊^[2024244066,计算机科学与技术]

toc: false

toc-title: "目录"

filters:

- scholarly-metadata.lua
- author-info-blocks.lua

Q01：Qname

- prompt：解释并举例说明Linear regression、Nonlinear regression以及Nonparametric regression
- model\_name：chatgpt
- model\_version：gpt-4o-min
- response：在统计建模和机器学习中，**回归分析**（Regression Analysis）用于探索自变量（输入）与因变量（输出）之间的关系。根据模型的形式和假设，回归可以分为**线性回归**（Linear Regression）、**非线性回归**（Nonlinear Regression）和**非参数回归**（Nonparametric Regression）。下面对这三者进行解释，并提供示例。

1. 线性回归 ( Linear Regression)

解释

线性回归假设因变量  $Y$  与自变量  $X$  之间的关系可以用**线性函数**来表示。基本形式为：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

其中：

- $\beta_0$  为截距 ( Intercept ) 。
- $\beta_1, \beta_2, \cdots, \beta_p$  为回归系数 ( Regression Coefficients ) 。
- $\epsilon$  为误差项 ( Error Term ) ，用来捕捉模型无法解释的部分。

特点

- 假设模型是线性的，即参数  $\beta_i$  在线性关系中出现。
- 可以使用**最小二乘法** ( Ordinary Least Squares, OLS ) 来估计参数。

示例

假设你想研究房屋面积 ( 平方英尺 ) 与房价之间的关系。一个简单的线性回归模型可以是：

$$\text{Price} = \beta_0 + \beta_1 \times \text{Area} + \epsilon$$

例如，模型拟合结果可能是：

$$\text{Price} = 50,000 + 200 \times \text{Area}$$

这表示每增加1平方英尺，房价增加200元。

---

## 2. 非线性回归 ( Nonlinear Regression)

### 解释

非线性回归假设因变量  $Y$  和自变量  $X$  之间的关系无法用线性函数来表示，而是通过**非线性函数**来建模。非线性回归模型的形式为：

$$Y = f(X, \theta) + \epsilon$$

其中  $f$  是一个非线性函数， $\theta$  是参数向量。

### 特点

- 模型中的参数以非线性方式出现。
- 参数估计需要使用迭代优化方法（如**梯度下降法**、**牛顿法**等）。
- 常见的非线性函数包括指数函数、幂函数、对数函数等。

### 示例

研究人口增长时，可能采用非线性回归模型，比如**逻辑增长模型**（Logistic Growth Model）：

$$Y = \frac{K}{1 + e^{-(\theta_1 + \theta_2 X)}}$$

其中：

- $K$  是最大人口容量。
- $\theta_1, \theta_2$  是待估参数。

---

## 3. 非参数回归 ( Nonparametric Regression)

### 解释

非参数回归不对因变量  $Y$  和自变量  $X$  之间的函数形式做任何假设。它通过数据驱动的方式来估计回归关系，例如**核回归**（Kernel Regression）、**局部加权回归**（Locally Weighted Regression, LOWESS）和**样条回归**（Spline Regression）。

### 特点

- 不需要预先指定回归函数的形式。
- 更灵活，可以捕捉复杂的关系。
- 需要大量数据来进行有效估计。

### 示例

假设你想研究温度和冰淇淋销售额之间的关系，但这种关系可能不是线性的。你可以使用**核回归**来估计  $Y$  与  $X$  之间的关系。非参数回归的结果会根据数据点自动拟合一条平滑曲线。

## 比较总结

类别	特点	适用场景
线性回归	假设关系是线性的，简单且易解释	关系简单、线性趋势明显
非线性回归	假设特定的非线性函数形式	已知存在特定非线性关系的场景，如人口增长模型
非参数回归	不假设特定函数形式，灵活但需要更多数据	关系复杂或无法预先设定函数形式的场景

以上就是线性回归、非线性回归和非参数回归的基本解释和示例。根据实际数据特征和研究需求，可以选择合适的回归方法。

### Q02：Qname

- prompt：解释最小二乘法的原理
- model\_name：chatgpt
- model\_version：gpt-4o-min
- response：最小二乘法（**Ordinary Least Squares, OLS**）是一种用于估计线性回归模型中参数的方法。其基本原理是**最小化预测值与实际值之间残差（误差）的平方和**，从而找到最佳拟合线。

## 1. 问题描述

假设我们有  $n$  个观测数据点  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ，我们试图用以下简单线性回归模型来描述这些数据：

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (i = 1, 2, \dots, n)$$

其中：

- $y_i$  是第  $i$  个观测值。
- $x_i$  是第  $i$  个自变量。
- $\beta_0$  是截距（Intercept）。
- $\beta_1$  是回归系数（Slope）。
- $\epsilon_i$  是第  $i$  个残差或误差项，表示实际值与预测值之间的差异。

我们希望找到参数  $\beta_0$  和  $\beta_1$ ，使得模型的预测值  $\hat{y}_i = \beta_0 + \beta_1 x_i$  与实际值  $y_i$  之间的误差尽可能小。

## 2. 残差和残差平方和

残差（Residual）

第  $i$  个数据点的残差（误差） $\epsilon_i$  定义为：

$$\epsilon_i = y_i - \hat{y}_i = y_i - (\beta_0 + \beta_1 x_i)$$

残差平方和 ( Sum of Squared Residuals, SSR)

最小二乘法的目标是最小化所有残差的平方和  $S$ ，即：

$$S = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

这样做的原因是，残差的平方消除了正负号的影响，并惩罚较大的误差。

### 3. 最小二乘估计的求解过程

通过最小化  $S$  来求解参数  $\beta_0$  和  $\beta_1$ 。数学上，我们可以通过对  $S$  分别对  $\beta_0$  和  $\beta_1$  求导，并令导数为零来找到最优解。

导数求解

1. 对  $\beta_0$  求偏导：

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0$$

2. 对  $\beta_1$  求偏导：

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

解出参数

将上述方程组求解，可以得到最小二乘法的闭合解：

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

其中：

- $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  是自变量的均值。
- $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  是因变量的均值。

### 4. 直观理解

图形解释

- 数据点分布**：数据点  $(x_i, y_i)$  在二维平面上散布。
- 拟合直线**：最小二乘法找到一条直线  $y = \beta_0 + \beta_1 x$ ，使得所有数据点到直线的垂直距离的平方和最小。
- 残差最小化**：通过调整直线的截距和斜率，最小二乘法使得这些垂直残差的平方和最小。

例子

假设我们有以下3个数据点： $(1, 2), (2, 3), (3, 5)$ 。我们希望找到最佳拟合直线 $y = \beta_0 + \beta_1 x$ 。

1. 计算均值：

◦  $\bar{x} = \frac{1+2+3}{3} = 2$

◦  $\bar{y} = \frac{2+3+5}{3} = 3.33$

2. 计算  $\beta_1$ ：

$$\beta_1 = \frac{(1-2)(2-3.33) + (2-2)(3-3.33) + (3-2)(5-3.33)}{(1-2)^2 + (2-2)^2 + (3-2)^2} = 1.5$$

3. 计算  $\beta_0$ ：

$$\beta_0 = 3.33 - 1.5 \times 2 = 0.33$$

4. 拟合直线：

$$y = 0.33 + 1.5x$$

---

5. 优点和局限性

优点

- 计算简单，容易理解和实现。
- 在误差项满足正态分布且方差相等的情况下，最小二乘法具有最优性。

局限性

- 对异常值 (Outliers) 敏感，异常点可能显著影响拟合结果。
- 仅适用于线性关系，若数据具有非线性关系，需要使用非线性回归或其他方法。

---

总结来说，最小二乘法通过最小化残差平方和来找到最佳的拟合直线，是线性回归中最常用的方法之一。

Q03：Qname

- prompt：广义最小二乘法和普通最小二乘法有什么区别
- model\_name：chatgpt
- model\_version：gpt-4o-min
- response：广义最小二乘法 (Generalized Least Squares, GLS) 和普通最小二乘法 (Ordinary Least Squares, OLS) 都是用于回归模型参数估计的方法，但它们在假设条件和适用场景上有所不同。下面详细解释这两者的区别，包括其原理、假设、适用场景和应用举例。

---

1. 普通最小二乘法 (OLS)

原理

OLS 通过最小化残差平方和来估计回归模型的参数。其目标函数为：

$$S = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

## 假设条件

OLS 在以下假设下具有最优性（无偏、方差最小）：

- 线性关系**：因变量  $y$  与自变量  $x$  之间存在线性关系。
- 同方差性**（Homoscedasticity）：误差项  $\epsilon_i$  具有相同的方差，即  $Var(\epsilon_i) = \sigma^2$ 。
- 无自相关性**：误差项之间相互独立，即  $Cov(\epsilon_i, \epsilon_j) = 0$ （当  $i \neq j$ ）。
- 正态性**：误差项服从正态分布（在大样本情况下，这一假设可以放宽）。

## 适用场景

- 数据满足上述假设条件时，OLS 是最佳线性无偏估计（BLUE）。
- 当误差项具有相同方差且相互独立时，OLS 效率最高。

## 优点和局限

- 优点**：简单直观，计算方便。
- 局限**：对异方差性和自相关性敏感，如果误差项不满足同方差性或独立性，OLS 参数估计仍然无偏，但不再是最优估计，可能导致标准误估计错误，影响统计检验。

## 2. 广义最小二乘法（GLS）

### 原理

GLS 是对 OLS 的推广，专门用来处理误差项具有异方差性或自相关性的情况。GLS 的目标是在误差项的协方差结构已知的情况下，最小化加权残差平方和。

GLS 的目标函数为：

$$S = (y - X\beta)^T \Omega^{-1} (y - X\beta)$$

其中：

- $y$  是  $n \times 1$  的因变量向量。
- $X$  是  $n \times p$  的自变量矩阵。
- $\beta$  是  $p \times 1$  的回归系数向量。
- $\Omega$  是  $n \times n$  的误差项的协方差矩阵。

## 假设条件

- 异方差性或自相关性**：误差项  $\epsilon$  的协方差矩阵  $\Omega$  非对角阵，可以有不同的方差或相关性，即  $Var(\epsilon) = \Omega$ 。
- 已知协方差结构**：误差项的协方差矩阵  $\Omega$  已知或可估计。

## 适用场景

- 当误差项存在异方差性（方差不相等）或自相关性（误差项之间存在相关）时，OLS 估计的效率会降低，而 GLS 可以提供更高效的估计。
- 例如：时间序列数据中存在自相关性，或者截面数据中不同观测点的误差方差不同。

## 优点和局限

- 优点：**在误差项存在异方差性或自相关性的情况下，GLS 提供无偏且最优的线性估计。
- 局限：**
  - 需要知道或估计误差项的协方差结构  $\Omega$ ，在实际应用中可能较难获取。
  - 如果误差协方差矩阵估计不准确，GLS 的结果也会受到影响。

## 3. OLS 与 GLS 的关系

### 1. 在特殊情况下，GLS 退化为 OLS：

如果误差项满足同方差性和无自相关性，即  $\Omega = \sigma^2 I$ （ $I$  是单位矩阵），那么 GLS 就简化为 OLS。

### 2. OLS 的效率问题：

如果误差项不满足 OLS 假设（存在异方差性或自相关性），OLS 估计仍然是无偏的，但不再具有最小方差，而 GLS 在这种情况下可以提高估计的效率。

## 4. 示例说明

### OLS 示例

假设我们有一组数据，表示房屋面积和房价之间的关系：

面积 (平方英尺) $x_i$	房价 (万元) $y_i$
1000	150
1500	200
2000	250
2500	300

可以使用 OLS 拟合线性模型：

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

### GLS 示例

如果在数据中发现残差的方差随着房屋面积的增加而增加（即存在**异方差性**），我们可以假设误差的协方差矩阵  $\Omega$  为：

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 & \sigma_2^2 & 0 & 0 & 0 & 0 & \sigma_3^2 & 0 & 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

在这种情况下，使用 GLS 可以更有效地估计回归系数。

## 5. 总结

比较维度	普通最小二乘法 (OLS)	广义最小二乘法 (GLS)
误差假设	同方差性、无自相关性	可以存在异方差性或自相关性
目标	最小化残差的平方和	最小化加权残差的平方和
协方差矩阵	$\Omega = \sigma^2 I$	$\Omega$ 为任意正定矩阵
计算复杂度	计算简单，直接解闭合解	计算较复杂，需要估计误差的协方差矩阵
应用场景	误差方差一致且相互独立	误差方差不一致或存在相关性
结果特性	在假设满足时是最佳线性无偏估计 (BLUE)	在误差结构已知时提供无偏且最优的线性估计

根据数据特性和误差的结构，合理选择 OLS 或 GLS 方法可以提高回归模型的估计效果和预测精度。

Q04：Qname

- prompt：采用极大似然估计来建模线性回归的一般步骤是什么
- model\_name：chatgpt
- model\_version：gpt-4o-min
- response：在线性回归模型中，**极大似然估计**（Maximum Likelihood Estimation, MLE）是一种用于估计回归系数的方法。MLE 通过最大化观测数据的似然函数来找到参数的估计值。与最小二乘法（OLS）不同，极大似然估计需要对误差项的分布做具体假设（通常假设误差项服从正态分布）。

以下是采用**极大似然估计建模线性回归**的一般步骤。

1. 定义线性回归模型

假设我们有  $n$  个观测数据点  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ，线性回归模型形式为：

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$$

其中：

- $y_i$  是第  $i$  个观测的因变量。
- $x_{ij}$  是第  $i$  个观测的第  $j$  个自变量。
- $\beta_0$  是截距（Intercept）。
- $\beta_j$  是回归系数。
- $\epsilon_i$  是误差项，通常假设  $\epsilon_i \sim N(0, \sigma^2)$ ，即误差项服从均值为0、方差为 $\sigma^2$ 的正态分布。

2. 写出似然函数

由于假设误差项  $\epsilon_i$  服从正态分布，因此  $y_i$  也服从正态分布：

$$y_i \sim N(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \sigma^2)$$

每个  $y_i$  的概率密度函数为：

$$f(y_i|\beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2}{2\sigma^2}\right)$$



假设观测值是独立的，所有观测数据的联合概率（即似然函数  $L$ ）为：

$$L(\beta_0, \beta_1, \dots, \beta_p, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2}{2\sigma^2}\right)$$

### 3. 取对数似然函数

为了简化计算，通常对似然函数取对数，得到对数似然函数  $\ln L$ ：

$$\ln L = \sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2}{2\sigma^2}\right)\right)$$

化简得到：

$$\ln L = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$$

### 4. 最大化对数似然函数

为了找到回归系数的极大似然估计，我们需要最大化对数似然函数  $\ln L$ 。这通常通过求对数似然函数关于参数的偏导数，并令其等于零来完成。

#### 步骤

1. 对回归系数  $\beta_j$  求偏导：

$$\frac{\partial \ln L}{\partial \beta_j} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip}) x_{ij}$$

令  $\frac{\partial \ln L}{\partial \beta_j} = 0$  来求解  $\beta_j$ 。

2. 对方差  $\sigma^2$  求偏导：

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$$

令  $\frac{\partial \ln L}{\partial \sigma^2} = 0$  来求解  $\sigma^2$ 。

### 5. 求解极大似然估计值

通过求解上述方程组，可以得到参数的极大似然估计值：

- 回归系数的估计值： $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$
- 误差方差的估计值： $\hat{\sigma}^2$

值得注意的是，在假设误差项服从正态分布的情况下，MLE 得到的回归系数估计与 OLS 得到的结果是相同的。

## 6. 总结步骤

1. 定义线性回归模型并假设误差项服从正态分布。
2. 写出似然函数，表示所有观测数据的联合概率。
3. 取对数似然函数，简化计算。
4. 对对数似然函数求导，对回归系数和方差分别求偏导。
5. 求解方程组，找到使对数似然函数最大的参数估计值。

---

## 7. 与最小二乘法的关系

- 当误差项服从正态分布时，极大似然估计和最小二乘法估计结果相同。
- MLE 提供了一个**概率统计基础**，可以在误差项不服从正态分布时推广到更复杂的模型（如广义线性模型、Logistic回归等）。

通过极大似然估计方法，线性回归模型的参数估计在统计学上有更强的理论支持，特别是在误差分布已知或需要进行统计推断的情况下。

### Q05：Qname

- prompt：如何判断一个线性回归模型建模的好坏？
- model\_name：chatgpt
- model\_version：gpt-4o-min
- response：判断一个线性回归模型建模的好坏需要综合考虑多个统计指标、诊断图表和模型假设的验证。以下是一些常用的方法和指标，可以帮助评估线性回归模型的质量和有效性。

---

## 1. 拟合优度指标

### 1.1 决定系数 $R^2$

- 定义： $R^2$  ( $R$ -squared) 表示模型解释因变量波动的比例。
- 公式：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

其中：

- $y_i$  是实际值。
- $\hat{y}_i$  是预测值。
- $\bar{y}$  是  $y_i$  的均值。
- 取值范围： $0 \leq R^2 \leq 1$ ，越接近 1，说明模型解释能力越强。
- 注意： $R^2$  不会惩罚模型中过多的变量，可能导致过拟合。

### 1.2 调整后的 $R^2$ (Adjusted $R^2$ )

- 定义：调整后的  $R^2$  对模型中的变量数量进行惩罚，适合用于多元回归。

$$R_{\text{adj}}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}$$

其中：

- $n$  是样本数量。
- $p$  是自变量数量。
- **优势：**当加入无用变量时，调整后的  $R^2$  会降低，有助于模型选择。

---

## 2. 误差分析

### 2.1 残差标准误 (Residual Standard Error, RSE)

- **定义：**RSE 是残差的标准差，用于衡量模型预测的误差大小。

$$RSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p - 1}}$$

- **解释：**RSE 越小，说明模型预测误差越小。

### 2.2 平均绝对误差 (Mean Absolute Error, MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

### 2.3 均方误差 (Mean Squared Error, MSE) 与均方根误差 (Root Mean Squared Error, RMSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad RMSE = \sqrt{MSE}$$

- **解释：**RMSE 和 MSE 越小，模型预测精度越高。

---

## 3. 回归系数的显著性检验

### 3.1 $t$ 检验

- **目的：**检验回归系数是否显著不为零。
- **假设：**
  - $H_0$ ：回归系数  $\beta_j = 0$  (该变量对因变量无显著影响)。
  - $H_1$ ：回归系数  $\beta_j \neq 0$ 。
- **解释：**如果  $p$ -值小于显著性水平 (如 0.05)，则拒绝  $H_0$ ，说明该回归系数显著。

### 3.2 F 检验

- **目的：**整体检验模型中所有自变量是否显著。

- 假设：
    - $H_0$ ：所有回归系数均为零，即模型无效。
    - $H_1$ ：至少有一个回归系数不为零。
  - 解释：如果  $p$ -值小于显著性水平，说明模型整体显著。
- 

## 4. 诊断图表

### 4.1 残差图 (Residual Plot)

- 目的：检查残差的分布，验证模型假设。
- 观察要点：
  - 残差应随机分布，均匀分布在 0 轴上下。
  - 如果残差图显示模式（如弧形、锥形），可能存在异方差性或非线性关系。

### 4.2 QQ 图 (Quantile-Quantile Plot)

- 目的：检查残差是否符合正态分布。
- 观察要点：
  - 残差点应接近一条直线。如果偏离直线，说明残差不满足正态性假设。

### 4.3 标准化残差与预测值图

- 目的：检查是否存在异方差性。
- 观察要点：
  - 如果残差的分散程度随预测值增大而变化，可能存在异方差性。

### 4.4 残差自相关图

- 目的：检查残差是否存在自相关性。
  - 方法：绘制残差的自相关函数 (ACF) 图或进行 Durbin-Watson 检验。
- 

## 5. 模型假设验证

### 5.1 线性性假设

- 方法：检查残差图，看是否存在非线性模式。

### 5.2 同方差性假设 (Homoscedasticity)

- 方法：残差图、Breusch-Pagan 检验或 White 检验。

### 5.3 正态性假设

- 方法：QQ 图、Shapiro-Wilk 检验、Kolmogorov-Smirnov 检验。

### 5.4 无自相关性假设

- 方法：Durbin-Watson 检验。

---

## 6. 过拟合与模型复杂度

### 6.1 交叉验证 ( Cross-Validation )

- 方法：如  $k$ -折交叉验证，评估模型在未见数据上的表现。
- 目的：判断模型是否过拟合。

### 6.2 信息准则

- AIC (Akaike Information Criterion)** 和 **BIC (Bayesian Information Criterion)**：
  - 用于模型选择，较小的 AIC 或 BIC 值表示模型更优。
  - AIC 和 BIC 对模型复杂度有惩罚。

---

## 7. 实际应用效果

### 7.1 预测效果

- 使用测试集或新数据评估模型的预测精度。

### 7.2 业务合理性

- 确保模型的回归系数和结果在实际业务情境中具有解释性和合理性。

---

## 总结

评估线性回归模型的好坏需要综合考虑多个方面，包括：

- 拟合优度指标 ( $R^2$ 、调整后的  $R^2$ )。
- 误差大小 (RSE、MAE、RMSE)。
- 回归系数和模型显著性检验 ( $t$  检验、 $F$  检验)。
- 残差分析和诊断图表。
- 模型假设的验证 (线性性、正态性、同方差性、无自相关性)。
- 避免过拟合 (交叉验证、AIC/BIC)。
- 实际应用效果。

通过综合评估这些指标和方法，可以判断模型是否适合实际应用，并进一步优化模型。