### 分析流程 数据源： normalized.csv 算法配置： 算法： 随机森林回归 变量： 变量X：{ x1，x2，x3 }；变量Y：{ y7 } 分析结果： 随机森林回归基于MSE、RMSE、MAE、MAPE、R²指标对模型进行评价，请看详细结论。

### 分析步骤 1. 通过训练集数据来建立随机森林回归模型。 2. 通过建立的随机森林来计算特征重要性。 3. 将建立的随机森林回归模型应用到训练、测试数据，得到模型评估结果。 4. 由于随机森林具有随机性，每次运算的结果不一样，若保存本次训练模型，后续可以直接上传数据代入到本次训练模型进行计算预测。 5. 注：随机森林无法像传统模型一样得到确定的方程，通常通过测试数据预测精度来对模型进行评价。

### 详细结论

**输出结果1：模型参数**

|  |  |
| --- | --- |
| 参数名 | 参数值 |
| 训练用时 | 0.448s |
| 数据切分 | 0.7 |
| 数据洗牌 | 是 |
| 交叉验证 | 5 |
| 节点分裂评价准则 | mse |
| 划分时考虑的最大特征比例 | None |
| 内部节点分裂的最小样本数 | 2 |
| 叶子节点的最小样本数 | 1 |
| 叶子节点中样本的最小权重 | 0 |
| 树的最大深度 | 10 |
| 叶子节点的最大数量 | 50 |
| 节点划分不纯度的阈值 | 0 |
| 决策树数量 | 100 |
| 有放回采样 | true |
| 袋外数据测试 | false |

**图表说明：**

上表展示了模型各项参数配置以及模型训练时长。

**输出结果2：特征重要性**

|  |  |
| --- | --- |
| 特征名称 | 特征重要性 |
| x1 | 10.70% |
| x2 | 8.00% |
| x3 | 81.30% |

**图表说明：**

上柱形图或表格展示了各特征（自变量）的重要性比例。

**输出结果3：模型评估结果**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | MSE | RMSE | MAE | MAPE | R² |
| 训练集 | 0.022 | 0.148 | 0.122 | 16.113 | 0.975 |
| 交叉验证集 | 0.253 | 0.452 | 0.412 | 81.157 | -2.953 |
| 测试集 | 0.258 | 0.508 | 0.436 | 53.891 | 0.764 |

**图表说明：**

上表中展示了交叉验证集、训练集和测试集的预测评价指标，通过量化指标来衡量随机森林的预测效果。其中，通过交叉验证集的评价指标可以不断调整超参数，以得到可靠稳定的模型。  
● MSE（均方误差）： 预测值与实际值之差平方的期望值。取值越小，模型准确度越高。  
● RMSE（均方根误差）：为MSE的平方根，取值越小，模型准确度越高。  
● MAE（平均绝对误差）： 绝对误差的平均值，能反映预测值误差的实际情况。取值越小，模型准确度越高。  
● MAPE（平均绝对百分比误差）： 是 MAE 的变形，它是一个百分比值。取值越小，模型准确度越高。  
● R²： 将预测值跟只使用均值的情况下相比，结果越靠近 1 模型准确度越高。  
● oob\_score：对于回归问题，oob\_score是袋外数据的R²。若在建立树过程中选择有放回抽样时，大约1/3的记录没有被抽取。没有被抽取的自然形成一个对照数据集，可用于模型的验证。所以随机森林不需要另外预留部分数据做交叉验证，其本身的算法类似交叉验证，而且袋外误差是对预测误差的无偏估计（当算法参数选择了“袋外测试数据”后，才会通过oob\_score来检验模型的泛化能力）。

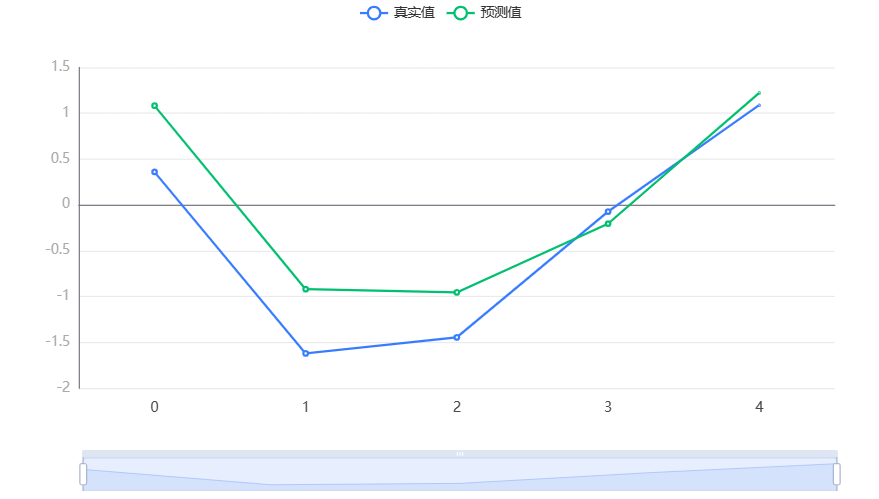
**输出结果4：预测结果**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 预测测试集结果Y | y7 | x1 | x2 | x3 |
| 1.0829782555700007 | 0.360097692 | -1.341640786 | 0.447213595 | 0.447213595 |
| -0.9203661028100012 | -1.621418665 | -0.447213595 | -0.447213595 | -1.341640786 |
| -0.9556047658000018 | -1.446195716 | 1.341640786 | 1.341640786 | -1.341640786 |
| -0.20560346129000004 | -0.073523633 | 0.447213595 | 1.341640786 | -0.447213595 |
| 1.2226603235600004 | 1.08871459 | -1.341640786 | 1.341640786 | 1.341640786 |

**图表说明：**

上表格为预览结果，只显示部分数据，全部数据请点击下载按钮导出。  
上表展示了随机森林模型对测试数据的预测情况。

**输出结果5：测试数据预测图**



**图表说明：**

上图中展示了随机森林模型对测试数据的预测情况。预测图最多只展现测试集前1000个样本预测信息，若需要绘制完整的预测图，可在预测结果中导出数据重新绘图。

**输出结果6：模型预测与应用**

**图表说明：**

如果预测的数据超过15条，系统会自动展示前15条，其余数据请点击预览数据右上方下载按钮查看全部。

### 参考文献 [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com. [2] 周志华. 机器学习[M]. 清华大学出版社, 2016.