

广义线性模型：线性回归

回归分析是一种预测性的建模技术，它研究的是因变量（目标）和自变量（特征）之间的关系，用于预测分析。学习目标：

- 理解最大似然估计及线性回归MSE损失函数的推导过程；
- 理解通过梯度下降优化方法，更新线性回归模型参数过程；
- 能在真实数据中正确应用线性回归解决问题；

知识点

- 最小二乘（最小平方方法）;见PDF
 - 最小二乘：基于经验的最小化误差平方根MSE；
 - 误差服从正态分布是最小二乘的充要条件，MSE损失函数的推导；
- 最大似然估计;见PDF
- [梯度下降求解回归方程参数](#)
代码实现可参考：[Linear Regression from scratch \(Gradient Descent\)](#)
- [\[进阶选读\]PRML 101页~106页](#)
- [\[进阶选读\]线性回归参数估计/假设检验/拟合优度](#)
- [\[进阶选读\]线性回归、岭回归、贝叶斯回归](#)

QA

通过知识学习、查找资料尝试回答：

- 如何将类别型变量引入线性回归，提出至少1种方案？
- 自变量进行标准化会对模型带来哪些影响（尝试从梯度、回归系数等角度）？尝试在项目过程中进行实验。

项目

线性回归练习题目：[Kaggle House Prices: Advanced Regression Techniques](#)，完成：

- 使用Numpy、Pandas进行数据处理，包括但不限于类别型变量处理，缺失值处理，数据标准化等；
- 使用sklearn.linear.LinearRegression进行线性回归建模，在Kaggle上提交预测结果；

项目参考：

- [sklearn 线性回归建模](#)
- [sklearn.linear_model.LinearRegression](#)
- [建模参考案例](#)

Refs

- [Kaggle House Prices: Advanced Regression Techniques](#)
- PRML

