

线性回归与正则化

PB11111111 张三

2022年2月2日

摘要：回归分析 (regression analysis) 指利用数据统计原理，对大量统计数据进行数学处理，并确定因变量与某些自变量的相关关系，建立一个相关性较好的回归方程（函数表达式），并加以外推，用于预测今后的因变量的变化的分析方法。在大数据分析中，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（预测器）之间的关系。这种技术通常用于预测分析，时间序列模型以及发现变量之间的因果关系。

关键词：线性回归；回归分析；序列建模

1 相关概念

解释与调研内容相关的、重要的但不会在章节里重点介绍的概念

损失函数 (loss function)：损失函数用来评价模型的预测值和真实值不一样的程度，损失函数越好，通常模型的性能越好。损失函数分为经验风险损失函数和结构风险损失函数。经验风险损失函数指预测结果和实际结果的差别，结构风险损失函数是指经验风险损失函数加上正则项。在一般的线性回归中，我们采用均方误差作为损失函数。而正则化则是在损失函数中加入一些惩罚项，从而解决过拟合的问题。

过拟合 (overfitting)：过拟合的问题通常发生在变量过多的时候，这种情况下训练出的模型能很好的拟合训练数据，损失函数可能非常接近于 0 或者就为 0。但是，这样的预测值虽然误差较小但方差很大，使其无法运用到新的数据样本中。

2 线性回归

(调研内容介绍)

3 最小二乘法

(调研内容介绍)

4 梯度下降法

(调研内容介绍)

5 岭回归

(调研内容介绍)

6 LASSO回归

(调研内容介绍)

7 学习心得

自己在调研过程中遇到了哪些挑战、怎样解决；有哪些收获、反思、总结？

参考文献

参考文献必不可少，格式一定要规范

推荐中文文献库：中国知网、万方；英文文献库：DBLP

[1]崔上书,杨炼,李婷.基于主成分分析及多元线性回归的空气质量预测算法研究[J].科技与创新,2022(06):97-100.DOI:10.15913/j.cnki.kjycx.2022.06.031.

[2]王璐,孙聚波.Lasso回归方法在特征变量选择中的应用[J].吉林工程技术师范学院学报,2021,37(12):109-112.

[3]Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need