

# 光伏发电预测第四次报告

张翼鹏 杨思敏

December 26, 2019

## 目录

<b>1</b>	<b>参数</b>	<b>2</b>
<b>2</b>	<b>特征</b>	<b>2</b>
<b>3</b>	<b>模型</b>	<b>2</b>

## 1 参数

这里准备推翻上次报告提出的用题目给出的误差函数（类似L1范数）代替线性回归均方误差代价函数的做法。

首先是对L1范数的优化比较难以实现，查阅资料只发现了一个相关方法，即近端梯度下降（PGD），用于解决具有L1正则化项的Lasso回归等优化问题，但与我们目前的情况也不太相符。

我认为这种做法是否能得到更好的结果这件事过于随机，受到具体数据的影响。首先我们始终使用线性回归，模型的复杂程度和拟合能力并未改变；其次我们用题目给出的误差函数，看似更加直击重点，但我们击的是训练集的重点，测试集很可能与之不符。

初步分析，如果数据分布比较均匀，那么用L1范数代价函数（MAE）或L2范数代价函数（MSE）得到的结果应该是一样的。这两个代价函数比较主要的区别就是对与数据中存在异常值的情况，MSE放大了直线上的点与异常值之间的偏差，导致拟合直线一定程度上会偏向异常值点；而MAE则受异常值影响较小。对于这种情况，我们改变代价函数，误差可能更小也可能更大。如果单纯为了避免异常值的影响，我们不如就直接删除掉异常值。

## 2 特征

将风向化成余弦值，对目前的线性回归模型确实能够减小误差。

表 1: 标准化风向后的误差比较

	普通	只标准化风向	只加入交叉项	同时
电站1	0.1293	0.1289	0.1290	0.1285
电站2	0.1517	0.1511	0.1507	0.1502
电站4	0.1556	0.1553	0.1527	0.1526
电站6	0.1261	0.1258	0.1277	0.1273
平均	0.1407	0.1403	0.1400	0.1396

## 3 模型

将训练集按小时，分割成24份，训练24个线性回归模型。预测时根据测试集数据的时间选择对应的模型进行预测。

表 2: 按小时训练结果（已标准化风向）

	普通	交叉
电站1	0.1249	0.1236
电站2	0.1716	0.1734
电站4	0.1418	0.1430
电站6	0.1218	0.1219
平均	0.1400	0.1404