

- 8, 复现效果差异较大; 尤其 NN2 与 NN5 效果差很多; 对 huber loss 的作用进行了分析很好;
- 9, 没有具体结果表格, 没有按照滑动窗口来训练和 predict;
- 10, R2 效果过好了或许用到未来数据;
- 11, 复现的结果很好, report 写的也很清楚, 文章末尾也提出新的根据 industry 来 fill missing value 的方法;
- 12, 94 个 feature 中只留下了 19 个 feature 进行建模, 或许有些过于激进; 但也从结果中发现对于 19 个 feature 建模反而 PCA+regression 效果最好, 或许这说明当 feature 较少时, 线性模型可能比复杂模型有优势;
- 13, 复现结果很好, report 写的也很清楚;
- 14, 因为计算机 memory 限制, 采取实时的构建 interaction feature 的方式, 很好的做法; 对有显著性预测作用的 feature 进行归类, 很好; 没解释 NN2 与 NN5 差别大的原因。
- 15, table1 的 ENET, RF, NN 的 R2 结果大于 1 了有明显异常;
- 16, data cleaning 中的 fill missing value 提到了不能先 fill missing value 再 split data 以防止用到未来数据; 表格结果过好了或许用到未来数据;
- 17, 绘制了按照时间的 R2 变化图, 很好, 能看到 predictive power 随着时间越接近现在而减弱; 探讨了 RF 和 NN 超过传统线性模型的原因可能在于前者能够刻画 feature 间 interaction 的能力, 很好。
- 18, 根据不同 feature 的 missing value 的占比不同而采取不同的处理方式, 很好的处理; 对于 OLS3 的效果超过所有 feature 的 OLS 有分析, 对于 ridge/lasso 的效果也有分析, 并且对于参数, R2 都按照时间序列的形式绘制出来很清晰; 对参数的 sensitivity 做了分析, 很好。
- 19, 对每种 model 都有写出自己的分析, feature importance 和相关性等图也很完整;