

思考 (2210)

Guangyao Zhao

2022-10-30

Contents

1. 画图就像堆积木，一点点『添加』最终成型。所以在 matplotlib 中，我推荐使用 Add 的写法。
2. Matplotlib 对 LaTeX 进行了满血支持，可以调用其进行高质量输出。只要不用 LaTeX 写 LaTeX，一切都是那么美好。
3. 面向对象其实非常好理解，只包括两个部分：属性和方法。如果拿人来比喻，属性就是一个人的状态，比如身高、体重、性别；方法是某些行为动作，比如吃饭、睡觉、运动。简单而言，属性是名词，方法是动词。
4. 喜欢 Polish 这个词，总有种由粗糙到精细的画面感。
5. Statistic is *math*; statistics and Python is *data analytics*; statistics, Python and model is *machine learning*; statistics, Python, model and domain knowledge is *data science*.
6. 一个大型项目要打包成一个文件夹，即 Package；文件夹里有不同的文件，即模块 (Module)；一个模块里又包含不同的对象，即类 (Class)；类里面当然又包括不同的方法 (Method)。包：小写 (sklearn)；模块：小写 (neural_network)；类：大写 (MLPClassifier)；方法：小写 (fit)。
7. 我做 PPT 第一件事便是删除所有默认生成元素，永不使用任何模版，自己做的除外。
8. 最近 3 个月最大的收获有 3 个：体重由 180 斤到 150，恢复正常体重；看了可解释性机器学习模型；Rstudio 公司的 Quarto。
9. 每个图、表格和公式都应该拥有自己的标注和编号。
10. 这段时间要做实验，实验室的 DO 计不够了，所以老师让我查查买哪家的比较好。我心想，这些东西他又怎么可能不了解呢，为什么非要让我查。于是我查了几个发了过去。他说让我给他比较下孰优孰劣，一下子我就懵了。原来他是想让我从 DO 计的原理到应用场景综合的考虑下。老师说，买实验仪器的时候是学习的大好机会。
11. Python 太通用，总觉得数据分析用 R 更高级点。
12. Why not combine the keyboard and mouse, I really tired to switch between them.

13. 化学反应里有个概念叫活化能 (Activation energy), 讲得是物质间发生反应需要克服的障碍。即便是放热反应, 也需要一定的初始助力。这就好像一个极具商业头脑的商人, 在财富积累的初始阶段, 也需要一定的启动资金。最近我时常觉得造物主在创造人类的时候也赋予每个人『活化能』这一属性: 初始阶段需要克服一定的苦难, 一旦超过了某种阈值, 便像放热反应一般一马平川。中国话好像叫『开窍』吧。
Overcome obstacles, overcome activation energy.
14. 长大后才发现中国的体量有多大, 整个欧洲也不过 8 亿人口。一个河南省 1 亿, 一个郑州市 1000 万, 单拎出来放在世界上也算是个大国了。
15. 『恐婚恐育, 不婚主义, 丁克』等等这些词俨然成为一种风尚了。我恶意猜测说者在用这些词的时候会觉得自己看透世事一般。但其实这些就像纹身, 打耳洞, 男生留长发等等行为类似, 这终究是世俗意义上所谓的『个性』啊。
16. 喜欢 Reconciliation 这个词, 万物都有自己与众不同的性格, 但结合一定需要调和。
17. 支持向量机 (Support Vector Machine, SVM) 里有一个概念叫松弛因子 (Relaxtion facor), 非常有哲学的一个想法。就是说啊, 不能限制太死, 不然根本没办法做出一个可用的模型, 要给他一个松弛空间, 允许他犯错来适应。
18. 最近幡然醒悟的一件事: 不要以为就自己讨厌现实社会的那些繁文缛节, 实则没有多少人会喜欢。但为什么有人游刃有余, 有人步履蹒跚? 前者是看透了那些不愉快的本质后仍然能保持最大限度的克制, 后者则有意无意地把自己困在那一亩三分地。在以往的 30 年中我总是拿自己情商低麻痹自己, 现在看来则是一种彻头彻尾的无能。
19. 机器学习里有一个名词叫泛化性能 (generalization ability), 讲得是依据训练集得到的模型, 在测试集上的实际性能。在此训练集可理解为生活环境为家庭和学校, 测试集则是实际社会环境。每个环境都有他独特的规则, 在一个人的成长过程中, 适当给其一定程度的惩罚项 (Penalty term), 能大大改善其在社会中的泛化性能。
20. 学校的环境会不会影响学生的学习效率? 以前我总觉得影响不大吧。最近改变了这一想法: 再粗犷的黑哥来了日本也是点头哈腰的, 环境对人的影响很大。
21. 感觉每天慌慌张张, 没有好好思考。
22. たまに頭を空っぽにして、ポケちゃんになりたい。
23. Seem to have found the sense of fitness.
24. 排版中有 3 个常用长度: point, cm, inche. 1 个 inche 是 72 point, 1 个 cm 是 28.346 poinnt。以上是绝对长度, 还有一个非常重要的概念是 pixel, 一般用 PPI (Pixel Per Inche) 表示 1 个 inche 里有多少个 pixel。此数值越高, 清晰度越高。但我还是更推荐直接导出 PDF 的做法。
25. 看着大家吐槽导师, 我真的感觉自己挺幸运的, 前几年花 5 万 RMB 给我配台式机, 前几天买实验器材又花了快 2 万。反正我导的宗旨就是不怕你花钱, 就怕你不学习。
26. 想到了一个比较不错的 idea, 一搜论文有人今年刚发的论文。一个观察: 水处理很多模型的想法都是从化学那边过来的。
27. 健身时当哪里疼了, 就代表练对地方了。人生也是如此, 当感到痛苦的时候, 那是成长。
28. SHAP 传入的数据不包括 y, 它只是基于给定的 model 和 X 计算特征重要性。这也是无论给定的模型

无论有多差, SHAP 总能给出一个特征重要性的原因。