

Sklearn 项目总结

1. 代码命名风格

- 模块名：小写字母，单词之间用_分割
- 包名：小写字母，单词之间用_分割
- 类名：单词首字母大写
- 普通变量：小写字母，单词之间用_分割
- 实例变量：以_开头，小写字母，单词之间用_分割
- 私有实例变量：以__开头（2个下划线），小写字母，单词之间用_分割
- 专有变量：开头，结尾，一般为python的自有变量，不要以这种方式命名
- 普通函数：小写字母，单词之间用_分割
- 私有函数：以__开头（2个下划线），小写字母，单词之间用_分割
- 常量名：大写字母，单词之间用_分割

错误例子

```
def Data_Access(self):  
    self.Name = 'error_example'  
    self.A = 1 + 2
```

2. 框架结构规范

- **FastResearchData**
可从多个不同的文件类型中读取数据，进行数据清洗、筛选等处理，组织生成更快更容易处理的数据格式
- **IndicatorGallexy**
存放各种指标的计算方法，传入数据并计算相应指标结果并返回
- **ModelEngine**
存放模型的构造函数、预测函数、模型评价指标计算函数等
- 一些小点
 - 可将多个类分写到多个文件里，并使用import进行调用
 - 函数内不能存在对数字或字符串等对象的直接使用的情况，可变变量尽量通过参数传递进函数，其余可变变量应放置在函数体前面，以便修改
 - 将具有相同或相似功能的函数进行合写，以便对函数功能进行修改

3. 股票的一些指标

- **EMA(Exponential Moving Average) 指数平均数指标**
 - EMA是一种趋向类指标，指数平均数指标是以指数式递减加权的移动平均。
 - 计算公式： $EMA_{today} = \alpha \times Price_{today} + (1 - \alpha) \times EMA_{yesterday}$
其中， α 为平滑指数。在计算MACD指标时，EMA计算中的N一般选取12和26天，因此 α 相应为 $2/13$ 和 $2/27$
在EMA指标中，每天价格的权重系数以指数等比形式缩小。时间越靠近当今时刻，它的权重越大，说明EMA函数对近期的价格加强了权重比，更能及时反映近期价格波动情况。所以EMA比MA更具参考价值，而EMA也不容易出现死叉和金叉
- **DIFF 差离值**
 - 股价的短期均价与长期均价的差
 - 首先分别计算出收市价SHORT日指数平滑移动平均线与LONG日指数平滑移动平均线，分别记为EMA(SHORT)与EMA(LONG)
 - 求这两条指数平滑移动平均线的差，即：DIFF=EMA(SHORT)-EMA(LONG)
- **DEA**
 - DEA是DIF的一定时间的平均值
- **MACD 指数平滑移动平均线**
 - 由快的指数移动平均线（EMA12）减去慢的指数移动平均线（EMA26）得到快线DIF，再用2×（快线DIF-DIF的9日加权移动均线DEA）得到MACD柱
 - MACD的意义和双移动平均线基本相同，即由快、慢均线的离散、聚合表征当前的多空状态和股价可能的发展变化趋势，但阅读起来更方便
 - 当MACD从负数转向正数，是买的信号。当MACD从正数转向负数，是卖的信号
 - 当MACD以大角度变化，表示快的移动平均线和慢的移动平均线的差距非常迅速的拉开，代表了一个市场大趋势的转变
- **BOLL 布林线指标**

- 布林线利用统计原理，求出股价的标准差及其信赖区间，从而确定股价的波动范围及未来走势，利用波带显示股价的安全高低价位
- 计算：
中轨线=N日的移动平均线
上轨线=中轨线+两倍的标准差
下轨线=中轨线-两倍的标准差

4. 个人框架设计

- 股票类：把股票单独作为一个类，类中存放有从.csv文件中读取出的数据，并提供计算上述其余几种指标的接口函数，以及将计算出的指标合并如总数据集中的接口函数
- 数据处理类：数据处理类中提供：选择预测对象、选择训练集属性列以及划分数据集的接口函数
- 模型选择类：模型选择类中提供：构建Nerual Network\Decision Tree\SVM\KNN\LR\NN模型的接口函数，结果预测函数，以及不同模型的比较函数

5. 个人总结

通过本次项目，熟悉了numpy以及pandas中的一些容器，如dataframe、list、array的用法，以及实现了sklearn中的一些机器学习模型。但是在项目实施过程中没有注意函数以及变量的命名规范，导致框架的可读性太差，另外，在框架的构建上，没有充分做到代码可重用性的要求。这些都是在以后的代码编写任务中需要多多注意的方面。