

太騷了，用Excel玩機器學習！

數據管道 2022-03-02 21:46

以下文章來源於機器學習算法與Python實戰，作者北海



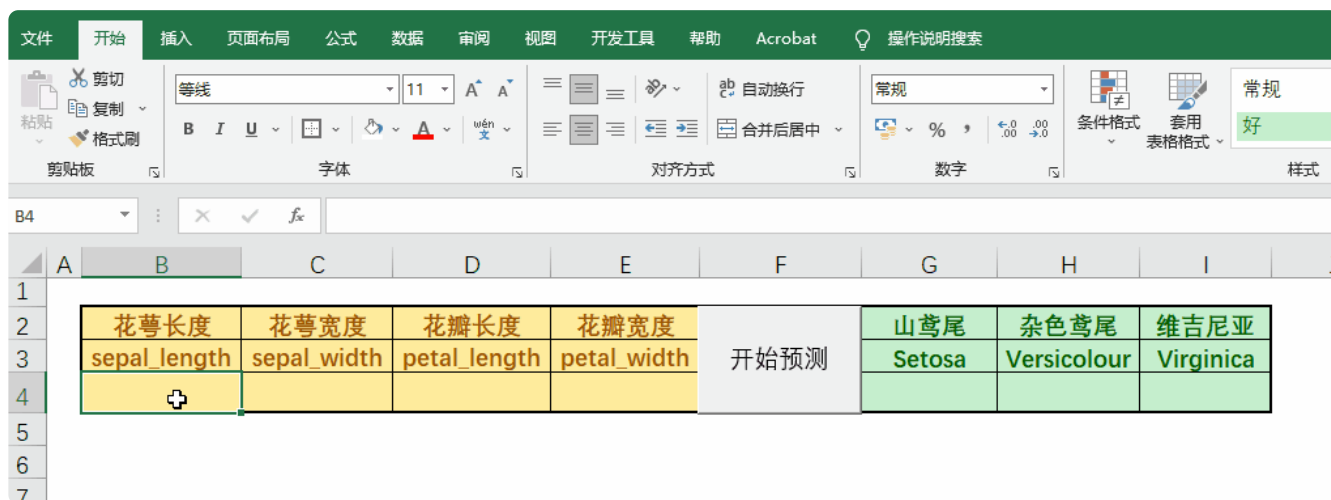
機器學習算法與Python實戰

長期跟踪關注統計學、數據挖掘、機器學習算法、深度學習、人工智能技術與行業發展...

你們好，我是寶器！

最近發現了一個好玩的Python庫，它可以將訓練好的機器學習模型轉換為Java、C、JavaScript、Go、Ruby、VBA本地代碼，可以讓連Python和機器學習一無所知的同學也能感受預測的神奇。

先看效果哈



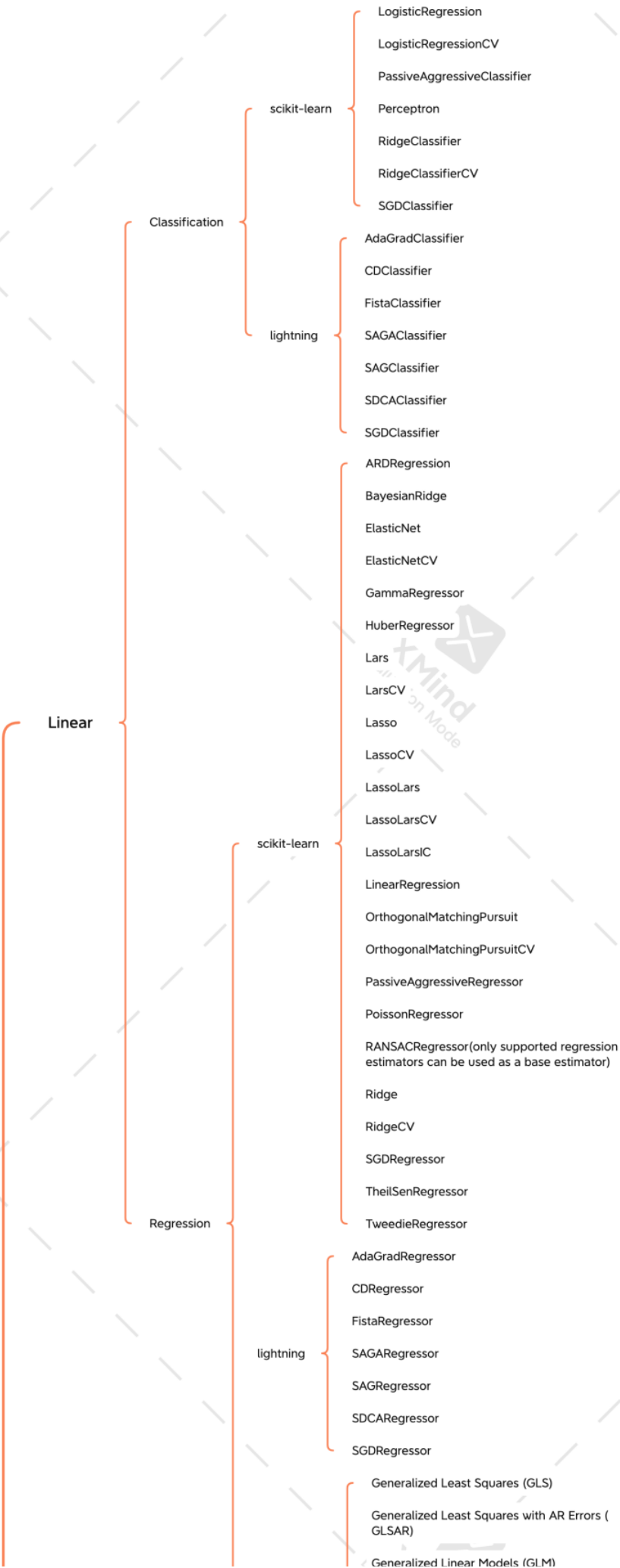
輸入特徵數據，點擊開始預測，即可輸出類別概率

這個小東西，感覺還能玩出別的花樣，但不清楚到底有什麼場景。寫這篇文章的時候，我對Excel的VBA一無所知，百度了幾個用法就開搞了。現在我也能用VBA寫一些函數和過程了，這也許是最大的收貨吧。

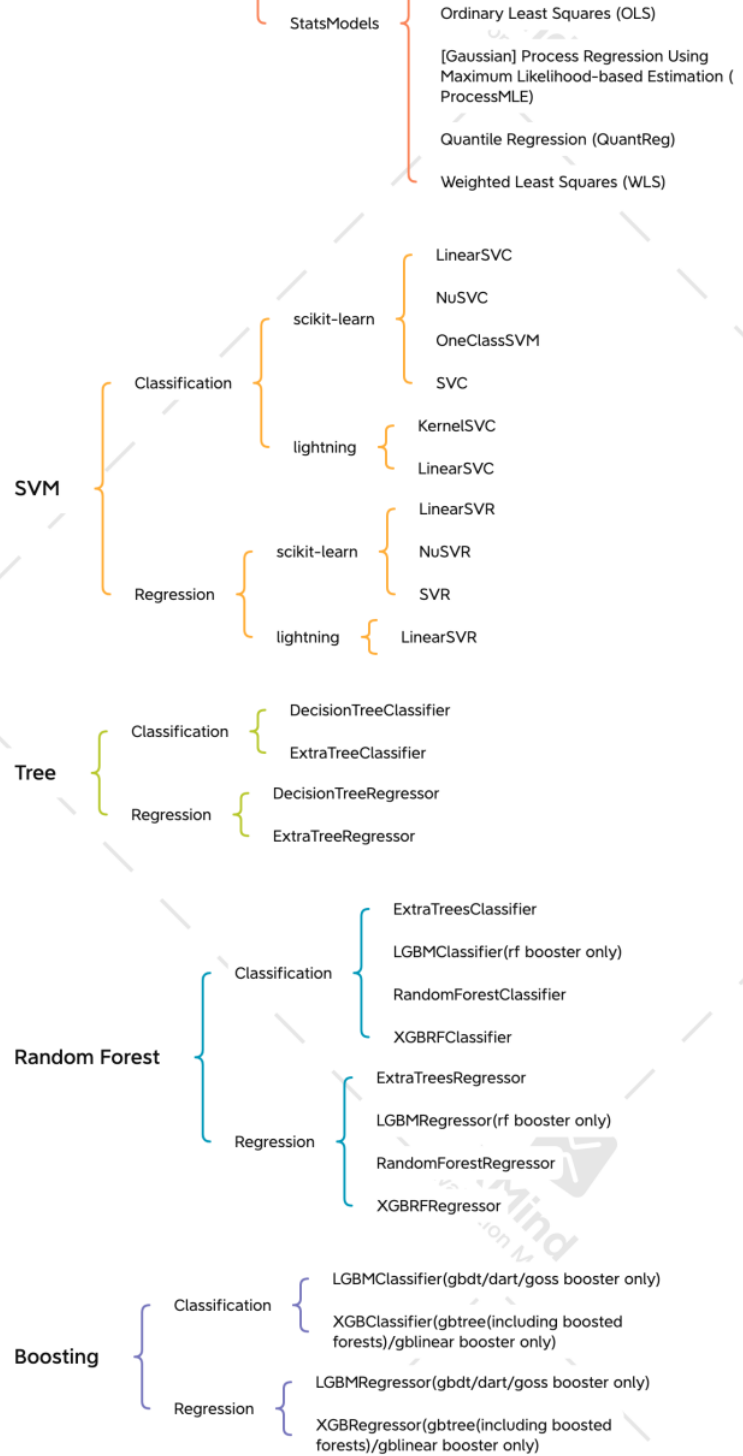
m2cgen

m2cgen (Model 2 Code Generator) 是一個輕量級代碼生成器，它可以將訓練好的機器學習模型轉換成無需依賴庫的本地代碼。

m2cgen目前支持的模型還蠻多的，常用常見的都包括了：



m2cgen Supported Models



使用方法

m2cgen的安裝非常方便，直接pip：

```
pip install m2cgen
```

使用，先用XGBClassifier訓練一個模型

```

# import packages
import pandas as pd
import numpy as np
import os
from random import sample
from sklearn import datasets
from xgboost import XGBClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
import pickle
import m2cgen as m2c

seed = 2020
test_size = 0.3
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=test_size, random_state=seed)
# fit model on training data
model = XGBClassifier()
model.fit(X_train, y_train)

```

將模型轉為VBA

```
code = m2c.export_to_visual_basic(model, function_name = 'pred')
```

VBA的英文全稱是Visual Basic for Applications，是一門標準的宏語言，通常使用來實現Excel中沒有提供的功能、編寫自定義函數、實現自動化功能等。VBA語言不能單獨運行，只能被office軟件（如：Word、Excel等）所調用。

生成的VBA長這樣：

```
print(code)
```

```

Module Model
Function pred(ByRef inputVector() As Double) As Double()
    Dim var0 As Double
    If (inputVector(2)) < (2.45) Then
        var0 = 0.41652894
    Else
        If (inputVector(3)) < (0.6) Then
            var0 = 0.257142842
        Else
            var0 = -0.209508225
        End If
    End If
End Function

```

實話實話，我對VBA還是小白，也懶得深究，於是就把代碼改成了過程，這裡就不貼了，我把改好之後的代碼放到了github。有VBA大神可以幫忙，歡迎提交PR。

<https://github.com/tjxj/excel2ml>

想省事的同學直接去copy即可

```
2475 lines (2475 sloc) | 66.5 KB

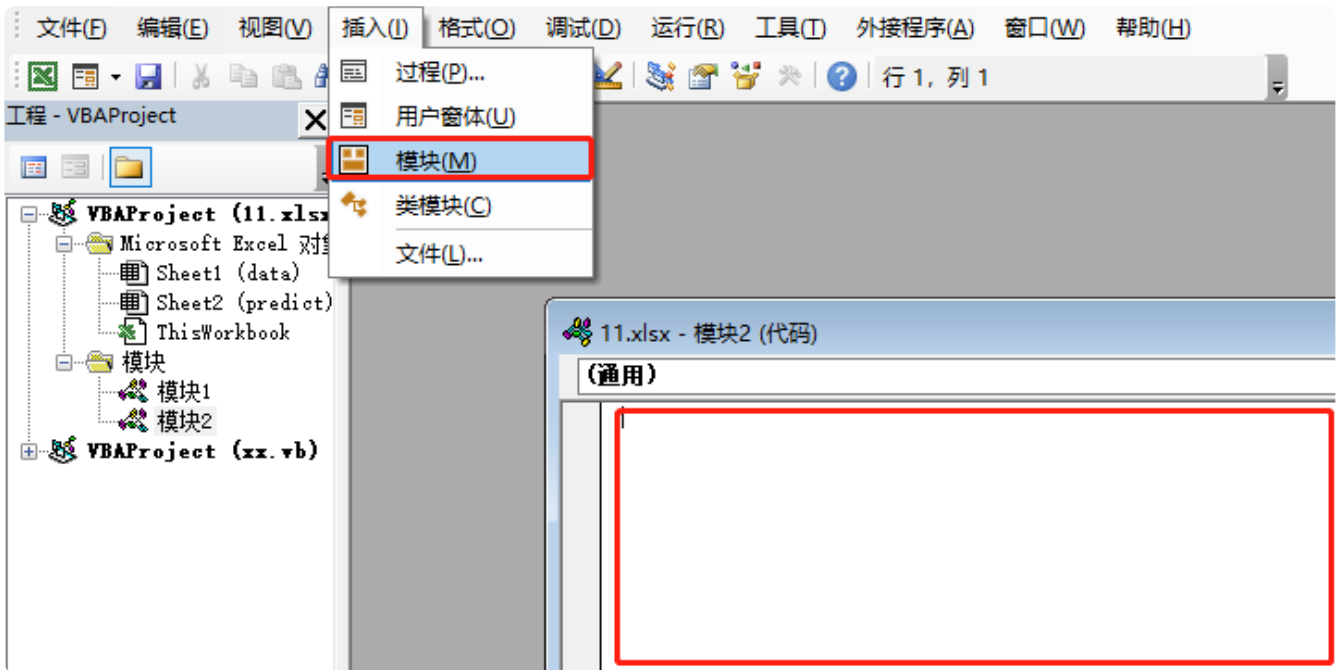
1 Sub xx()
2     sepal_length = Cells(4, 2)
3     sepal_width = Cells(4, 3)
4     petal_length = Cells(4, 4)
5     petal_width = Cells(4, 5)
6     Dim var0 As Double
7     If (petal_length) < (2.45) Then
8         var0 = 0.41652894
9     Else
10         If (petal_width) < (0.6) Then
```

Excel

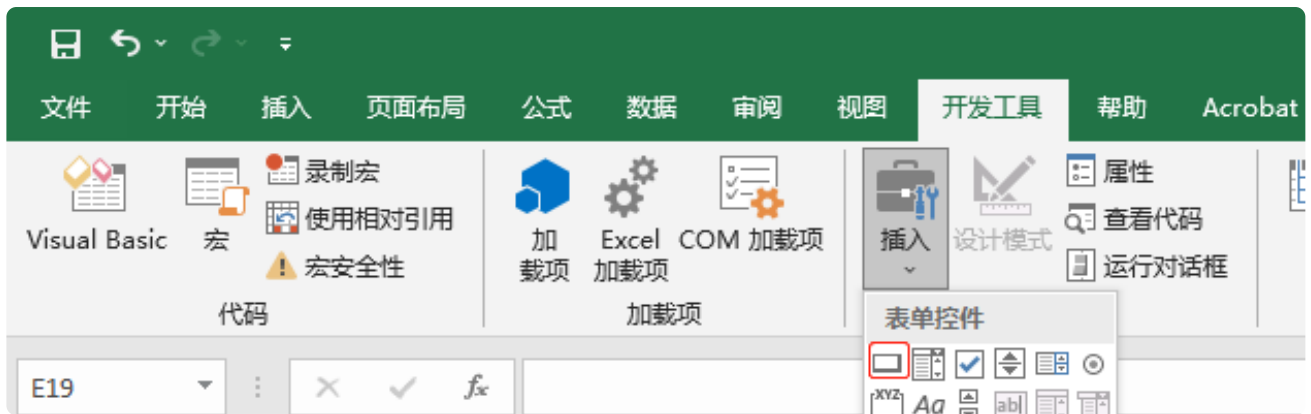
點擊excel菜單中的【開發工具】（如果沒有開發工具選項卡，請參考：excel開發工具選項卡在哪？原來需要自己添加），然後在代碼選項卡中點擊【Visual Basic】即可，也可以直接Alt + F11 喚出：



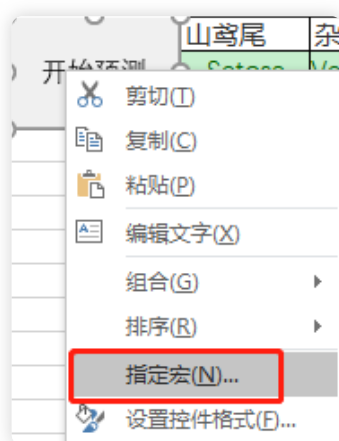
點擊插入-模塊，把改好的VBA 代碼貼進去即可



保存後退出，然後回到sheet頁，開發工具-插入-按鈕



編輯好文字後，右鍵指定宏，選中我們剛才保存的那個即可。



然後就一切OK了。

文件 开始 插入 页面布局 公式 数据 审阅 视图 开发工具 帮助 Acrobat 操作说明搜索									
剪贴板	剪切	复制	格式刷	等线 11 A A			自动换行		
	粘贴	格式刷		B I U			合并后居中	常规	条件格式 套用 表格格式
剪贴板				字体			对齐方式		
B4							数字		
	A	B	C	D	E	F	G	H	I
1									
2		花萼长度	花萼宽度	花瓣长度	花瓣宽度	开始预测	山鸢尾	杂色鸢尾	维吉尼亚
3		sepal_length	sepal_width	petal_length	petal_width		Setosa	Versicolour	Virginica
4									
5									
6									
7									

.....END.....

推薦閱讀

- 1. 我在字節做了哪些事
- 2. 寫給所有數據人。
- 3. 從留存率業務案例談0-1的數據指標體系
- 4. 數據分析師的一周
- 5. 超級菜鳥如何入門數據分析？



歡迎長按掃碼關注「數據管道」

喜歡此內容的人還喜歡

極市CV技術開學乾貨！從論文解讀到算法實踐、50+數據集資源、近3年頂會文章分類匯總

極市平台

Python實現替換照片人物背景，精細到頭髮絲（附代碼）

Jack Cui

Python數據處理入門教程！

Datawhale