



deepshare.net

深度之眼

0基础1小时完成一场AI比赛

流程：

Process



deepshare.net

深度之眼

- 1 比赛官网注册报名
- 2 安装开发环境
- 3 编辑代码，并执行
- 4 向比赛官网提交结果，并查看比赛排名



deepshare.net

深度之眼

比赛官网注册报名

官网登陆

Official website landing



1) 登陆比赛官网:

www.pkbigdata.com

点击DC竞赛

注册并进行实名认证



报名参加

sign up



2)

- 点击上方菜单栏“全部竞赛”
- 在搜索框中搜索“达观杯”
- 找到【“达观杯” 文本智能处理挑战赛】，报名参加



下载数据集至本地

Download dataset to local



3) 下载数据



[查看样本](#)

[View sample](#)



4) 用excel打开下载好的[train_set.csv], 查看第一个样本的' word_seg'的值然后关闭文件即可(只是看看数据张啥样~)

[illegible]



deepshare.net

深度之眼

2 安装开发环境anaconda

下载数据集至本地

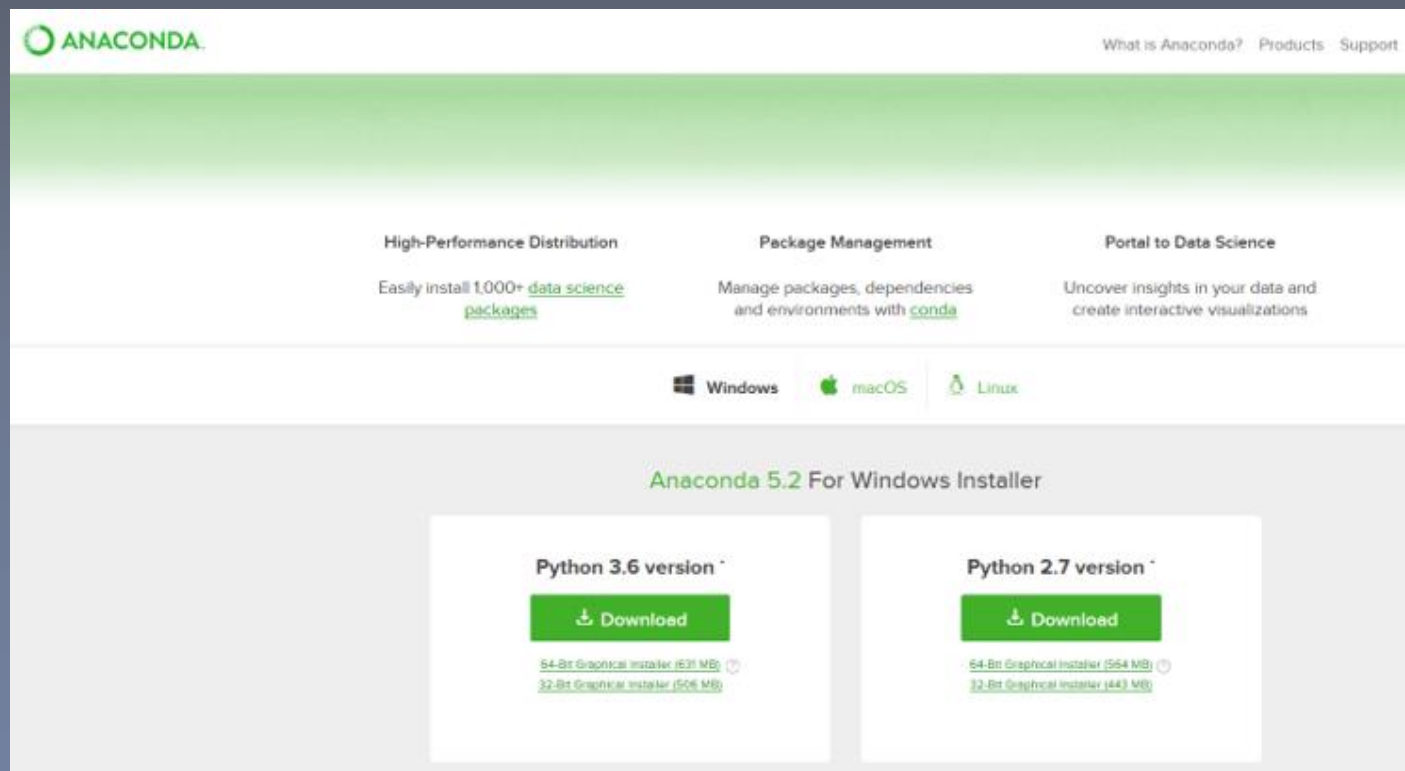
Download dataset to local



1) 登陆

www.anaconda.com/download, 下载
anaconda至本地

1) 下载完成后, 安装即可
(一直默认设置即可)





deepshare.net

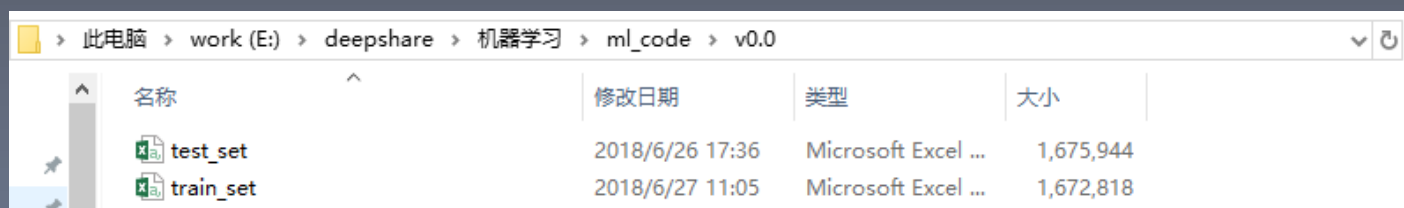
深度之眼

3 编辑代码，并执行

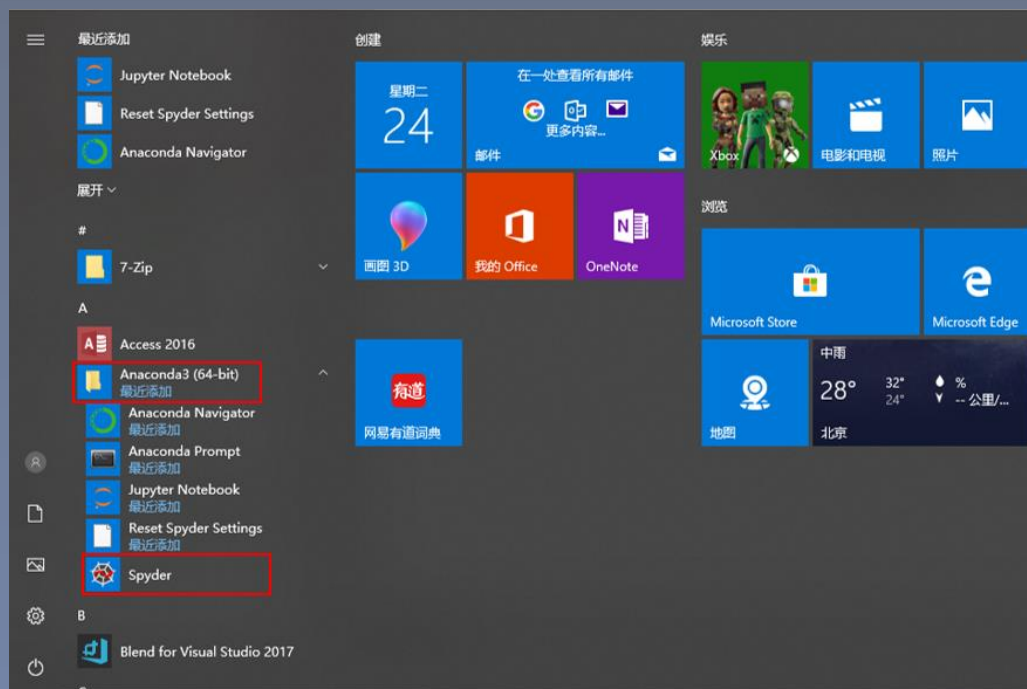
复制数据

Copy data

1) 新建 [v0.0] 文件夹,并将下载好的数据复制到该目录下



2) 在windows的[开始]菜单中, 打开anaconda中的spyder)

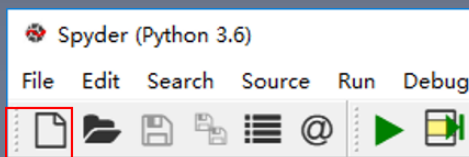


编辑代码

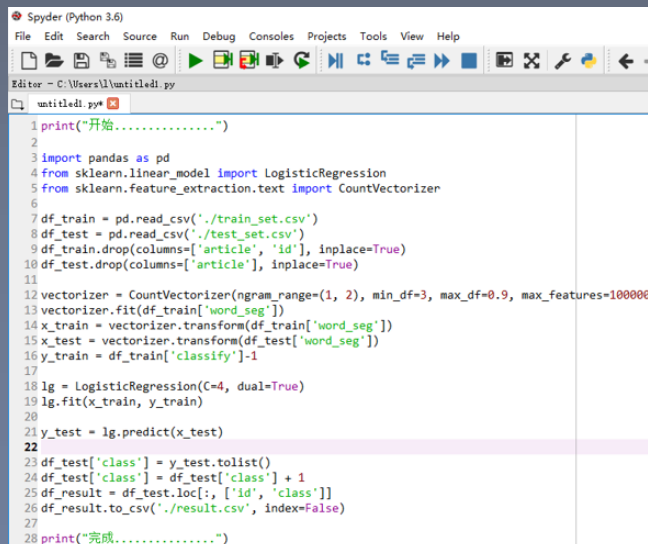
Edit code



3) 新建文件 → 编辑代码 → 保存至刚刚建好的文件夹[v0.0]，并命名为[com2018.py] → 运行代码



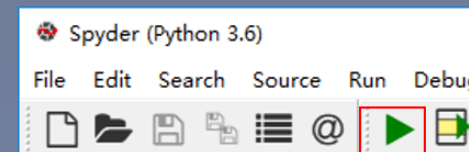
S1



S2



S3



S4

(注：清晰代码在下一张ppt，根据电脑性能的不同，大约需要执行20分钟左右)


```
1 print("开始.....")
2
3 import pandas as pd
4 from sklearn.linear_model import LogisticRegression
5 from sklearn.feature_extraction.text import CountVectorizer
6
7 df_train = pd.read_csv('./train_set.csv')
8 df_test = pd.read_csv('./test_set.csv')
9 df_train.drop(columns=['article', 'id'], inplace=True)
10 df_test.drop(columns=['article'], inplace=True)
11
12 vectorizer = CountVectorizer(ngram_range=(1, 2), min_df=3, max_df=0.9, max_features=100000)
13 vectorizer.fit(df_train['word_seg'])
14 x_train = vectorizer.transform(df_train['word_seg'])
15 x_test = vectorizer.transform(df_test['word_seg'])
16 y_train = df_train['class']-1
17
18 lg = LogisticRegression(C=4, dual=True)
19 lg.fit(x_train, y_train)
20
21 y_test = lg.predict(x_test)
22
23 df_test['class'] = y_test.tolist()
24 df_test['class'] = df_test['class'] + 1
25 df_result = df_test.loc[:, ['id', 'class']]
26 df_result.to_csv('./result.csv', index=False)
27
28 print("完成.....")
```




deepshare.net

深度之眼

4 提交结果，并查看成绩与排名

提交文件

Submit file



1) 运行完成后，查看文件目录，生成的[result]文件为要提交至比赛官网的文件

此电脑 > work (E:) > deepshare > 机器学习 > ml_code > v0.0				
名称	修改日期	类型	大小	
com2018	2018/7/25 0:09	Python File	1 KB	
result	2018/7/25 0:01	Microsoft Excel ...	945 KB	
test_set	2018/6/26 17:36	Microsoft Excel ...	1,675,944	
train_set	2018/6/27 11:05	Microsoft Excel ...	1,672,818	

查看得分与排名

View scores and rankings

2) 提交[result]至官网,
查看得分与排名





deepshare.net

深度之眼

联系我们：

电话：18001992849

邮箱：service@deepshare.net

Q Q：2677693114



公众号



客服微信

