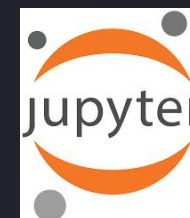
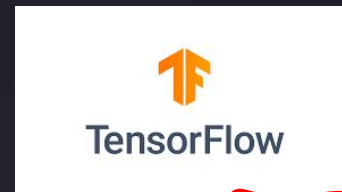




CSDN学院 IT实战派

第十章：自如应对之法-项目实战的经验2

上一章回顾



| 面试模拟：请介绍你经历过的一个实战项目？

1. 面试官不期待有标准的答案
2. 面试官期待你有实战经验
3. 面试官希望你能提供实现的细节和对细节的理解
4. 面试官期待你有解决问题的能力

面试模拟：请介绍你经历过的一个实战项目？

1. 面试官不期待有标准的答案
2. 面试官期待你有实战经验
3. 面试官希望你能提供实现的细节和对细节的理解
4. 面试官期待你有解决问题的能力

1 By Data
2 CV
3 NLP

开发工具

三个实战项目

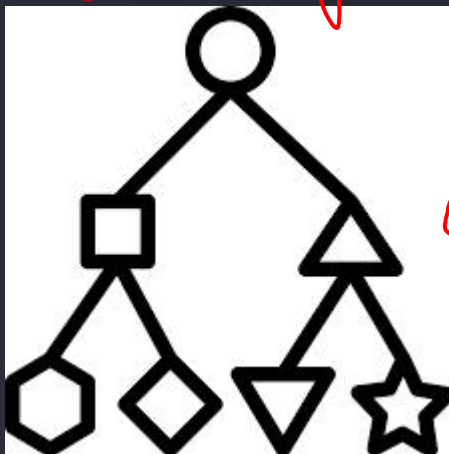
实战项目遇到的
问题和改进
的思路

深度学习项目实战-金融/大数据分析

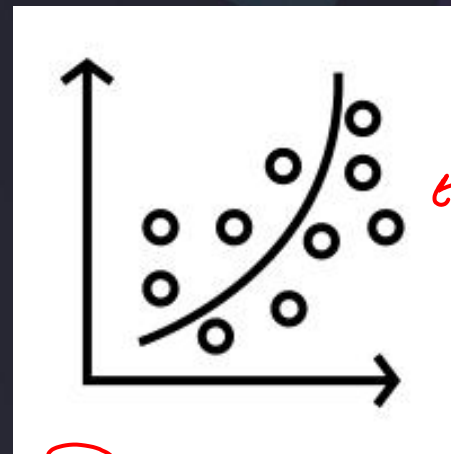
应用:

量化交易
风险评估
高价值客户定位
广告投放

Classification



Regression



输入A	输入B	输入C	目标
XXX	XX	X	TTT
XXX	XX	X	T
XXX	XX	X	TT

目标2
1/0

深度学习项目实战-量化选股案例介绍

输入A	输入B	输入C	→ 目标
XXX	XX	X	TTT
XXX	XX	X	T
XXX	XX	X	TT

| 深度学习项目实战-案例实现流程

1. 任务和模型定义
2. 数据处理与特征工程
3. 模型搭建与训练
4. 模型测试及优化

深度学习项目实战-案例实现流程

1. 任务和模型定义
2. 数据处理与特征工程
3. 模型搭建与训练
4. 模型测试及优化

MLP

深度学习项目实战-案例实现流程

1. 任务和模型定义

2. 数据处理与特征工程

3. 模型搭建与训练

4. 模型测试及优化



#数据读取

```
data = pd.read_csv('data/data_train.csv' )
```

#数据清洗, 去掉NaN

```
data=data.dropna()
```

#对数据进行归一化处理

```
nm_scaler = MinMaxScaler()  
x = nm_scaler.fit_transform(x)
```

使用主成分分析, 对数据作降维处理, 提取k个关键的特征

```
pca = PCA(n_components=k)  
x = pca.fit_transform(x)
```

#划分训练和测试数据集

```
train_x=x[0:M]  
test_x=x[M:]
```

深度学习项目实战-案例实现流程

1. 任务和模型定义
2. 数据处理与特征工程
3. 模型搭建与训练
4. 模型测试及优化



```
# 构建最基础的机器学习KNN模型, 对模型进行训练
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=3, weights='distance')
#KNN模型训练
```

```
knn.fit(train_x, train_y)
```

```
#对新的输入结果进行预测
```

```
predict_train_y = knn.predict(test_x)
```

```
#搭建ANN模型
```

```
from keras.models import Sequential
from keras.layers import Dense, Activation
from keras.optimizers import Adam
model= Sequential([Dense(10,input_dim=10), Activation('tanh'),
Dense(1), Activation('sigmoid')])
```

```
adam=Adam(lr=0.01)
```

```
model.compile(optimizer=adam, loss='mean_squared_error',
metrics=['accuracy'])
```

```
#ANN模型训练
```

```
model.fit(train_x,train_y,nb_epoch=100,batch_size=200)
```

深度学习项目实战-案例实现流程

1. 任务和模型定义
2. 数据处理与特征工程
3. 模型搭建与训练
4. 模型测试及优化



#对模型表现进行分析

```
from sklearn.metrics import accuracy_score,  
confusion_matrix
```

```
predict_test_y = model.predict(test_x)
```

```
theta=0.5
```

(0-1)

```
predict_test_y[predict_test_y>=theta]=1
```

```
predict_test_y[predict_test_y<theta]=0
```

```
acc=accuracy_score(test_y,predict_test_y)
```

```
cm = confusion_matrix(test_y,predict_test_y)
```

过了若干时间后。

面试模拟：请介绍你经历过的一个实战项目？

- 1. **任务和模型定义**：一个金融量化选股的案子，用机器学习和深度学习都进行了实验
- 2. **数据处理与特征工程**：为了提高表现，进行了特征工程，对数据进行了归一化处理和PCA降维
- 3. **模型搭建与训练**：搭建了一个三层的神经网络，用了ADAM优化器
- 4. **模型测试及优化**：相比与经典kNN算法，我的准确率提高了3个点，同时采用回归器的模型底层，通过阈值来调整风险的偏好

面试模拟：金融/大数据分析其他需要学习准备的知识

1. 数据处理：爬虫，PCA，独热编码One-hot-coding，Bootstrap
2. 监督学习模型：SVM，Random Forest，Xgboost
3. 非监督学习：kmeans，BIRCH，DBSCAN
4. 推荐算法：协同过滤collaborative filtering
5. 其他：卡尔曼滤波Kalman filter

| 本章回顾

一个金融量化选股的大数据分析案例

1. 任务和模型定义
2. 数据处理与特征工程
3. 模型搭建与训练
4. 模型测试及优化

课程相关资料





欢迎大家扫码或者添加微信好友ai_flare（学习小助手），加入学习群，老师会在群里帮大家解答学习、职业发展与求职问题（名额有限、人满即止）