



## D51-D53：Kaggle期中考



PDF 下載

全螢幕

## Sample Code & 作業內容

請至以下Kaggle競賽平台期中考連結參加考試，請按join competition加入測驗  
並將任一次提交完成畫面截圖，上傳到github並回到官網完成提交。(提交畫面截圖範例)  
<https://www.kaggle.com/c/ml100marathon-02-01/>

2 submissions for Jimmy Yang

Sort byMost recent

AllSuccessfulSelected

Submission and Description	Private Score	Public Score	Use for Final Score
<div>basic.csv</div> <div>a year ago by Jimmy Yang</div> <div>add submission details</div>	0.84309	0.84947	<input type="checkbox"/>
<div>basic.csv</div> <div>a year ago by Jimmy Yang</div> <div>add submission details</div>	0.84309	0.84947	<input type="checkbox"/>
No more submissions to show			

競賽頁面

InClass Prediction Competition

**ml100marathon-02-01**  
Midterm exam for ML 100 marathon by Cupoy

Host Overview Data Kernels Leaderboard Rules Team My Submissions

This competition hasn't been launched. Only hosts and Kaggle admins can see it.

Overview

Description

Evaluation

Timeline

Prizes

Kernels

Requirements

+ Add Page

機器學習百日馬拉松期中練習測驗

Overview

隨著移動設備的完善和普及，零售與電子商務進入了高速發展階段，這其中以 O2O（Online to Offline）消費最為熱絡。據統計，O2O 行業估值上億的創業公司至少有 10 家，也不乏百億巨頭的身影。O2O 行業每日自然流量有數億消費者，各類 APP 每天記錄了超過百億條用戶行為和位置記錄，因而成為大數據科研和商業化運營的最佳結合點之一。  
以優惠券活化老用戶或吸引新客戶消費是 O2O 的一種重要營銷方式。然而，隨機投放的優惠券對多數用戶造成無意義的干擾。對商家而言，濫發的優惠券可能降低品牌聲譽，同時難以估算營銷成本。個性化投放是提高優惠券核銷率的重要技術，它可以讓具有一定偏好的消費者得到真正的實惠，同時賦予商家更強的營銷能力。本次練習數據擷取自電商之部分數據，希望各位透過分析建模，預測用戶是否會在規定時間內使用相應優惠券。

Data

本賽題提供用戶在2016年1月1日至2016年5月31日之間真實線下消費行為，預測用戶在2016年6月領取優惠券後15天以內的使用情況。

Evaluation

本賽題目標是預測投放的優惠券是否在規定時間內核銷。針對此任務及一些相關背景知識，以該用戶使用於某日取得之優惠券核銷預測 AUC（ROC 曲線下面積）作為評價標準，即對將 User\_id + Date\_received + Coupon\_id 為一組計算核銷預測的AUC值，若某使用者於同一日取得多張相同優惠券，則任一張核銷皆為有效核銷。

使用與參考資料可於Data頁面觀看下載

Host Overview Data Kernels Leaderboard Rules Team My Submissions

This competition hasn't been launched. Only hosts and Kaggle admins can see it.

Data Description

Edit

Data (16 MB)

API kaggle competitions download -c ml100marathon-02-01 ? Download All

Data Sources

+ New Version

column\_description.csv7 x 2

sample\_submission.csv304k x 2

test\_offline.csv594k x 6

train\_offline.csv1.16m x 7

About this file

Help us describe this file

Columns

A Column

A Description

同學提交檔案前可參考Baseline Kernel作法  
<https://www.kaggle.com/ml100marathon/baseline-example>

```
In [1]:
import os
import numpy as np
import pandas as pd
from datetime import date

from sklearn.model_selection import KFold, train_test_split, StratifiedKFold, cross_val_score,
GridSearchCV
from sklearn.pipeline import Pipeline
from sklearn.linear_model import SGDClassifier, LogisticRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import log_loss, roc_auc_score, auc, roc_curve
from sklearn.preprocessing import MinMaxScaler

DATA_ROOT = "../input/"

In [2]:
dfoff = pd.read_csv(os.path.join(DATA_ROOT, 'train_offline.csv'))
dftest = pd.read_csv(os.path.join(DATA_ROOT, 'test_offline.csv'))
dftest = dftest[~dftest.Coupon_id.isna()]
dftest.reset_index(drop=True, inplace=True)
print(dfoff.shape)
print(dftest.shape)
dfoff.head(20)

(1160742, 7)
(306313, 6)
```

## 熱門問答

郭怡廷 · 2019.06.10 · 1 回答

無法進入kaggle期中考頁面

您好，點選期中考頁面連結出現以下錯誤訊息：「這個網頁無法正常運作www.kaggle.com 將您重新導向的次數過多。試試看清除 Cookie.ERR\_TOO\_MANY\_REDIRECTS」已嘗試清除Cookie 依然無法進入連結...

到 Cupoy 問答社區提問，讓教練群回答你的疑難雜症

向專家提問

如何提問