

DESAFÍO 3: Predicción de clicks



Gastón Domínguez



Elaborar un predictor de clicks

Aplicación de modelos de clasificación

Onsiderar la probabilidad de la predicción



GENERALIDADES



Dataset - Archivos

TOTAL	8,644,592
ctr_21.zip	1,523,221
ctr_20.zip	1,356,066
ctr_19.zip	1,326,180
ctr_18.zip	1,013,581
ctr_17.zip	1,093,607
ctr_16.zip	1,092,413
ctr_15.zip	1,239,524

TRAIN

ctr_test_labeled.zip 1,139,639

TEST

Dataset - Features

(#: Features Hasheados)

Label

Etiqueta click (1) o no click (0)

Nums. (6)

- ['auction_age','auction_bidfloor','auction_time',
- 'creative_height','creative_width','timezone_offset']

String (41) - #

- ['action_categorical_xx','auction_boolean_xx','auction_categorical_xx',
- 'creative_categorical_xx', 'device_id', 'device_id_type', 'gender']

Listas (3) -

• ['action_list_1','action_list_2','auction_list_0']

Bool (1)

• ['has_video']

51 Features



APLICACIÓN



Pre-Procesamiento

- Cálculo de media y desvío
 - Para el total del dataset
 - Normalización de variables numéricas

- Reducción del dataset
 - o a) Raw Data
 - b) Random Sample (p=25%) -> adoptada
 - o c) DownSample (Labels=0)



- Extractor
- Eliminación de features de listas
- Estandarización
- Reemplazo de NaNs

- Vectorizer
- Feature Hasher
- Combinaciones entre features (n_degree)

- •
- SGDClassifier (partial_fit, con batches)
 - Regresión logística
- Classifier
- Score: r² / log-loss



Grid

• [[:15], [:]] **Features N**_features • [2^20, 2^21, 2^22] degree [102] • [0.15, 0.5, 0.85] **I1_ratio** alpha • [0.0001, 0.001, 0.01, 0.1]



TRAIN - Modelos fitteados

features	n_features 🗸	degree	I1_ratio	alpha 🔻	score
				0.0001	0.999
			0.15	0.001	0.945
				0.01	0.312
				0.1	-0.044
			0.5	0.0001	0.998
		1		0.001	0.945
		1	0.5	0.01	0.311
				0.1	-0.044
			0.85	0.0001	0.999
				0.001	0.944
				0.01	0.311
[.45]	2^20			0.1	-0.044
[:15]	2~20	2	0.15	0.0001	0.961
				0.001	0.997
				0.01	0.923
				0.1	0.402
			0.5	0.0001	0.980
				0.001	0.988
				0.01	0.923
				0.1	0.352
				0.0001	0.989
			0.85	0.001	0.997
				0.01	0.923
				0.1	0.384



features 🔻	n_features 🖵	dograe	l1 ratio	alpha	coro		0.999
reatures _ n_reatu	n_reatures ,	atures degree .	0.15	0.0001	score 0.999		0.945
				0.001	0.945		
				0.01	0.312		0.312
				0.1	-0.044		0.011
				0.0001 0.001	0.998 0.945		-0.044
		1	0.5	0.001	0.945		0.000
				0.1	-0.044		0.998
		2^20		0.0001	0.999		0.045
			0.85	0.001	0.944	_	0.945
				0.01	0.311		0.311
[:15]	2^20		0.15	0.1 0.0001	-0.044 0.961	\	0.511
				0.001	0.997	- \	-0.044
				0.01	0.923		-0.044
				0.1	0.402	i	0.999
			0.5	0.0001	0.980	``	0.555
				0.001	0.988	į	0.944
				0.01 0.1	0.923 0.352		0.544
				0.0001	0.989	, i	0.311
			0.05	0.001	0.997	``	
			0.85	0.01	0.923	``	-0.044
				0.1	0.384	•	

score



[:15]

2^20

							0.999
features	n_features 🔻	degree	l1_ratio 🔻	alpha	score		0.045
			0.15	0.0001	0.999		0.945
				0.001	0.945		0.212
				0.01	0.312 -0.044		0.312
			0.5	0.0001	0.998		-0.044
				0.001	0.945		- <mark>0.044</mark>
		1		0.01	0.311		(0.998)
				0.1	-0.044		0.996
				0.0001	0.999	\	0.945
			0.85	0.001	0.944		0.943
				0.01	0.311 -0.044		0.311
[:15]	2^20	2^20	0.15	0.0001	0.961		0.511
				0.001	0.997		-0.044
				0.01	0.923		0.044
				0.1	0.402		0.999
			2 0.5	0.0001	0.980	Ň	0.555
		2		0.001	0.988	Ņ	0.944
				0.01	0.923 0.352		0.544
			0.85	0.0001	0.989	``	0.311
				0.001	0.997	``	
				0.01	0.923	\ \	-0.044
				0.1	0.384	'	3.3

0.5

0.0001

score

0.998



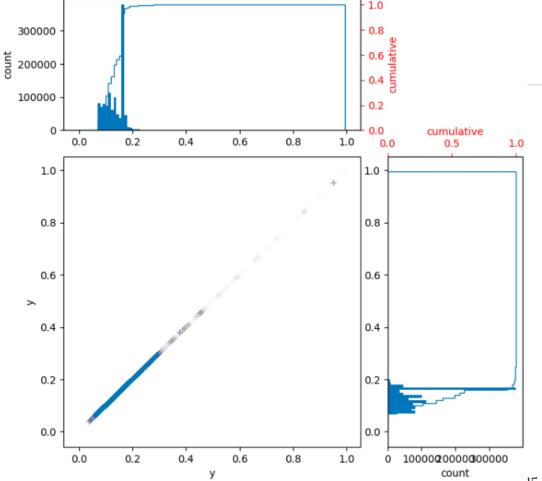


Probabilidad de click

Mean $(y_proba | Label=0) = 0.13436$

 $Min (y_proba | Label=0) = 0.00068$

 $Max (y_proba | Label=0) = 0.99358$



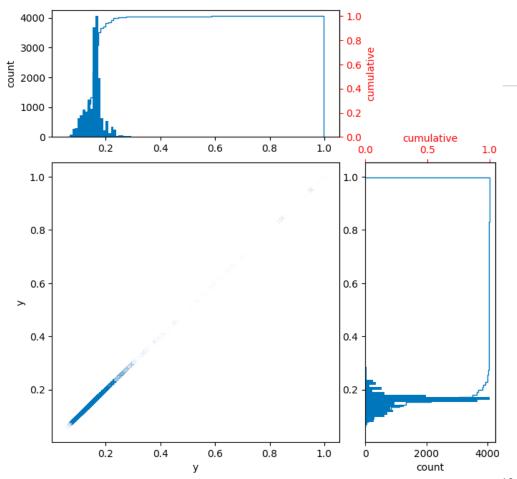


Probabilidad de click

Mean $(y_proba | Label=0) = 0.15948$

 $Min (y_proba | Label=0) = 0.05866$

Max $(y_proba | Label=0) = 0.99885$





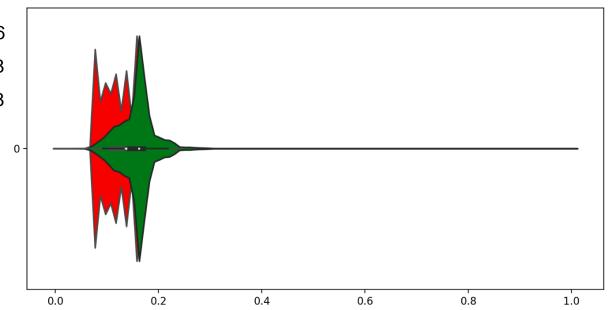
TEST - Label = [0, 1]

Label = 0

Mean = 0.13436

Min = 0.00068

Max = 0.99358



Label = 1

Mean = 0.15948

Min = 0.05866

Max = 0.99885

TEST - Report

	precision	recall	f1-score	support
0	0.98413	0.99971	0.99186	1121523
1	0.11233	0.00226	0.00444	18116
accuracy	0 54022	0 50000	0.98386	1139639
macro avg	0.54823	0.50099	0.49815	1139639
weighted avg	0.97028	0.98386	0.97617	1139639



Muchas Gracias!

Preguntas?