UNIVERSIDAD GALILEO
STATISTICAL LEARNING
Catedratico: Ing. Preng Biba
Auxiliar: Ing. Gladys Rodriguez

PROYECTO 2

Human Activity Recognition using Smartphone Data

MARVIN A. DIAZ CASTILLO
Carnet: 9516008

OBJETIVO

- Clasificar la Actividad Humana por medio de la data de un Smartphone
- El objetivo es clasificar las actividades en una de las seis actividades realizadas
  - Descansando
  - Parado
  - Sentado
  - Caminando
  - Subiendo escaleras
  - Bajando escaleras

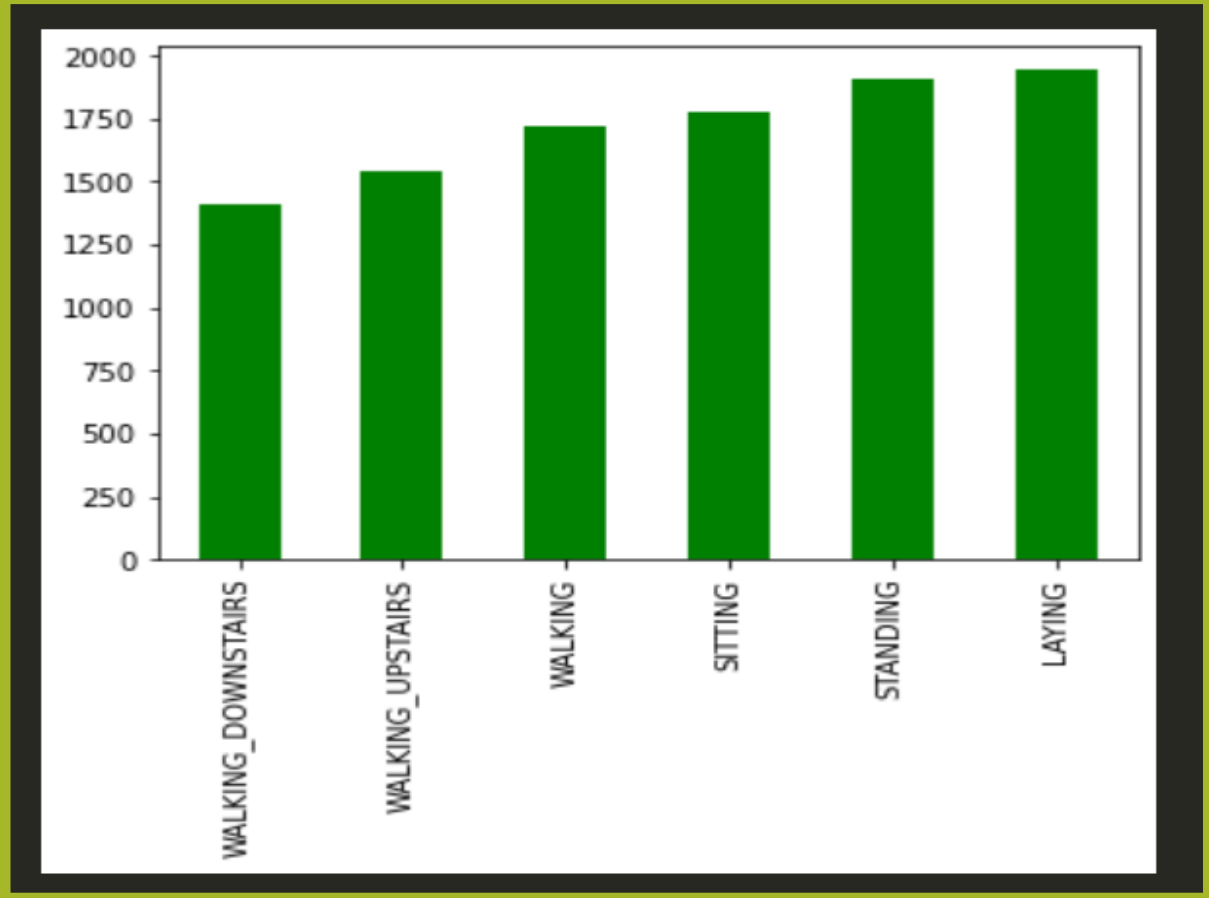
DATA

- Data Set: 10299 muestras y 562 características
- 30 personas, 6 actividades.
- Información del sensor: Velocidad angular triaxial del giroscopio.
- Variables de dominio de tiempo y frecuencia
- Su etiqueta de actividad.
- Un identificador del sujeto que llevó a cabo el experimento.

DISTRIBUCION DE LAS ACTIVIDADES

```
df.Activity.value_counts()
✓ 0.1s

LAYING                1944
STANDING              1906
SITTING               1777
WALKING               1722
WALKING_UPSTAIRS      1544
WALKING_DOWNSTAIRS    1406
Name: Activity, dtype: int64
```



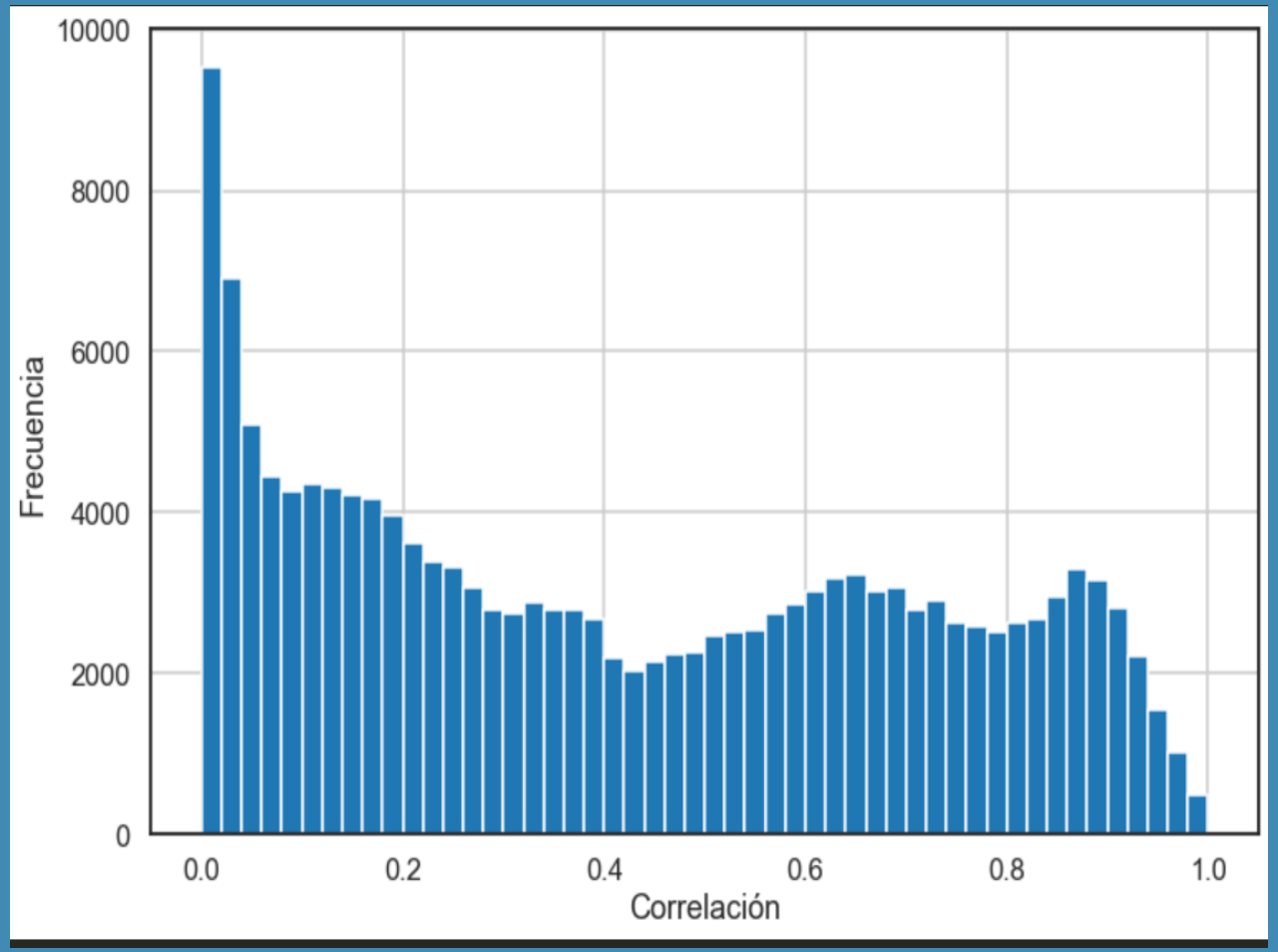
ANALISIS DE CORRELACION

```
# VAlores con mayor correlación
corr_values.sort_values('correlation', ascending=False).query('abs_correlation>0.8')
```

✓ 0.2s

	feature1	feature2	correlation	abs_correlation
156894	fBodyBodyGyroJerkMag-mean()	fBodyBodyGyroJerkMag-sma()	1.000000	1.000000
93902	tBodyAccMag-sma()	tGravityAccMag-sma()	1.000000	1.000000
101139	tBodyAccJerkMag-mean()	tBodyAccJerkMag-sma()	1.000000	1.000000
96706	tGravityAccMag-mean()	tGravityAccMag-sma()	1.000000	1.000000
94257	tBodyAccMag-energy()	tGravityAccMag-energy()	1.000000	1.000000
...	...	...	...	...
22657	tGravityAcc-mean()-Y	angle(Y,gravityMean)	-0.993425	0.993425
39225	tGravityAcc-arCoeff()-Z,3	tGravityAcc-arCoeff()-Z,4	-0.994267	0.994267
38739	tGravityAcc-arCoeff()-Z,2	tGravityAcc-arCoeff()-Z,3	-0.994628	0.994628
23176	tGravityAcc-mean()-Z	angle(Z,gravityMean)	-0.994764	0.994764
38252	tGravityAcc-arCoeff()-Z,1	tGravityAcc-arCoeff()-Z,2	-0.995195	0.995195

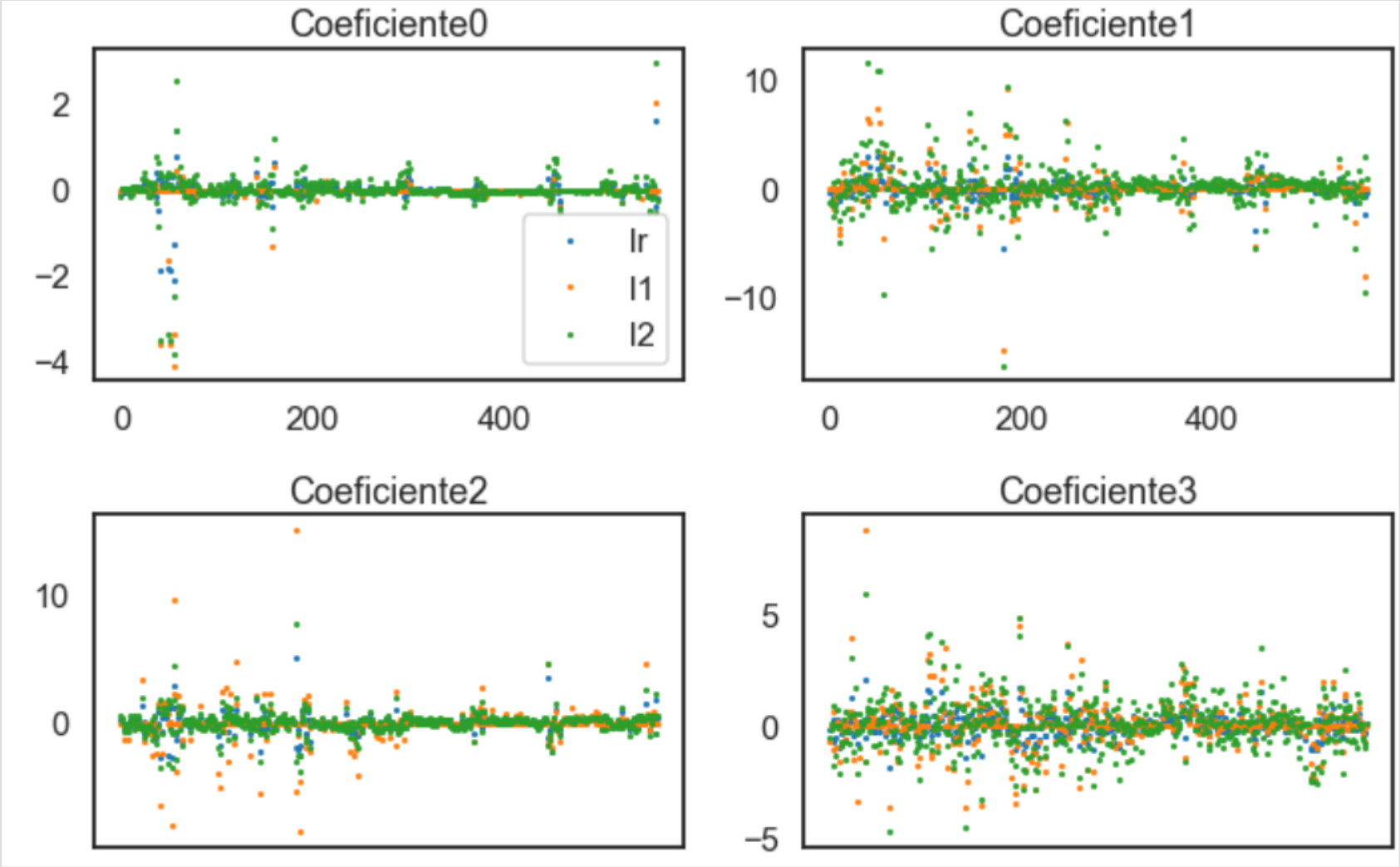
22815 rows × 4 columns

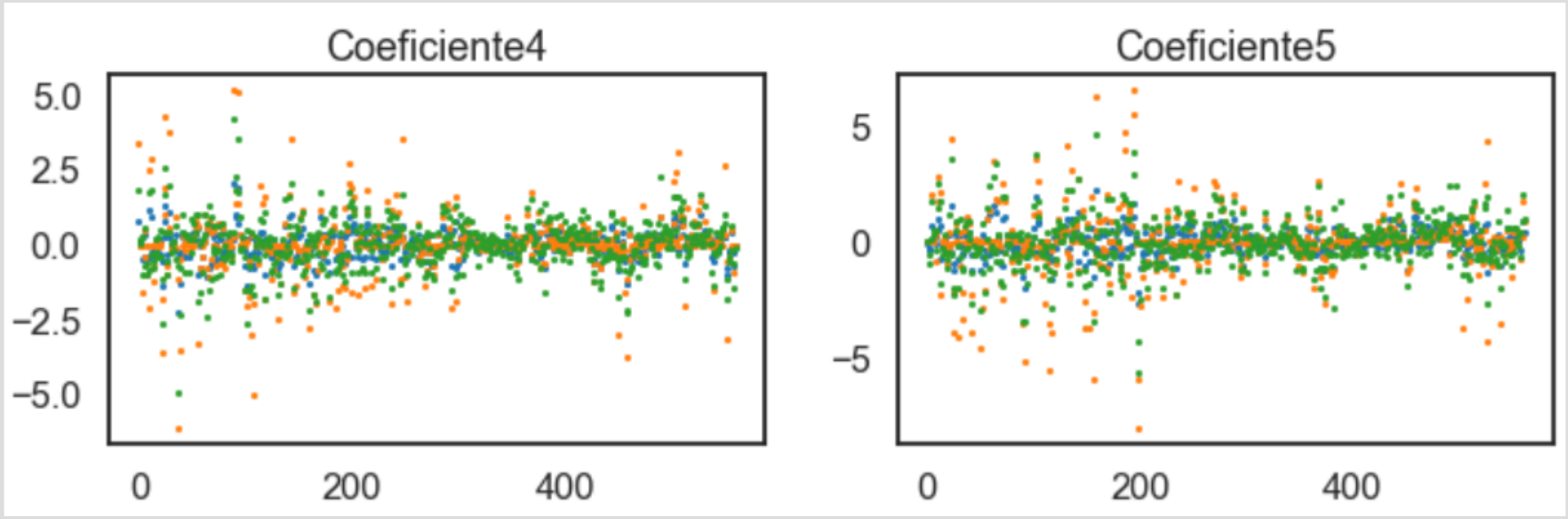


	lr					
	0	1	2	3	4	5
467	0.002993	0.217266	0.308574	0.054700	0.098334	0.237068
549	0.134871	-1.306534	1.478447	0.255047	0.282069	-0.354369
219	0.024451	-0.113410	-0.022481	-0.621019	-0.027569	-0.490645
81	-0.025714	-0.133342	0.071428	-0.080978	0.044292	-0.202523
245	-0.020129	0.084135	-0.158133	-0.348243	0.212745	-0.018933
55	-1.260545	2.287108	1.111516	0.094611	-0.975173	0.948330
78	0.125709	-0.057108	-0.000535	0.123442	-0.340251	0.224885
79	-0.141387	0.019343	0.006189	-0.193702	0.390725	-0.304583
147	-0.129983	-1.146467	1.122597	-0.054408	0.361250	-0.441206
493	0.001605	0.220263	0.274081	0.133273	0.253586	0.267638

l2					
0	1	2	3	4	5
-0.015393	0.527925	0.427618	-0.054149	0.288340	0.234280
0.318195	-5.517847	2.565188	0.439207	0.529715	-1.069425
0.042324	-0.080316	-0.019160	-1.437535	-0.193968	-1.004554
-0.050799	-1.813820	0.205014	-0.407655	-0.018429	-0.546569
-0.039755	0.332173	-0.290340	-0.753963	0.545235	-0.199476
-2.473193	2.878359	1.742269	0.212317	-1.856742	1.460618
0.350567	-0.199276	-0.031623	0.316007	-0.403523	0.334993
0.118011	-0.020612	-0.035518	-0.341141	0.670793	-0.674167
-0.290298	2.038250	1.319681	-0.282643	0.656109	-1.233966
-0.018180	0.448042	0.365899	0.383623	0.648394	0.180536

I1					
0	1	2	3	4	5
0.000000	0.444868	0.288052	0.072614	0.000000	0.227805
0.000000	-3.085037	4.635855	0.521366	0.447211	-1.542963
0.000000	0.000000	0.000000	-0.684324	0.000000	-0.687159
0.000000	0.000000	0.000000	0.000000	0.053969	-0.630920
0.000000	0.000000	0.000000	-0.266923	0.351680	0.000000
-3.349345	3.259702	2.307449	0.000000	-3.291337	1.004223
0.000000	-0.144005	0.000000	0.153800	-0.646908	0.523292
-0.099450	0.000000	-0.065511	-0.234506	0.524329	-0.805003
0.000000	0.000000	0.000000	0.000000	0.000000	-2.611601
0.000000	0.315724	0.489997	0.121645	0.398609	0.123822


Coeficiente0
2
0
−2
−4
0
200
400
lr
l1
l2
Coeficiente1
10
0
−10
0
200
400
Coeficiente2
10
0
Coeficiente3
5
0
−5


Coeficiente4
5.0
2.5
0.0
−2.5
−5.0
0
200
400
Coeficiente5
5
0
−5
0
200
400

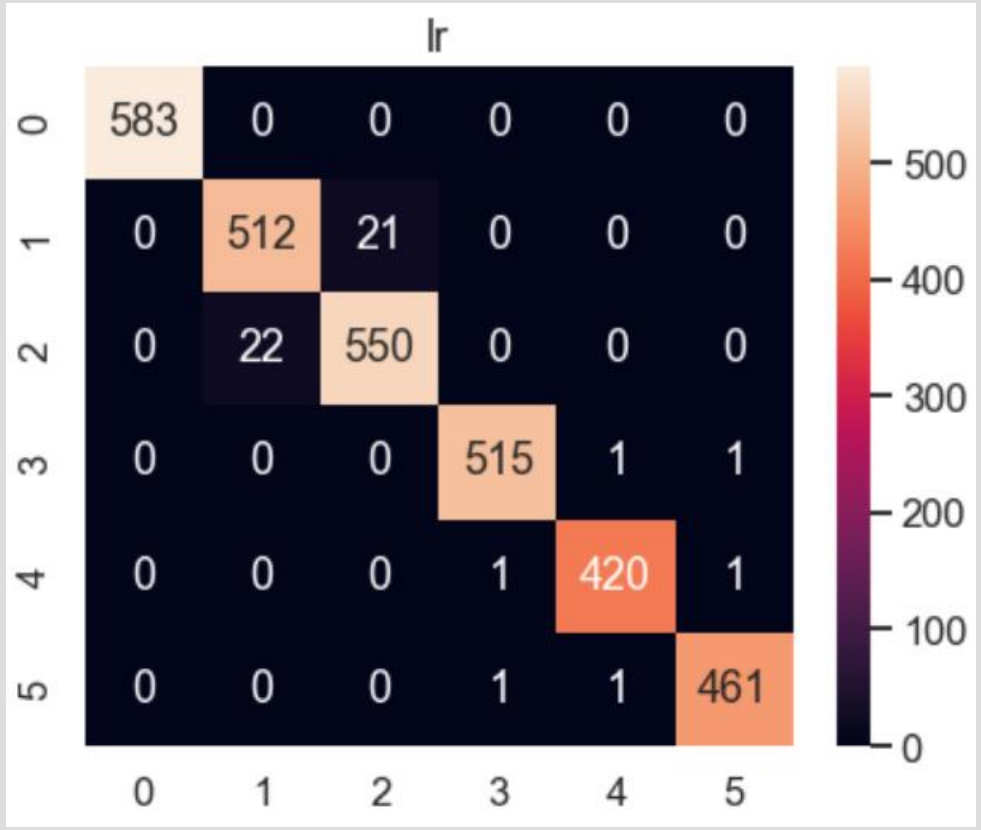
	lr	l1	l2
0	3	3	3
1	5	5	5
2	3	3	3
3	1	1	1
4	0	0	0

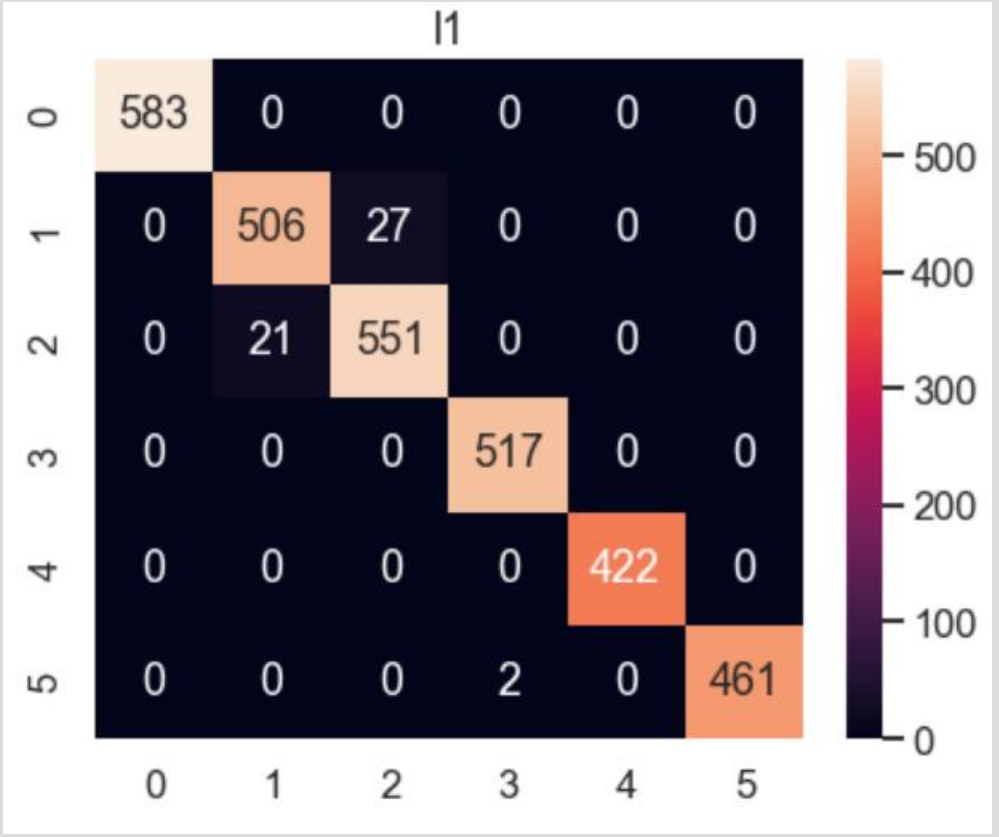
	lr	l1	l2
0	0.998939	0.998910	0.999757
1	0.988165	0.999462	0.999477
2	0.987592	0.995559	0.999671
3	0.981381	0.999137	0.994338
4	0.998277	0.999918	0.999997

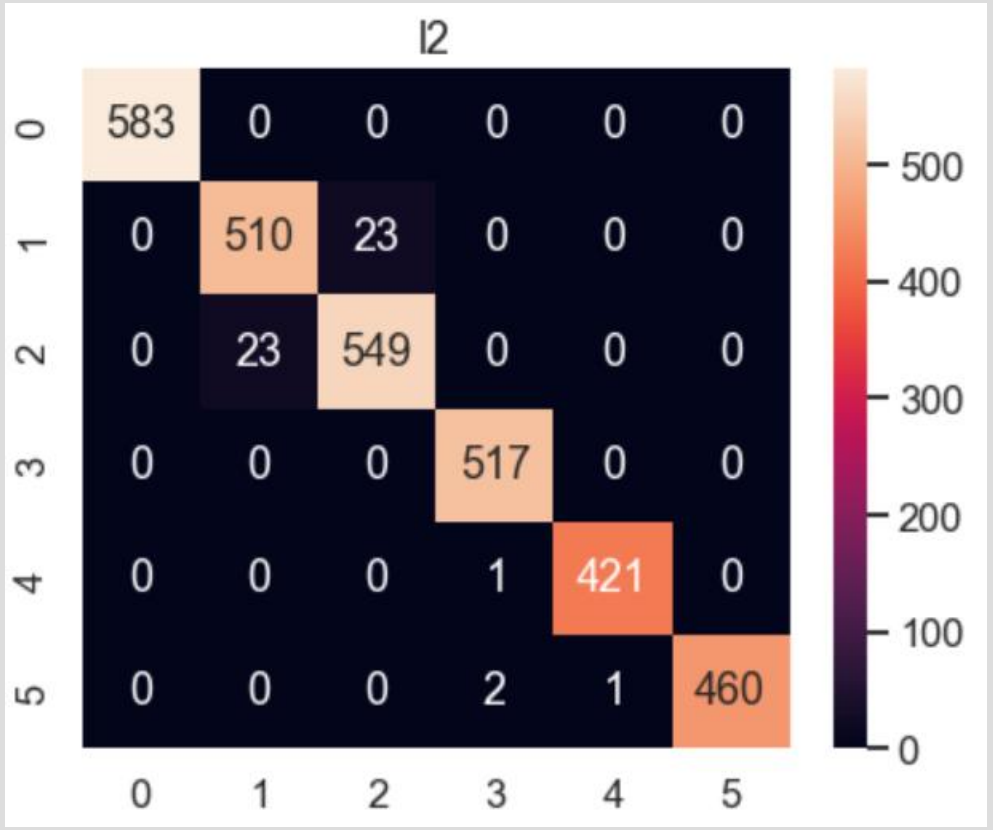
`metrics`

✓ 0.1s

	lr	l1	l2
precision	0.984144	0.983835	0.983824
recall	0.984142	0.983819	0.983819
fscore	0.984143	0.983815	0.983819
accuracy	0.984142	0.983819	0.983819
auc	0.990384	0.990144	0.990165


lr
0 583 0 0 0 0 0
1 0 512 21 0 0 0
2 0 22 550 0 0 0
3 0 0 0 515 1 1
4 0 0 0 1 420 1
5 0 0 0 1 1 461
0 1 2 3 4 5
500
400
300
200
100
0

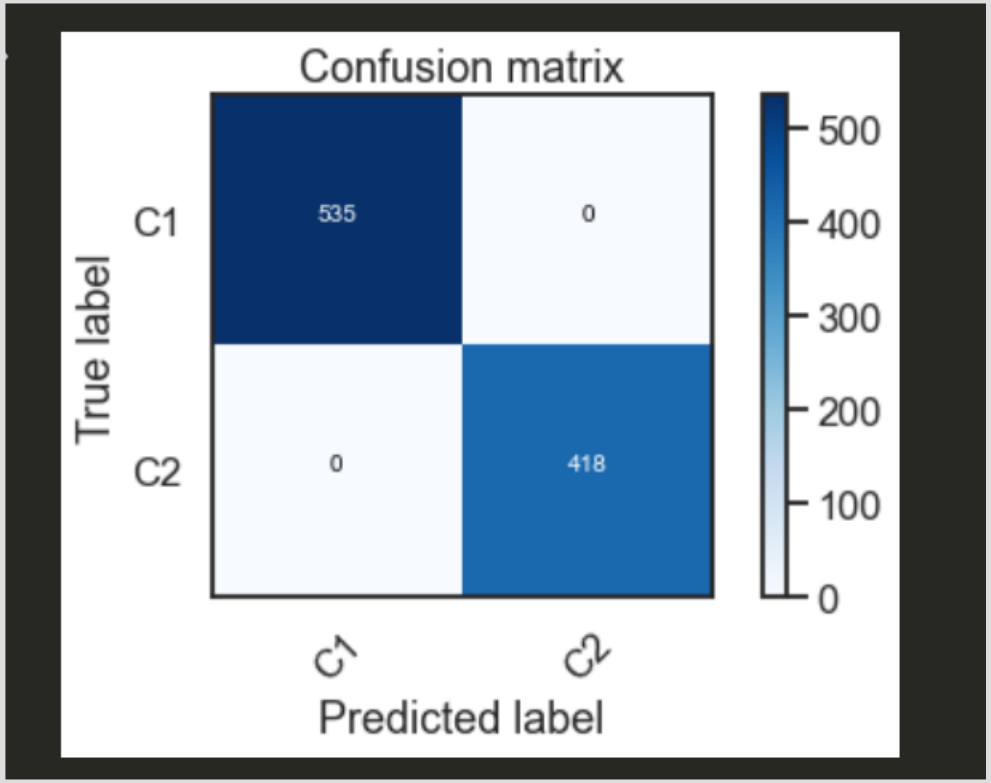

l1
0 583 0 0 0 0 0
1 0 506 27 0 0 0
2 0 21 551 0 0 0
3 0 0 0 517 0 0
4 0 0 0 0 422 0
5 0 0 0 2 0 461
0 1 2 3 4 5
500
400
300
200
100
0


l2
583 0 0 0 0 0
0 510 23 0 0 0
0 23 549 0 0 0
0 0 0 517 0 0
0 0 0 1 421 0
0 0 0 2 1 460
0 1 2 3 4 5
0 1 2 3 4 5
500
400
300
200
100
0

	precision	recall	f1-score	support
0	1.00	1.00	1.00	583
1	0.93	0.94	0.94	533
2	0.95	0.94	0.94	572
3	1.00	1.00	1.00	517
4	0.99	0.99	0.99	422
5	0.99	0.99	0.99	463
accuracy			0.97	3090
macro avg	0.98	0.98	0.98	3090
weighted avg	0.97	0.97	0.97	3090

Confusion matrix, without normalization

```
[[535   0]
 [  0 418]]
```

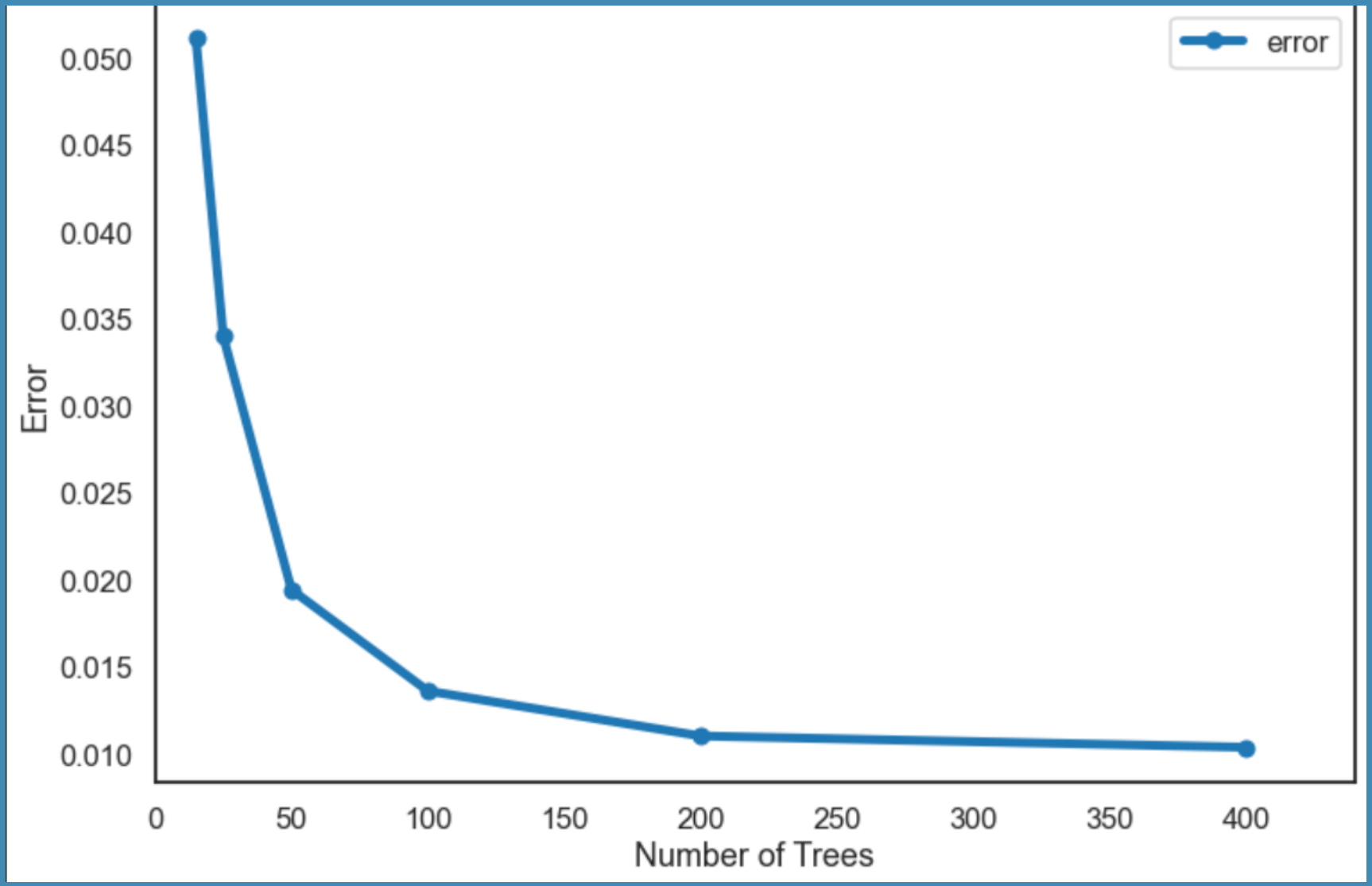


Error Data Frame
Index número de trees
Cuando incrementa el numero de trees disminuye el error

```
Fitting model with 15 trees
Fitting model with 25 trees
Fitting model with 50 trees
Fitting model with 100 trees
Fitting model with 200 trees
Fitting model with 400 trees
```

n_trees	error
15.0	0.051133
25.0	0.033981
50.0	0.019417
100.0	0.013592
200.0	0.011003
400.0	0.010356

GRAFICANDO EL NUMERO DE TREES



Gradient Boost

El mejor estimador fue con el numero de estimadores de 100 con una learning rate de 0.01

```
GV_ABC.best_estimator_
```

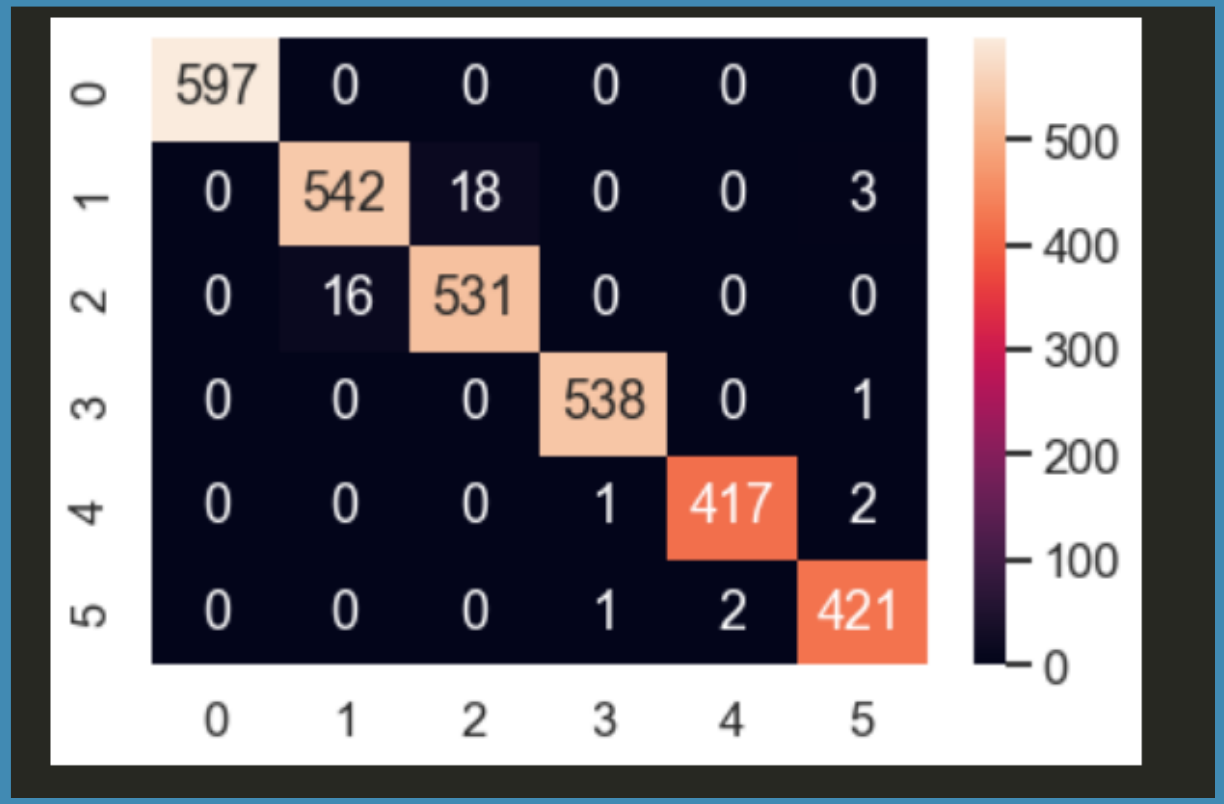
```
AdaBoostClassifier(base_estimator=DecisionTreeClassifier(max_depth=1),
                   learning_rate=0.01, n_estimators=100)
```

Utilizando el mejor estimador para predecir el X-test

	precision	recall	f1-score	support
0	1.00	1.00	1.00	597
1	0.96	0.97	0.97	558
2	0.97	0.97	0.97	549
3	1.00	1.00	1.00	540
4	0.99	1.00	0.99	419
5	0.99	0.99	0.99	427
accuracy			0.99	3090
macro avg	0.99	0.99	0.99	3090
weighted avg	0.99	0.99	0.99	3090

Se obtuvieron valores altos cercanos a 1

MATRIZ DE CONFUSION



La actividad 1 se confunde con la actividad 2, que son las actividades sitting and standing

Se clasificaron en diagonal lo que quiere decir que se clasificaron correctamente

Ada Boost Clasificador

```
GV_ABC.best_estimator_
```

✓ 0.1s

```
AdaBoostClassifier(base_estimator=DecisionTreeClassifier(max_depth=1),
                   learning_rate=0.01, n_estimators=100)
```

Class 2 fue muy bajo

	precision	recall	f1-score	support
0	1.00	1.00	1.00	598
1	0.00	1.00	0.00	1
2	1.00	0.49	0.66	1108
3	0.92	0.84	0.88	589
4	0.73	0.95	0.82	320
5	0.89	0.80	0.84	474
accuracy			0.75	3090
macro avg	0.76	0.85	0.70	3090
weighted avg	0.94	0.75	0.81	3090

Class 1 y 2 fueron clasificados mal al igual que la clase 3