



Ingeniería en Robótica y Sistemas Digitales

TE3002B.501

Implementación de Robótica Inteligente

Actividad

Ajuste de modelos de clasificación

Alumnos:

Ariadna Minerva Solís Naranjo A01639943

Barbara Nicole Vidal Sandoval A01635233

Luis Paulo Flores Arzate A01275194

Manuel Eduardo Ochoa Obezo A00227718

Tec de Monterrey, Campus Guadalajara

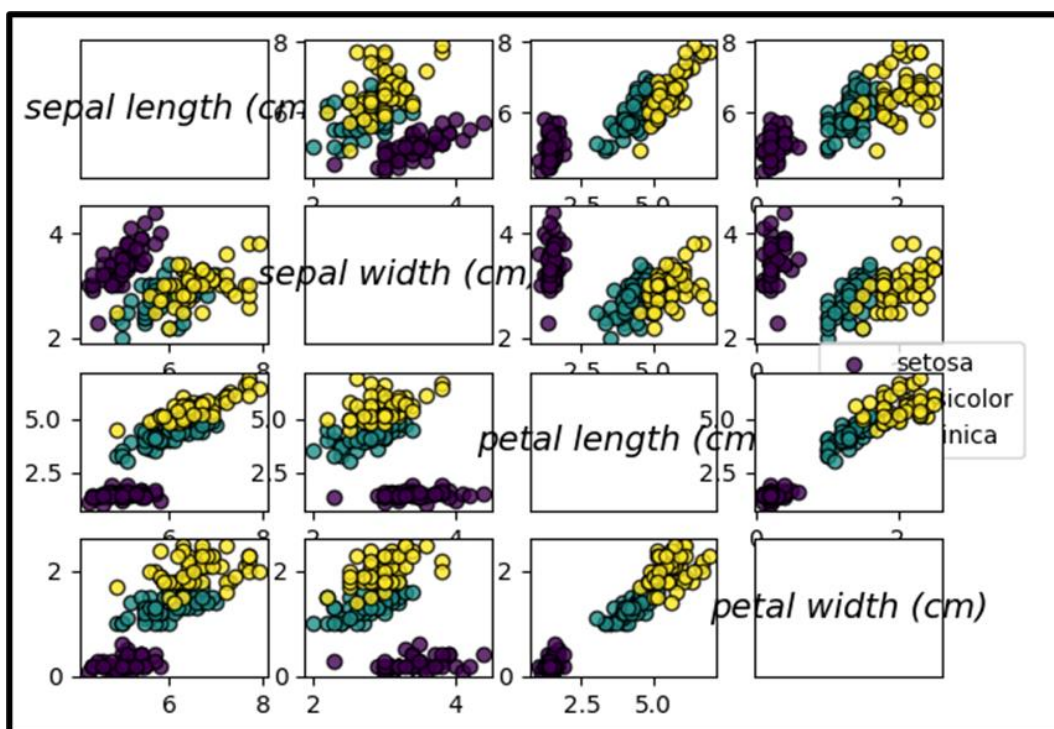
30 de mayo de 2023

Introducción

Los modelos de clasificación son una herramienta muy valiosa y utilizada dentro del campo de aprendizaje profundo, ya que puede aprender automáticamente las características más relevantes de los datos que se están considerando, y a su vez establecer relaciones que permitan clasificar los mismos datos. Esto nos puede ayudar a predecir el tipo o clase de nuevas observaciones, sobre todo cuando se trata de conjuntos de datos muy grandes. A su vez, es importante poder ser capaces de evaluar dichos modelos de clasificación, estimando su capacidad para predecir de forma correcta la clase para las nuevas observaciones. Es por esto, que durante la siguiente actividad se estarán probando y ajustando distintos modelos de clasificación, con el objetivo de evaluar su rendimiento utilizando, entre otras cosas, validación cruzada.

Ejercicio 1

- 1.1



¿Qué representan las variables incluidas en la base de datos? ¿Consideras que las variables predictoras tienen información suficiente para determinar la clase de cada uno de los tipos de datos?

Las variables que utiliza la base de datos de Iris representan características que son útiles para diferenciar tipos de especie de plantas: setosa, virginica y versicolor. En cuanto a la información que posee cada variable, consideramos que contienen suficiente información para detectar y clasificar cada dato dentro de dicha base de datos.

- 1.2

```
[6] # Train SVM classifier with all the available observations
clf = SVC(kernel = 'linear')
clf.fit(x, y)

# Predict one new sample
print("New evaluations", targets[clf.predict([
    [4.4, 2.9, 1.4, 0.2],
    [0.4, 0.1, 0.1, 0.4],
    [7.7, 3.0, 6.1, 2.2],
    [5.3, 2.9, 3.2, 1.3],
    [4.8, 3.8, 2.8, 0.8],
    [6.5, 4.1, 6.1, 2.3],
    [5.3, 3, 2, 0.2],
    [5, 4, 3, 2],
    [6.2, 2.5, 1.9, 2.4],
    [5.4, 3, 4.5, 1.5]
])])

New evaluations ['setosa' 'setosa' 'virginica' 'versicolor' 'setosa' 'virginica' 'setosa'
'versicolor' 'versicolor' 'versicolor']
```

En esta sección del código declaramos 10 nuevas evaluaciones con valores aleatorios para que pueda reconocer y clasificar cada parte de datos con la clase a la que más se asemejan. Observando, nuestro clasificador predice de manera correcta las clases a las que pertenece cada sección de datos.

- 1.3

	precision	recall	f1-score	support
0	1.00	1.00	1.00	5
1	1.00	1.00	1.00	5
2	1.00	1.00	1.00	5
accuracy			1.00	15
macro avg	1.00	1.00	1.00	15
weighted avg	1.00	1.00	1.00	15
Acc: 1.9466666666666668				
Precision:	[2.	2.	1.86666667]	
Recall:	[2.	1.84 2.]	

En esta parte del código, realizamos una matriz de confusión con los valores de las predicciones de los targets y su testeo para que de esa manera se desplieguen los datos de la imagen anterior.

Link a collab: <https://colab.research.google.com/drive/1XmwKMwjtFhUoC3Eb1JDB5IJ-P1nKQgNd#scrollTo=EkEanlQQeUi0>

Ejercicio 2

Modelo KNN con k=3:					Modelo de árbol de decisión:				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.81	0.85	0.83	59	0	0.95	0.95	0.95	59
1	0.70	0.69	0.70	71	1	0.93	0.87	0.90	71
2	0.52	0.50	0.51	48	2	0.88	0.96	0.92	48
accuracy			0.69	178	accuracy			0.92	178
macro avg	0.68	0.68	0.68	178	macro avg	0.92	0.93	0.92	178
weighted avg	0.69	0.69	0.69	178	weighted avg	0.92	0.92	0.92	178
Exactitud del modelo: 0.6906349206349207					Exactitud del modelo: 0.9214285714285714				

Modelo clasificador con proceso Gaussiano:				
	precision	recall	f1-score	support
0	0.89	0.29	0.44	59
1	0.77	0.38	0.51	71
2	0.34	0.88	0.49	48
accuracy			0.48	178
macro avg	0.67	0.51	0.48	178
weighted avg	0.70	0.48	0.48	178
Exactitud del modelo: 0.4838095238095238				

- 2.3:

Indicar con qué clasificador se obtuvieron mejores resultados

Los dos modelos que alcanzaron más del 90% de exactitud fueron el SVM (con kernel lineal) y el de árbol de decisión. Sin embargo, entre ambos modelos, el que tiene la mayor exactitud es el SVM con kernel lineal. Por lo tanto, de entre los cinco modelos de clasificación implementados, el que obtuvo los mejores resultados fue el SVM con kernel lineal, con un 95%.

Link a collab:

<https://colab.research.google.com/drive/1pDq7qfhXMus67oGLVPcaAYX3cWK6igcy#scrollTo=eJbOszXZ1DH>

Ejercicio 3

- 3.1

Modelo SVM con kernel lineal:				
----- K = 5 -----				
	precision	recall	f1-score	support
1	0.78	0.75	0.77	264
2	0.76	0.78	0.77	264
accuracy			0.77	528
macro avg	0.77	0.77	0.77	528
weighted avg	0.77	0.77	0.77	528
Exactitud del modelo: 0.768984725965858				

- 3.2

Modelo SVM con kernel lineal:

----- test_size = 10% -----

----- K = 5 -----

	precision	recall	f1-score	support
1	0.77	0.77	0.77	264
2	0.77	0.77	0.77	264
accuracy			0.77	528
macro avg	0.77	0.77	0.77	528
weighted avg	0.77	0.77	0.77	528

Exactitud del modelo: 0.7690745732255166

• 3.3

Modelo SVM con kernel lineal:

----- test_size = 15% -----

	precision	recall	f1-score	support
1	0.78	0.78	0.78	264
2	0.78	0.78	0.78	264
accuracy			0.78	528
macro avg	0.78	0.78	0.78	528
weighted avg	0.78	0.78	0.78	528

Exactitud del modelo: 0.7840251572327044

Modelo SVM con kernel lineal:

----- test_size = 20% -----

	precision	recall	f1-score	support
1	0.78	0.76	0.77	264
2	0.77	0.78	0.78	264
accuracy			0.77	528
macro avg	0.77	0.77	0.77	528
weighted avg	0.77	0.77	0.77	528

Exactitud del modelo: 0.7726864330637917

Modelo SVM con kernel lineal:

----- test_size = 25% -----

	precision	recall	f1-score	support
1	0.75	0.78	0.76	264
2	0.77	0.75	0.76	264
accuracy			0.76	528
macro avg	0.76	0.76	0.76	528
weighted avg	0.76	0.76	0.76	528

Exactitud del modelo: 0.7613836477987421

Modelo SVM con kernel lineal:

----- test_size = 30% -----

	precision	recall	f1-score	support
1	0.79	0.78	0.79	264
2	0.78	0.80	0.79	264
accuracy			0.79	528
macro avg	0.79	0.79	0.79	528
weighted avg	0.79	0.79	0.79	528

Exactitud del modelo: 0.7879424977538185

Modelo SVM con kernel lineal:

----- test_size = 35% -----

	precision	recall	f1-score	support
1	0.79	0.80	0.79	264
2	0.79	0.79	0.79	264
accuracy			0.79	528
macro avg	0.79	0.79	0.79	528
weighted avg	0.79	0.79	0.79	528

Exactitud del modelo: 0.7936028751123091

Modelo SVM con kernel lineal:

----- test_size = 40% -----

	precision	recall	f1-score	support
1	0.79	0.77	0.78	264
2	0.77	0.79	0.78	264
accuracy			0.78	528
macro avg	0.78	0.78	0.78	528
weighted avg	0.78	0.78	0.78	528

Exactitud del modelo: 0.7784366576819408

Modelo SVM con kernel lineal:

----- test_size = 45% -----

	precision	recall	f1-score	support
1	0.81	0.79	0.80	264
2	0.80	0.81	0.80	264
accuracy			0.80	528

Modelo SVM con kernel lineal:

----- test_size = 50% -----

	precision	recall	f1-score	support
1	0.79	0.77	0.78	264
2	0.77	0.79	0.78	264
accuracy			0.78	528

Modelo SVM con kernel lineal:

----- test_size = 55% -----

	precision	recall	f1-score	support
1	0.79	0.77	0.78	264
2	0.77	0.79	0.78	264
accuracy			0.78	528
macro avg	0.78	0.78	0.78	528
weighted avg	0.78	0.78	0.78	528

Modelo SVM con kernel lineal:

----- test_size = 60% -----

	precision	recall	f1-score	support
1	0.79	0.79	0.79	264
2	0.79	0.79	0.79	264
accuracy			0.79	528
macro avg	0.79	0.79	0.79	528
weighted avg	0.79	0.79	0.79	528

Modelo SVM con kernel lineal:

	precision	recall	f1-score	support
1	0.79	0.77	0.78	264
2	0.78	0.80	0.79	264
accuracy			0.78	528
macro avg	0.78	0.78	0.78	528
weighted avg	0.78	0.78	0.78	528

Exactitud del modelo: 0.7841509433962264

Modelo SVM con kernel lineal:

	precision	recall	f1-score	support
1	0.78	0.78	0.78	264
2	0.78	0.77	0.78	264
accuracy			0.78	528
macro avg	0.78	0.78	0.78	528
weighted avg	0.78	0.78	0.78	528

Exactitud del modelo: 0.7784546271338725

Modelo SVM con kernel lineal:

	precision	recall	f1-score	support
1	0.77	0.78	0.78	264
2	0.78	0.77	0.78	264
accuracy			0.78	528
macro avg	0.78	0.78	0.78	528
weighted avg	0.78	0.78	0.78	528

Exactitud del modelo: 0.776513926325247

Modelo SVM con kernel lineal:

	precision	recall	f1-score	support
1	0.76	0.79	0.77	264
2	0.78	0.75	0.76	264
accuracy			0.77	528
macro avg	0.77	0.77	0.77	528
weighted avg	0.77	0.77	0.77	528

Exactitud del modelo: 0.7690206648697215

Modelo SVM con kernel lineal:

	precision	recall	f1-score	support
1	0.77	0.78	0.78	264
2	0.78	0.77	0.77	264
accuracy			0.77	528
macro avg	0.77	0.77	0.77	528
weighted avg	0.77	0.77	0.77	528

Exactitud del modelo: 0.77455525606469

Modelo SVM con kernel lineal:

	precision	recall	f1-score	support
1	0.79	0.80	0.80	264
2	0.80	0.79	0.79	264
accuracy			0.80	528
macro avg	0.80	0.80	0.80	528
weighted avg	0.80	0.80	0.80	528

Exactitud del modelo: 0.7953459119496855

Modelo SVM con kernel lineal:

	precision	recall	f1-score	support
1	0.79	0.78	0.79	264
2	0.79	0.79	0.79	264
accuracy			0.79	528
macro avg	0.79	0.79	0.79	528
weighted avg	0.79	0.79	0.79	528

Exactitud del modelo: 0.7879245283018868

- 3.4

De acuerdo a los resultados del apartado anterior, ¿Cuántas observaciones creen que son necesarias para entrenar el modelo para el tipo de datos probados?

De acuerdo al análisis realizado, observamos que la exactitud del modelo es mayor con el 45% de los datos u observaciones, obteniendo un 0.80 de exactitud del modelo. Por lo que, se puede concluir que incluyendo más datos no necesariamente mejora la exactitud del modelo, y en este caso hacer uso del 45% de los datos probados sería el número que nos daría mejores resultados.

Link a collab:

<https://colab.research.google.com/drive/14cMXarsf60fQp9aggJgTSyo8GhvEOrPE#scrollTo=PgCcks3p1DHv>

Ejercicio 4

• 4.1

De acuerdo a los resultados del apartado anterior, ¿Qué valor de k es el óptimo para los datos probados?

De acuerdo a los resultados obtenidos nuestro valor de exactitud decrece con forme la k es mayor a partir del número k = 17. Sin embargo, tomando la mejor k de nuestro análisis, la más óptima sería K = 6 con un valor de exactitud del modelo de 0.9527289546716003

K = 6

-----Clasificador K-NN con k = 6 -----				
-----Evaluación usando k-folds cross validation con 5 pliegues-----				
	precision	recall	f1-score	support
1	0.91	0.95	0.93	78
2	0.95	0.99	0.97	78
3	1.00	0.92	0.96	78
accuracy			0.95	234
macro avg	0.95	0.95	0.95	234
weighted avg	0.95	0.95	0.95	234
Exactitud del modelo: 0.9527289546716003				

• 4.2

De acuerdo a los resultados del apartado anterior, ¿qué valor de este parámetro es el óptimo?

Observando y analizando, el valor de C, el modelo más óptimo y más preciso es el que tiene $C = 0.000009$ ya que como observamos dicho valor, 0.47391304347826085, es el más alto.

Modelo SVM con kernel lineal:				
----- C = 9e-06 -----				
	precision	recall	f1-score	support
1	0.40	0.51	0.45	78
2	0.53	0.44	0.48	78
3	0.54	0.47	0.50	78
accuracy			0.47	234
macro avg	0.49	0.47	0.48	234
weighted avg	0.49	0.47	0.48	234
Exactitud del modelo: 0.47391304347826085				

- 4.3

De acuerdo a los resultados del apartado anterior, ¿Qué valor seleccionarías para este parámetro?

Observando y analizando, seleccionaríamos el valor del kernel gamma = 0.0001 ya que es el valor más alto que toma como exactitud del modelo, 0.6706753006475485.

Modelo SVM con kernel lineal:				
----- gamma = 1e-05 -----				
	precision	recall	f1-score	support
1	0.64	0.69	0.66	78
2	0.67	0.59	0.63	78
3	0.71	0.73	0.72	78
accuracy			0.67	234
macro avg	0.67	0.67	0.67	234
weighted avg	0.67	0.67	0.67	234
Exactitud del modelo: 0.6706753006475485				

Link a collab:

https://colab.research.google.com/drive/13-UbSW51c06k_XRl8BoGBuTScp6V-8Lv?usp=sharing

Conclusión

Gracias a esta actividad pudimos llevar a cabo el análisis de diferentes bases de datos para entrenar diferentes modelos de clasificación y en base a eso poder obtener un modelo optimizado ajustando los parámetros de la función de SVC como el kernel, el factor de regularización C, gamma y algunos parámetros fuera de éste como la variable k, y el porcentaje del test size. También, tuvimos la oportunidad de realizar el análisis con funciones y archivos .txt que portaban las bases de datos dentro de nuestro código, de esa manera aprender las diferentes funciones y sus atributos para la obtención de los targets y features, así como la información base.