

Ejercicio 2

Los ejemplos de iris setosa están bien separados pero los de iris virginica y versicolor no, lo cual puede hacer que el modelo sea poco preciso distinguiendo entre esas dos clases.

Preparación del dataset:

Es necesario llevar los distintos atributos al mismo rango para que valores de distintas magnitudes no afecten en la función de distancia euclidiana.

	Min	Max	Mean
sepal length:	4.3	7.9	5.84
sepal width:	2.0	4.4	3.05
petal length:	1.0	6.9	3.76
petal width:	0.1	2.5	1.20

Rango Normalizado:

	Min	Max	Mean
sepal length:			
sepal width:			
petal length:			
petal width:			

Parametros de k-NN:

- k: cantidad de ejemplos más cercanos que votan sobre la clase del elemento que se quiere predecir
- Weighted vote: si se aplica, hace que el voto de los ejemplos más cercanos cuente mas
- Measure types: define como calcular la distancia entre los ejemplos.
- Funciones de distancia:
 - Distancia euclidean
 - Distancia Manhattan
 - Cosine
 - Correlation
 - Jaccard
 - Chebychev

Performance del modelo k-NN con k=5, weighted vote y distancia euclidea:

PerformanceVector:

accuracy: 98.67%

ConfusionMatrix:

True:	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa:	25	0	0
Iris-versicolor:	0	25	1
Iris-virginica:	0	0	24