



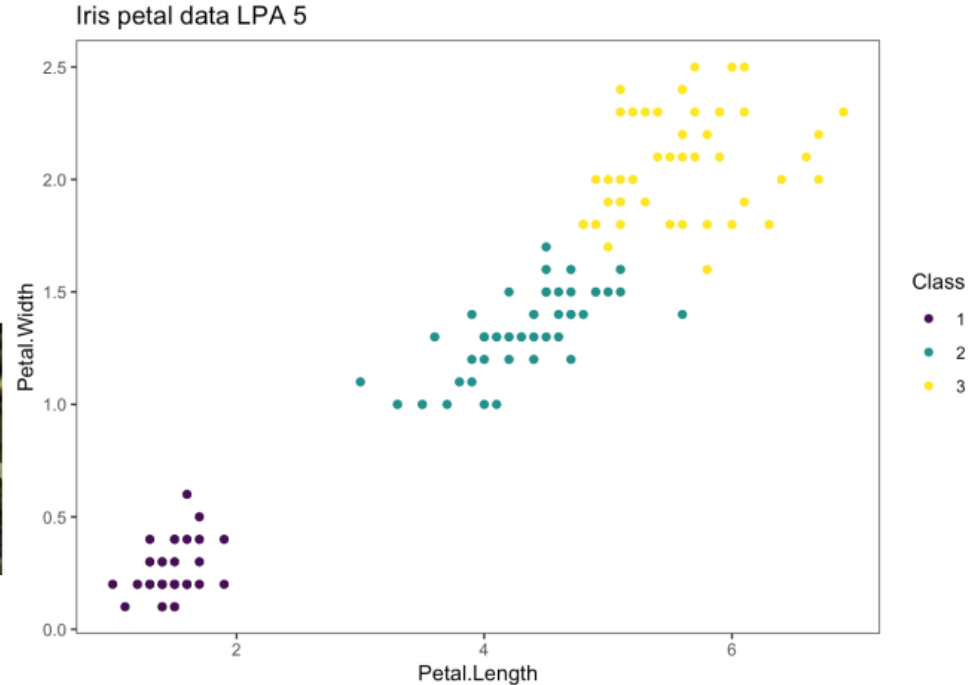
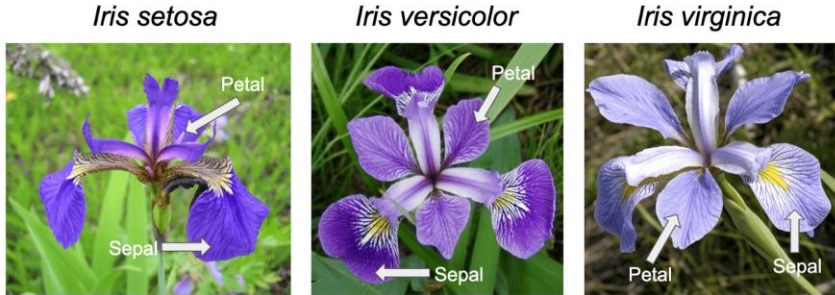
K-nearest neighbors

K-vecinos más cercanos

Ing. Juan M. Rodríguez

K-nearest neighbors

Tenemos un conjunto de datos clasificado, por ejemplo Iris:



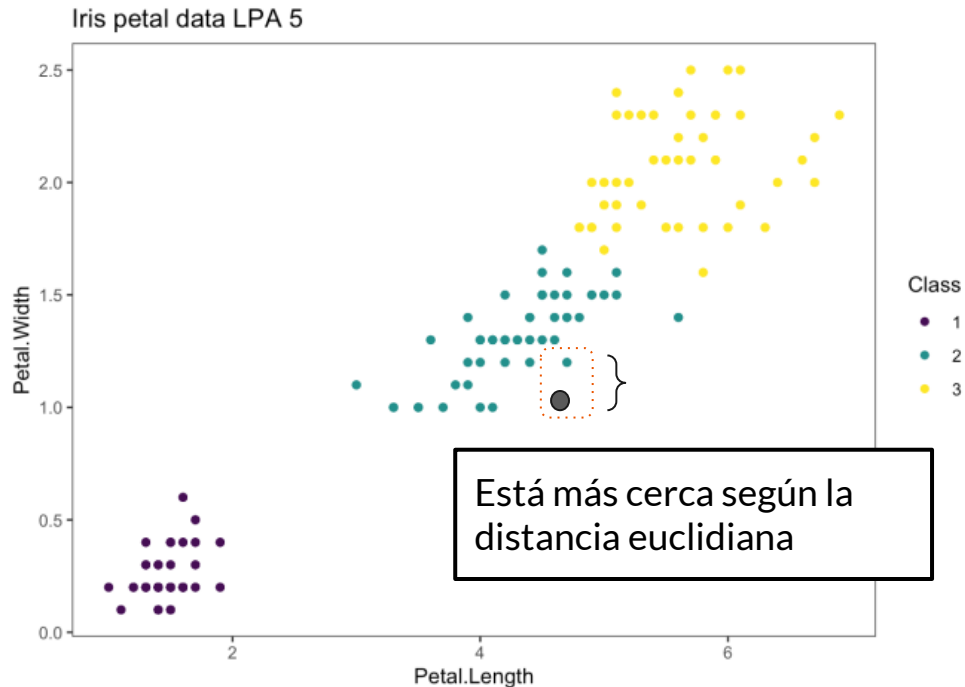
K-nearest neighbors

Aparece una nueva observación

¿A qué clase pertenece?

Veamos a su vecino más cercano

Esta nueva observación está más cerca de un verde, entonces es verde



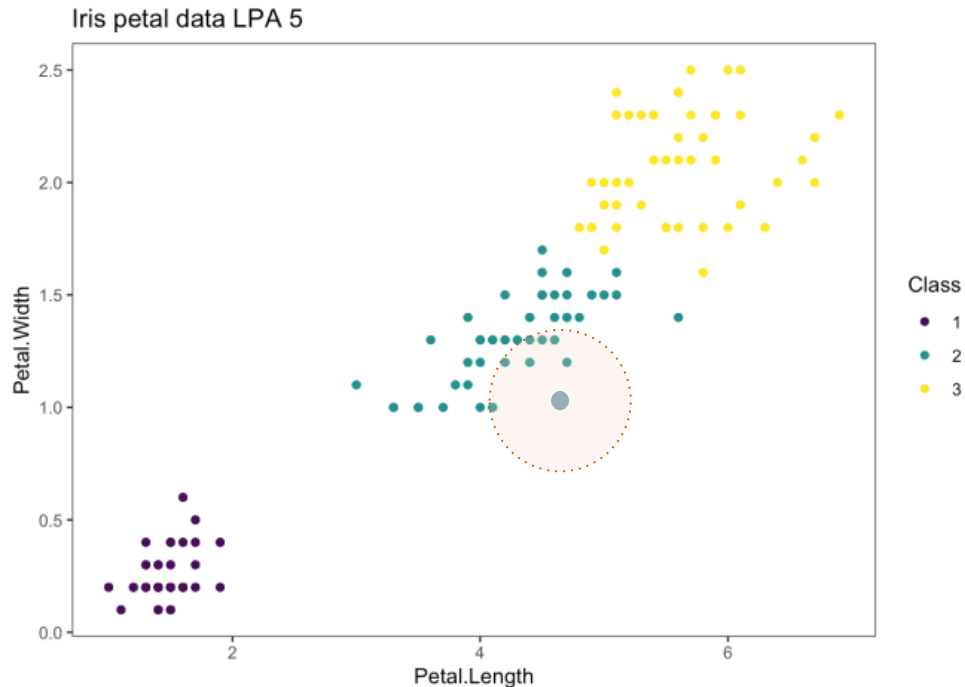
K-nearest neighbors

Aparece una nueva observación

¿Pero es suficiente con tomar el vecino más cercano?

Qué tal si tomamos k vecinos más cercanos.
 $k=7$, por ejemplo

Son todos verdes, no cambia nada.



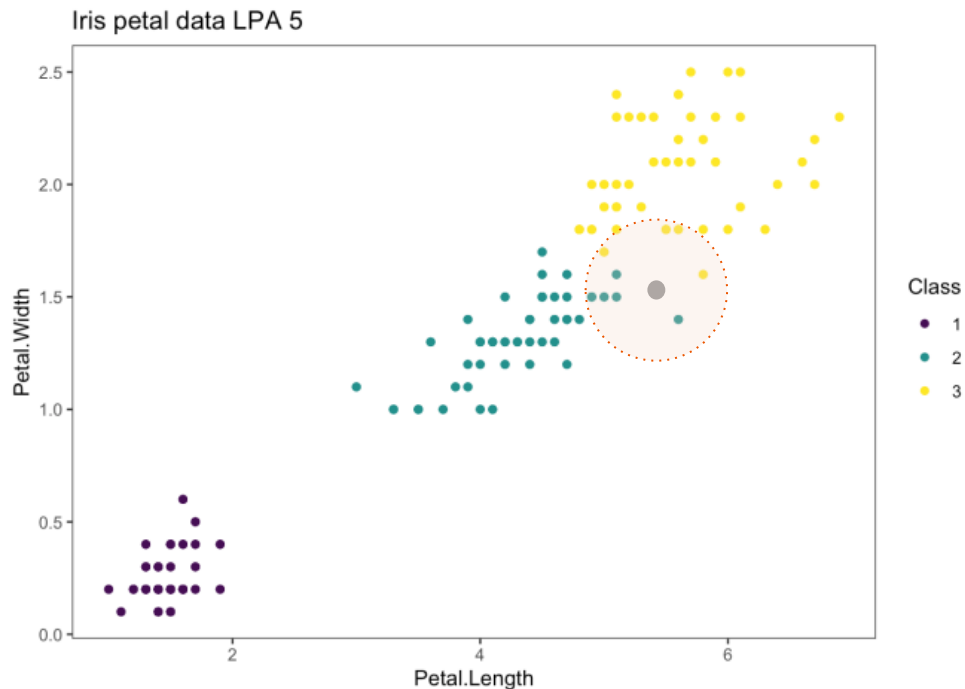
K-nearest neighbors

¿Qué pasa si la nueva observación está en otra ubicación?

Usemos esta vez $k=9$

- 5 vecinos verdes
- 4 vecinos amarillos

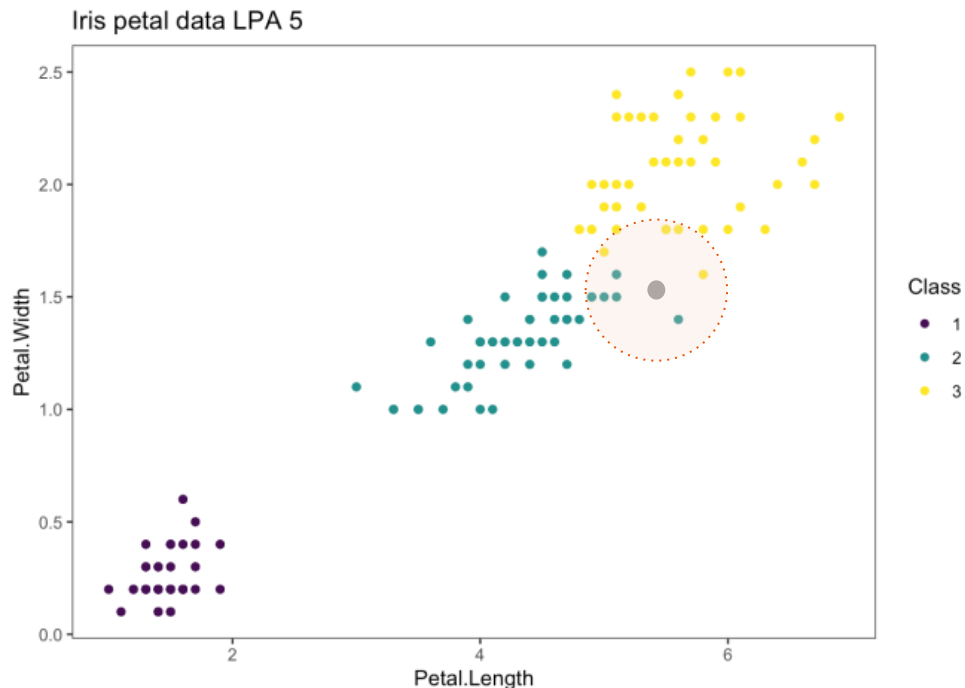
¡Es verde!



K-nearest neighbors

hiperparámetros de K-NN

- El valor **K**
- El tipo de distancia: Euclídea, Manhattan, etc.
- Peso: Los vecinos pueden estar ponderados, siendo que los más cercanos tengan más peso en la votación y los más lejanos menos.
 - ¿Es lo mismo si $k=2$, que uno de los vecinos esté a 1 cm y el segundo a 2m?



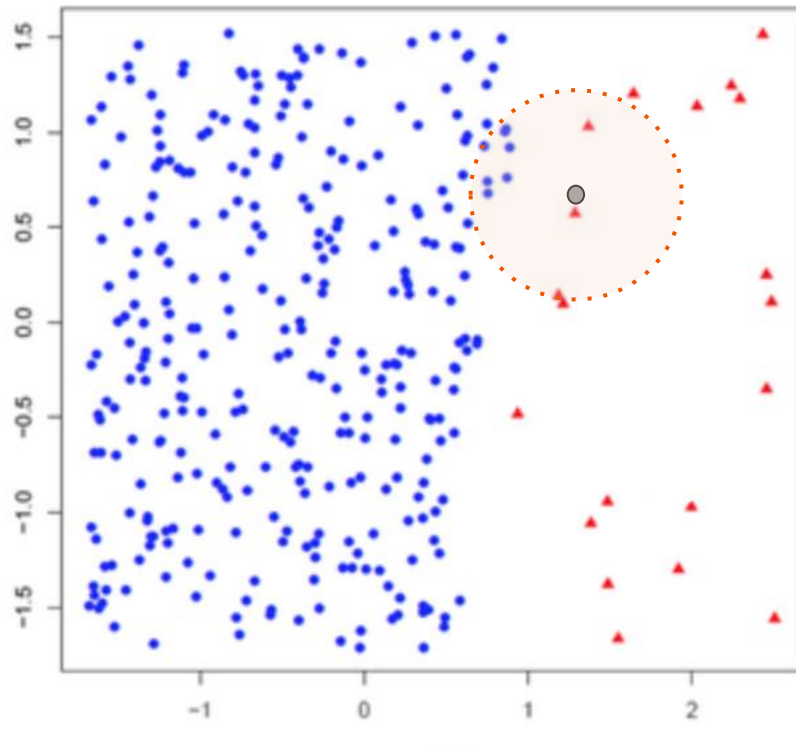
K-nearest neighbors

Conjuntos desbalanceados

k=9

- 3 rojos
- 6 azules

Es azul



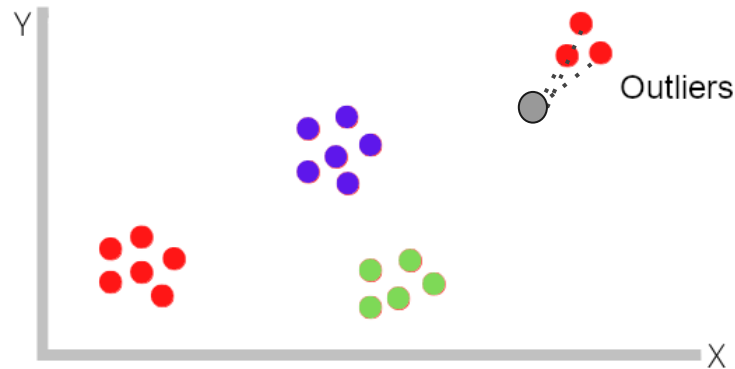
K-nearest neighbors

Outliers

k=3

• 3 rojos

Es rojo





K-nearest neighbors

Resumen:

- Método de clasificación
- Es sensible a conjuntos de datos no balanceados
- Es sensible a outliers
- En **scikit-learn**, $k=5$ por defecto