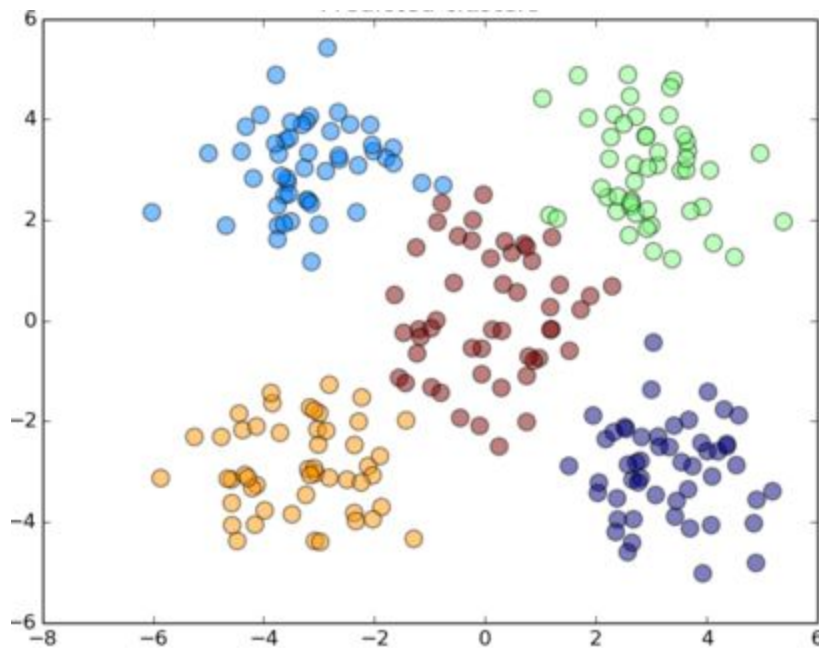


Carlos Catalán

Clasificador de Iris

K-Means



Este informe muestra el desarrollo y resultado de un clasificador de especies de Iris:

- Setosa
- Versicolour
- Virginica

Tomando como referencia la **Longitud y Anchura** del **Sépalo** y el **Pétalo**.

Dataset:

<http://archive.ics.uci.edu/ml/datasets/Iris?fbclid=IwAR28iPffCY5baMbxBc27B50Y7YEytmXBE0WDdN3gAr3nv2ACBx2i8Tuoqo>

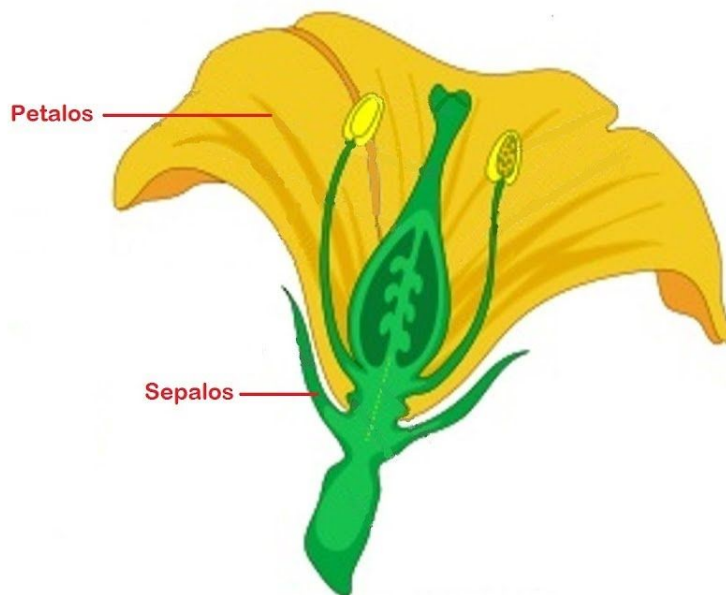
Conclusiones

En el desarrollo del modelo podemos observar que hay cosas que resaltan solamente visualizando los datos. Para la visualización se tomó el pétalo como principal característica.

Dispersión de datos

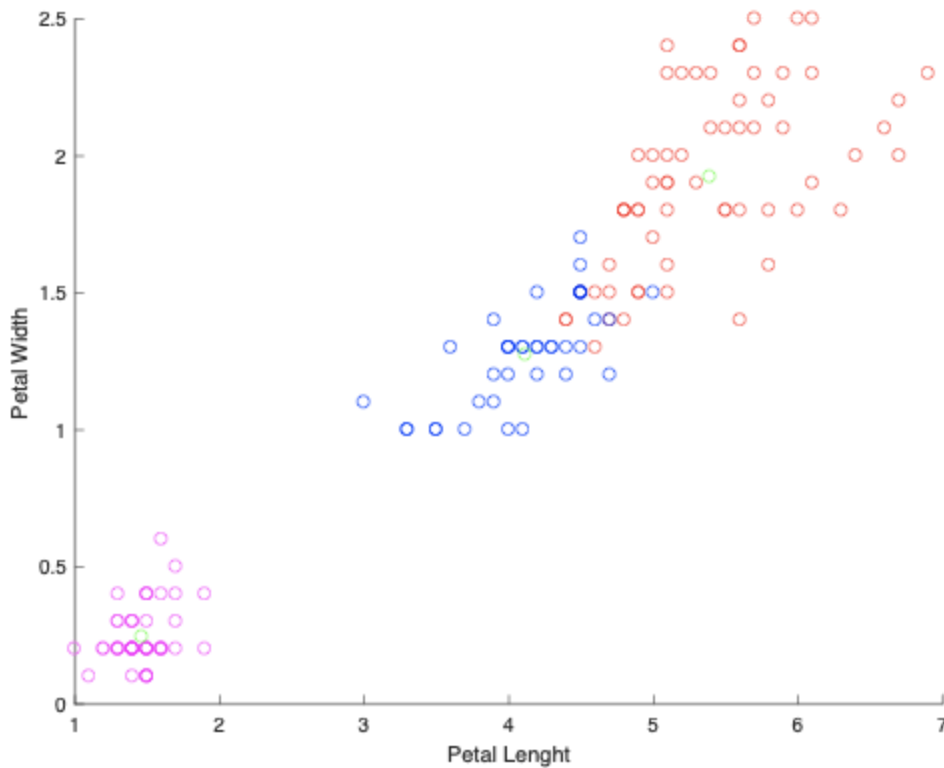
Longitud de Sépalo	Anchura de Sépalo	Longitud de Pétalo	Anchura de Pétalo
0.8281	0.4336	1.7644	0.7632

Como se puede observar hay una diferenciación notoria en cada una de las clases tomando como referencia su pétalo (Longitud y anchura).



Por lo tanto podemos decir que es más **determinante** las medidas del pétalo para poder clasificar los iris.

Entre las especies podemos apreciar que la más diferenciable es la Setosa (Color magenta) ya que las medidas de su pétalo son mucho menores que las de las otras dos.



Las otras dos (Versicolour y Virginica) no tienen un margen tan claro, pero podemos apreciar que la media de la longitud de los pétalos de cada una tienen aproximadamente 2 cm de diferencia, siendo la Virginica la del pétalo más grande (Color rojo).

Datos técnicos:

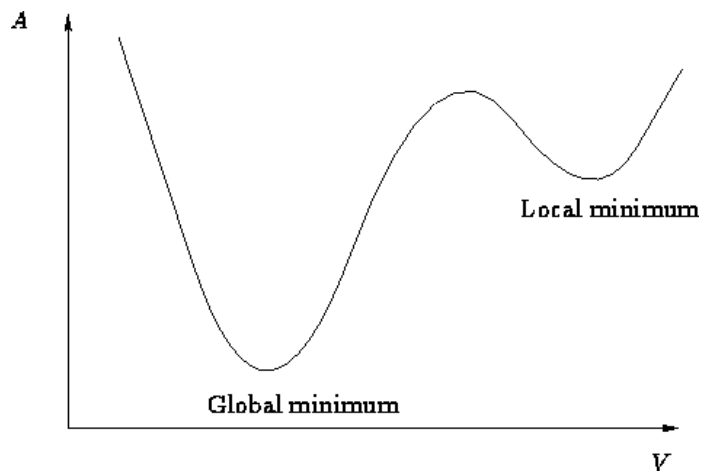
Visualización

Para la visualización de los datos se tomo 2 de las 4 características que se utilizaron en el modelo, la selección de estas se dio mediante el análisis de la desviación estándar (Dispersión) del conjunto de datos de cada característica.

¿Por que K-Means?

K-Means ayuda a encontrar las sumas de las distancias y de acuerdo a las distancias mínimas agruparlos en grupos (Centroides) diferentes. De esta manera creamos grupos a partir de los datos.

Su principal ventaja (con respecto a otros) es que es sencillo y rápido de implementar, mientras que una de sus desventajas es que su resultado final depende de la inicialización de los centroides y en algunas ocasiones tiende a converger en el mínimo local y no en el global.



Una solución a esto es no inicializar aleatoriamente los centroides.