

Actividad Práctica Integradora

Introducción al machine learning

Actividad 4

Situación

Anteriormente, trabajamos con temas relacionados al aprendizaje supervisado, el cual tiene la particularidad de contar con etiquetas de salida para realizar las predicciones de datos del Titanic.

En esta ocasión, nos enfrentamos al desafío de generar predicciones de tipo no supervisado, por lo cual no contamos con una variable de salida objetivo. Para ello, tendremos como foco la agrupación de datos que cumplan con patrones similares y que nos permitan tomar decisiones a posteriori.

En este caso, el “Vivero de San Juan” necesita realizar un estudio que le permita clasificar su nueva flor del iris dentro del vivero, y para ello se han solicitado nuestra ayuda.

Iris es un género de plantas rizomatosas de la familia *Iridaceae*. Constituye el mayor género de la familia, con más de 300 especies, además de muchos híbridos y cultivares. Además del nombre del género, iris se usa comúnmente para referirse a todas las especies, así como a otros varios géneros estrechamente emparentados y a una subdivisión dentro del género.

La flor del iris es de especial interés como ejemplo de la relación entre las plantas florales y los insectos polinizadores. La forma de la flor y la posición de los receptores del polen y los estigmas en los pétalos exteriores forman una pista de aterrizaje para un insecto volador, el cual, al sondear el néctar, entrará primero en contacto con el periantio y después con los tres estambres sobre una superficie en espiral que nace en el ovario formado por los tres carpelos.

La proyección transversal en forma de repisa sobre la espiral interior, situada a los lados de los estambres y por debajo el estilo arqueado bajo el estigma,

Actividad Práctica Integradora

permite que el insecto entre en contacto con la superficie cubierta de polen solamente tras pasar este, mientras que, al salir de la flor, toma contacto con la cara inferior no receptiva del estigma. De este modo, un insecto que transporte polen de una flor visitada anteriormente y entre en otra, depositará el polen sobre el estigma y, cuando salga de ella, el polen que lleva no se frota ni se quedará adherido al estigma de la misma flor.

Consignas

El desafío actual consiste en poder generar agrupaciones mediante K-means que permitan calcular el grado de precisión de una predicción no supervisada. Para ello, nos solicitan analizar la base de datos de la flor del iris y clasificar información según el orden que se presenta a continuación.

1. Debemos describir brevemente la información que nos entrega la base de datos. Para ello, puede ser útil cargar la base mediante el siguiente código:

```
from sklearn import datasets
iris = datasets.load_iris()
```

2. Ahora, debemos graficar las relaciones con las cuales buscaremos realizar el agrupamiento de *clusters* (por ejemplo, largo del sépalo versus ancho del sépalo).
3. Lo siguiente es entrenar el modelo y determinar los centros de cada *cluster*, su inercia y etiqueta.
4. Debemos, además, graficar los *clusters*.
5. Por último, mediante el método del codo, determinemos cuál es el número de *clusters* ideales para realizar la agrupación.

Formato de entrega:

La entrega debe ser mediante un archivo "ipynb" (formato de extensión de *notebook* para Python). Para cada una de las preguntas debe contener los códigos de cómo

Actividad Práctica Integradora

se lograron resolver las consignas solicitadas.

Entrega

¡Llegaste al final de la actividad de este módulo! Recuerda guardar tus respuestas y luego subirlas clicando en el botón "Enviar tarea".

Puedes consultar tus dudas con tus compañeros en el foro de la materia o con tu tutor.