Errores de clasificaciones

Seguro que algunos y algunas de ustedes se han dado cuenta de que no hemos utilizado las primeras características de los albaricoques en nuestra generación de datos:

| **Diámetro mínimo (mm)** | **Diámetro máximo (mm)** | **Peso medio (g)** |
| --- | --- | --- |
| 35 | 39 | 27 |

Le invitamos a modificar el programa de generación de frutas (generacionFrutas.py) incluyendo ahora la primera característica para el albaricoque:

caracteristicasAlbaricoques =

[**[35,39,27]**,[40,44,41],[45,49,54],[50,54,74],[55,59,100]]

No olvide eliminar el archivo frutas.csv presente en el directorio datas.

Una vez realizadas estas operaciones, podemos generar datos nuevos ejecutando el script.

Ahora que disponemos de observaciones nuevas, vamos a verificar que la máquina ha creado dos grupos distintos con éxito y ha clasificado nuestras frutas correctamente. Todo ello con la ayuda de un gráfico.

Gráfico, Gráfico de burbujas

Descripción generada automáticamente

*Errores de clasificaciones*

En la figura anterior podemos comprobar que la máquina ha tenido éxito creando dos grupos distintos, pero ha cometido errores en las clasificaciones. En efecto, podemos notar que algunos albaricoques se han convertido en cerezas.

Esto es debido al hecho de que las distancias de las observaciones calculadas con el algoritmo K-Mean respecto a los dos centroides han hecho que estos datos estuvieran más próximos del centroide de las cerezas que del de los albaricoques. En este caso, es conveniente probar otro algoritmo dedicado al clustering para ver si es posible corregir estos errores.