

## Práctica 1. Herramientas.

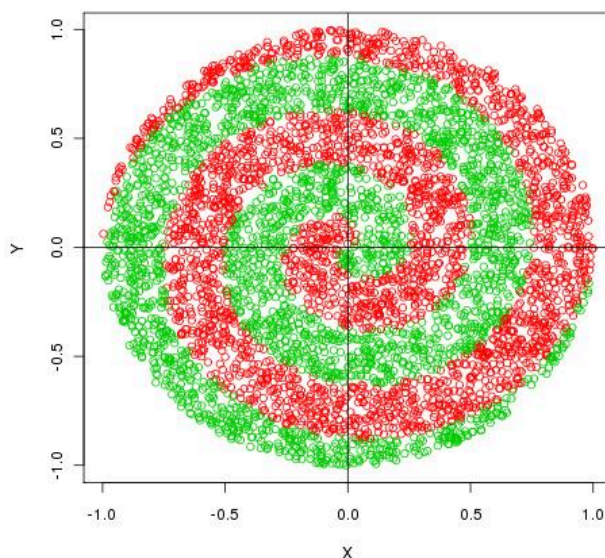
1. Instalar y probar jupyter notebook, o acceder a Google Collab. Desarrollar un Hello world en Python, con comentarios previos sobre el clima del día. Busquen las librerías panda y numpy.
2. Generar datos con python. Diagonal, Paralelo, Espirales. Gráficas

En el entorno elegido, preparen funciones en Python que generen dataframes panda (de longitud dada  $n$ ) de acuerdo a las siguientes descripciones:

a) Los datos tienen  $d$  inputs, todos valores reales, correspondientes a la posición del punto en un espacio  $d$ -dimensional. El output es binario, y corresponde a la clase a la que pertenece el ejemplo. La clase 1 corresponde a puntos generados al azar, provenientes de una distribución normal, con centro en el  $(1, 1, 1, \dots, 1)$  y matriz de covarianza diagonal, con desviación estándar igual a  $C * \text{SQRT}(d)$ . La clase 0 tiene la misma distribución, pero centrada en el  $(-1, -1, -1, \dots, -1)$ . Se puede encontrar información sobre Gaussianas multidimensionales y el caso especial de una matriz diagonal en <http://cs229.stanford.edu/section/gaussians.pdf> (secciones 1 y 3). Los parámetros que se deben ingresar a la función son  $d$  y  $n$  (enteros) y  $C$  (real). De los  $n$  puntos generados,  $n/2$  deben pertenecer a cada clase.

b) Igual al punto anterior, pero las distribuciones tienen centro en el  $(1, 0, 0, \dots, 0)$  y en el  $(-1, 0, 0, \dots, 0)$ , respectivamente y la desviación estándar es igual a  $C$  independientemente de  $d$ .

c) Espirales anidadas: Los datos tienen 2 inputs,  $x$  e  $y$ , que corresponden a puntos generados al azar con una distribución UNIFORME (en dicho sistema de referencia  $x$ - $y$ ) dentro de un círculo de radio 1. El output es binario, correspondiendo la clase 0 a los puntos que se encuentran entre las curvas  $r = \theta/4\pi$  y  $r = (\theta + \pi)/4\pi$  (en polares) y la clase 1 al resto. De los  $n$  puntos generados,  $n/2$  deben pertenecer a cada clase. La siguiente figura es un ejemplo:



(notar la paleta de colores que evita combinaciones desagradables a la vista)

Para verificar los problemas a) y b), generen conjuntos con  $d=2$ ,  $n=200$  y  $C=0.75$ , y gráfíquenlos. También genere conjuntos con  $d=4$ ,  $n=5000$  y  $C=2.00$ , y verifiquen en el código que las medias y desviaciones estándar sean correctas.

Para el problema c), generen un gráfico con  $n=2000$  y compárenlo con el que está arriba.

