

Clasificación

Usaremos la base de datos de Iris de Plantas [1] y usaremos solo las clases de Iris-Setosa e Iris-Versicolor. En total son 100 ejemplares, 50 de cada clase y todos serán posicionados aleatoriamente. El 60 % de los ejemplares se usará para entrenar los algoritmos y el 40 % restante lo usaremos para hacer le testing y se usaran todas las características (en tota 4). Los resultados son los siguientes.

Regresión logística

Ya que esta ejecuta la gradiente descendiente para optimizar el error, el resultado varía entre ejecuciones como en 2 y 3 . Siendo el mayor un 100 % de acierto 1.

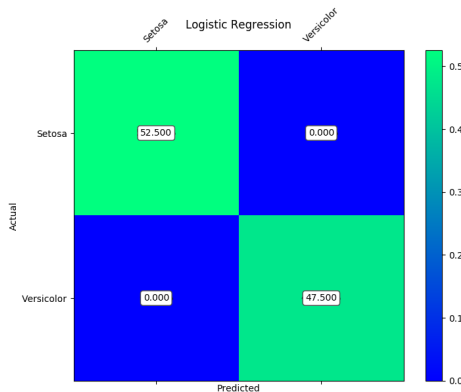


Figura 1: Mejor Resultado

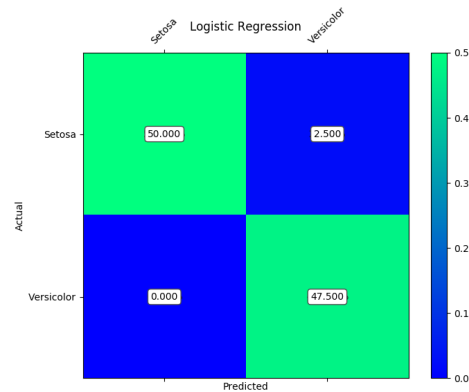


Figura 2: Resultado variante

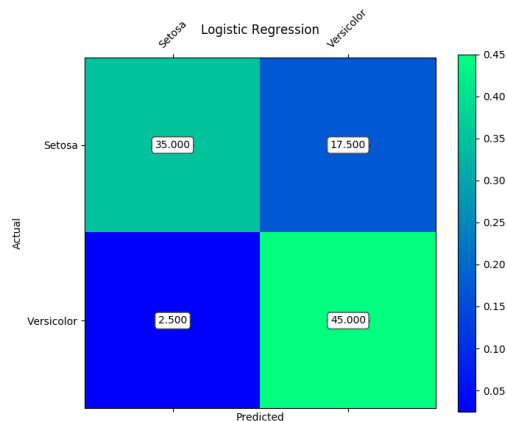


Figura 3: Otro resultado variado

SVM

Aquí la variable w y b serán siempre los mismos por lo que el resultado será el mismo también.

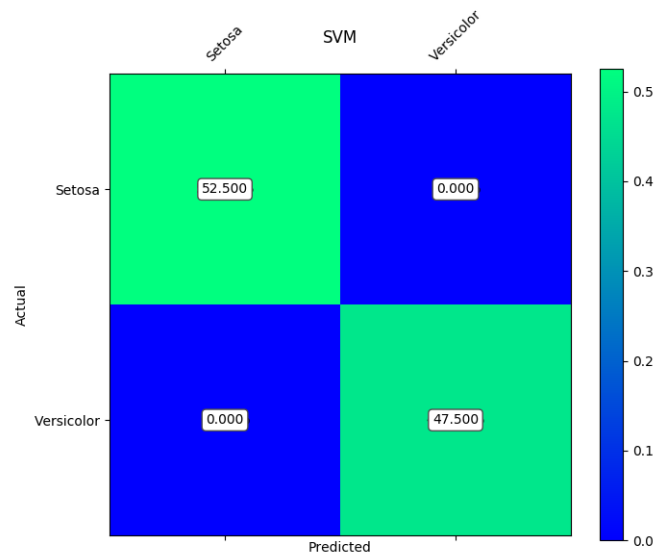


Figura 4: Resultado exitoso con SVM

Como vemos, en el mejor de los casos las dos matrices son iguales. Sin embargo, en la regresión logística el resultado varía ya que como inicializamos las variables θ aleatoriamente, y a esto se suma al hecho que nos detendremos cuando detectemos que el error no este cambiando mucho por lo que el error que nos quede al final va a variar.

Referencias

- [1] Iris Data Set. From Fisher, 1936. <https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>.