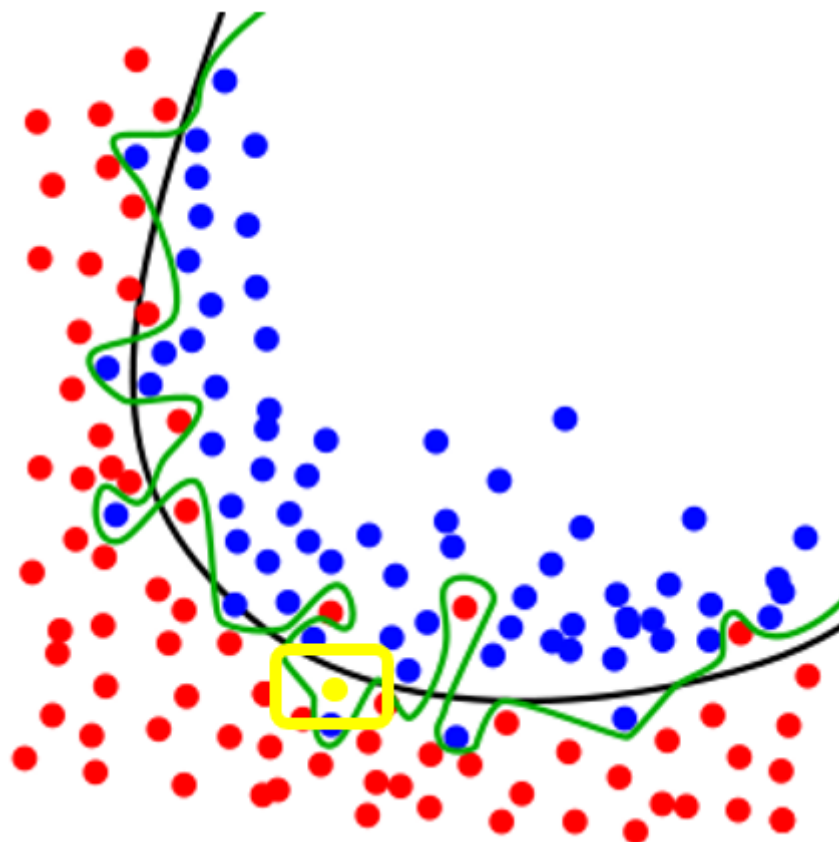


过拟合，正则化

1. 什么是过拟合



來自wiki，藍色跟紅色為當時訓練分類器的資料，黃色為新加入要分類的資料

以上圖來看，綠線就是Overfitting的結果，黑線代表正常的分類模型，綠線雖然完全把訓練資料分類出來，但如果現在有一個新的資料進來(黃色點點)，就會造成分類錯誤，因為綠色線的模型在訓練資料的準確率是非常高的，不過在新資料的分類下錯誤率變會提升，以下有個方法可以偵測是否有Overfitting的情況發生。

防止过拟合，划分验证集。

2. 造成Overfitting的原因與解決方式

- 訓練資料太少
- 擁有太多的參數，功能太強的模型
 - 減少參數或特徵或者是減少神經層數
 - 在相同參數跟相同資料量的情況下，可以使用Regularization(正規化)
 - 在相同參數跟相同資料量的情況下，可以使用Dropout

3. Regularization (正規化)

Weight decay(權重衰減)

Weight decay的意思就是對擁有較大權重的參數，課以罰金，藉此控制Overfitting的情況，因為Overfitting就是**Weight 太大的時候可能會發生**的問題。

Weight decay的方式就是在loss function (損失函數)加入**參數權重的L2 norm**，就可以抑制權重變大我直接用以下公式來介紹 —

$$L_{weight_decay} = L + \frac{1}{2} \lambda \sum_w W^2$$

Weight's L2 norm

L2 norm就是把全部weight的平方和除2

$$W \leftarrow W - \eta \frac{\partial L_{weight_decay}}{\partial W}$$
$$W \leftarrow (1 - \eta \lambda) W - \eta \frac{\partial L}{\partial W}$$

加入**weight decay**後的**gradient decent** 更新

以上是使用Weight decay後的Gradient decent的參數更新推導，L是loss function，也就是損失函數，做Weight decay就是在loss function上加上Weight的L2 norm，進而推導出weight 更新就是上面的那個公式，可以看出跟原本的gradient decent更新比較，在一開始原本的weight乘上了(1- ηλ)，因為 η，λ 都為正，因此可以減少原本的weight的影響，越大的weight就變越小，越小的weight改變就不大，這就是Weight decay的由來。