

In [ ]: 【ハイパーパラメータ・チューニング】  
これまで機械学習の手法として、SVMやランダムフォレストを使って、分類や予測をやってきましたが、実は、SVMやランダムフォレストには、ハイパーパラメータという「手動」で設定する必要がある変数があります。前にやった時は、ハイパーパラメータをデフォルト値のまま実行したことになります。  
今回は、また再びタイタニックを使い、このハイパーパラメータを見ていきましょう。

In [ ]: 問1  
機械学習アルゴリズムのハイパーパラメータがどういう役割か、また実際にSVM・ランダムフォレストの主なハイパーパラメータにどのようなものがあるか、またデフォルトの値が何か、調べてください。

In [20]: # ハイパーパラメータとは、機械学習の実行による推論や予測では決定されない、  
# プログラム実施者が任意に決定するパラメータを言う。なお、「パラメータ」  
# は、機械学習モデルが学習の過程で自動的に最適化を行うものを指す。  
#  
# <主なハイパーパラメータ>  
# 1 SVM  
# ① kernel(デフォルト:rbf - radial basis function)  
# カーネルは、SVMの演算を行う'関数'、他に、'linear','poly'等の  
# 選択肢がある  
# ② C(コスト、デフォルト:1.0 <-実数)  
# SVMの損失関数の誤差の係数(誤分類を許容する程度)  
# ③ gamma( $\gamma$ 、デフォルト:1/(特徴量の数\*分散) <- 実数)  
# kernel が 'rbf'・'poly'・'sigmoid'の際のカーネル関数の係数  
# 2 ランダムフォレスト  
# ① n\_estimators(デフォルト:100-整数)  
# ランダムフォレストは、「決定木」に設定した条件で元の集合を2分割し  
# さらに別れた先で再度2分割する、を繰り返すが、その決定木の数  
# ② max\_depth(デフォルト:None)  
# 決定木のノードの深さ(=階層)  
# ③ max\_features(デフォルト:特徴量nの平方根)  
# 決定木それぞれに対し、ランダムに指定する特徴量の数。