

# 宿題ガイドライン

## 1. 分類問題： <https://www.kaggle.com/ronitf/heart-disease-uci>

データファイル： heart.csv

目的変数： target (2値問題)

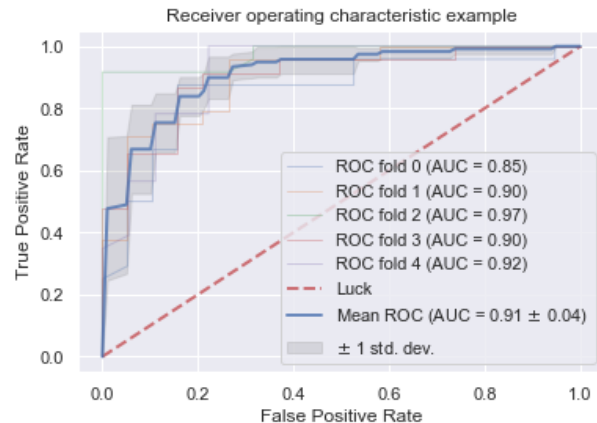
学習データ： テスト(検証)データ比率： 7:3

評価指数： 正解率(Acuracy)、再現率(Recall)およびAUC

ヒント： 説明指数を統計的に処理する

参考：井伊が Python で解いた結果

	precision	recall	f1-score	support
False	0.82	0.90	0.86	40
True	0.91	0.84	0.88	51
accuracy				0.87
macro avg	0.87	0.87	0.87	91
weighted avg	0.87	0.87	0.87	91



# 宿題ガイドライン

## 2. 回帰問題: <https://www.kaggle.com/c/new-york-city-taxi-fare-prediction/overview>

データファイル: NYC\_taxi\_10000\_cleansed.csv

目的変数: fare\_amount

学習データ: テスト(検証)データ比率: 7:3

評価指数: 相関係数 (R)、決定係数 (R2)

ヒント:

- (1) 時間データの取り扱い
- (2) 経度・緯度データの取り扱い
- (3) 異常値(ノイズ)の処理

参考: 井伊が Python で解いた結果

相関係数 (R): 0.9502797777290971

決定係数 (R2): 0.9030316559608622

