

コンピューターシミュレーション2 第4回 演習課題

Dataset: auto_mpg.csv

(1) Sample code: S4-Sample_auto_mpg-knnReg.ipynb

- (a) データを学習データと評価データに分ける。
- (b) *horsepower* を独立変数 X 、*mpg* を従属変数 Y として KNN 回帰モデル($k=3$)を作成する。
- (c) 評価データを用いて平均二乗誤差 (MSE)と R^2 値を計算し、モデルの精度検証を行う。
- (d) k の値や独立変数 X を変更して(b) ~ (c)を繰り返す。3つ以上のモデルを作成すること。

Note: KNN 回帰モデルは複数の独立変数 X も使用できる。(d)ではデータセット内の *horsepower*, *weight*, *displacement* の3つの中から1~3個の独立変数を選んで使用する。

(2) Sample code: 5-Sample_auto_mpg-lm.ipynb

Part-1

- (a) *horsepower* を独立変数 X 、*mpg* を従属変数 Y として線形回帰モデルを作成する。
- (b) X^2 の項を含めたモデルを作成する (X^2 の項を含めて(a)を繰り返す)。
- (c) データをプロットし、線形回帰モデルと X^2 の項を含めたモデルを重ねる。データの相関などを考察する。

追加課題：

別の独立変数 X を1つ選んで(a)~(c)を繰り返す。

Part-2

- (d) データを学習データと評価データに分ける。
- (e) *horsepower*、*weight* と *displacement* の3つの中から1つ独立変数 X として選び、 X^2 の項を含めた場合と含めない場合で、学習データの線形回帰モデルを作成する。
- (f) 評価データを用いて平均二乗誤差(MSE)と R^2 値を計算し、モデルの精度検証を行う。

追加課題：

複数の独立変数の組み合わせを用いてモデルを作成する。評価データを用いて平均二乗誤差 (MSE) と R^2 値を計算することで、モデルの精度検証を行う。 X^2 の項を含めた場合、含めない場合、変数の組み合わせを変更した場合など、3つ以上のモデルを作成する。

(3) Discussion

- (a) 構築したすべてのモデル (KNN 回帰モデル・線形回帰モデル) のうち、MSE を基に最適な3つのモデルを選ぶ。ただし MSE は小さいほうがより良い。
- (b) 構築したすべてのモデルから R^2 値を基に最適な3つのモデルを選ぶ。ただし R^2 値は高いほうがより良い。
- (c) (a)と(b)で得られた結果から、使用したモデルや特徴変数についてどれが最も良いか比較し、その結果について考察しなさい。