

ベイズ推定による 多項式回帰モデルでの 自動車の燃費予測

工学院大学

予測モデリング 最終課題

概要

- 自動車の重さ（weight）を説明変数として、燃費（MPG, Miles per Gallon）を予測する回帰モデルを構築

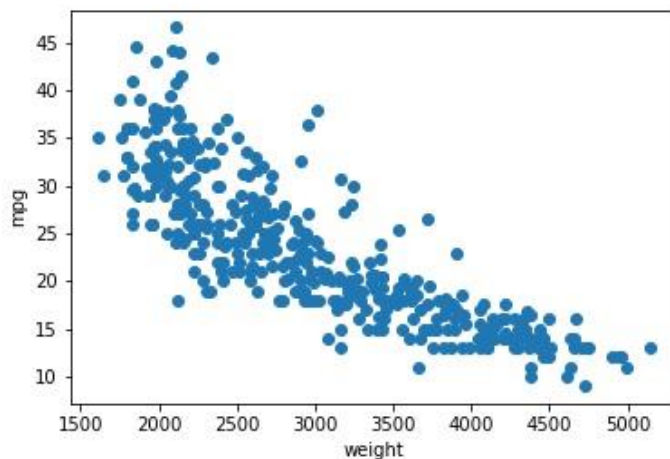
- 3次曲線のフィッティングを行った

$$y = w_0 + w_1x + w_2x^2 + w_3x^3$$

- データセット

- UCIのAuto MPG Data Setを利用

➤ <https://archive.ics.uci.edu/ml/datasets/auto+mpg>



ベイズ推定によりパラメータを推定

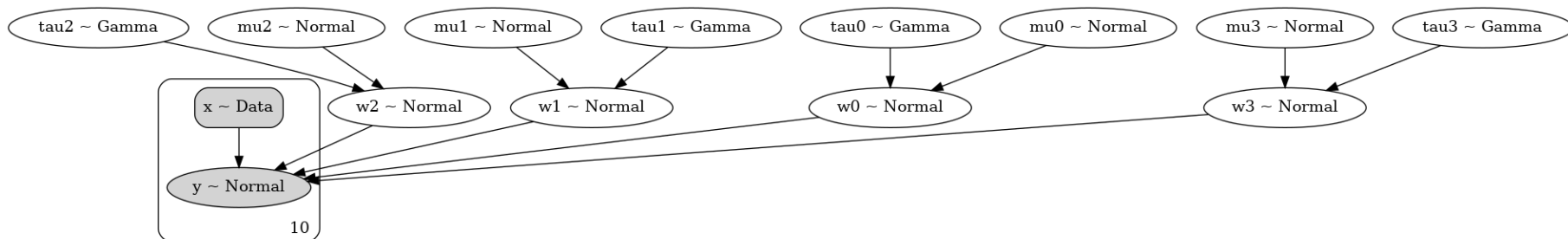
- 重みパラメータが正規分布に従い、
その正規分布に関して
平均 μ_i が正規分布、精度 τ_i がガンマ分布に従う ($i = 0, 1, 2, 3$)

$$w_i \sim \mathcal{N}(\mu_i, \tau_i), \mu_i \sim \mathcal{N}(\mu, \tau), \tau_i \sim \text{Gamma}(\alpha, \beta)$$

$$y \sim \mathcal{N}(w_0 + w_1x + w_2x^2 + w_3x^3, \tau_y)$$

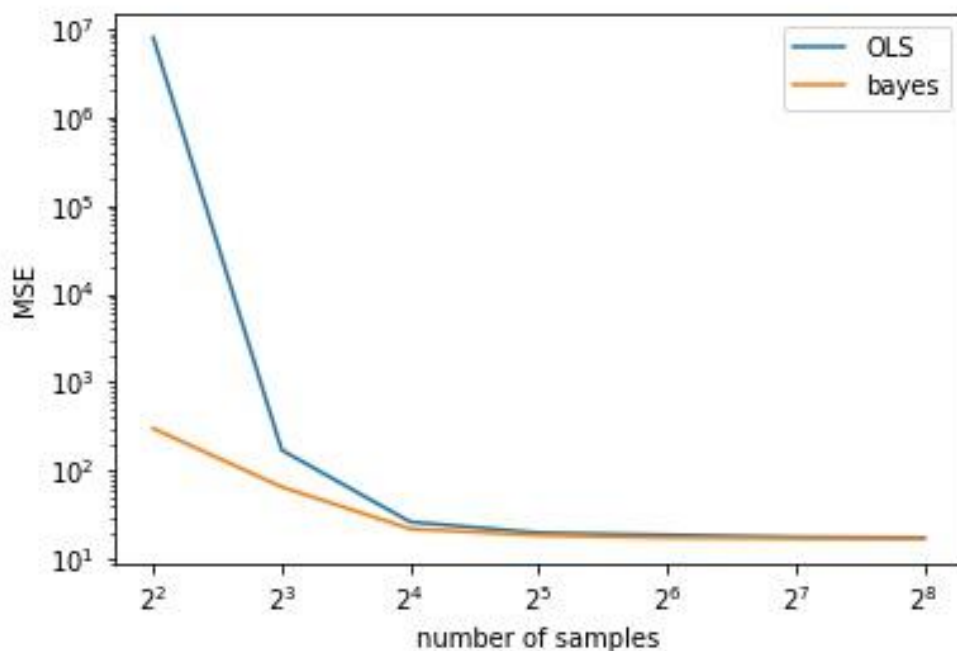
- ベイズ推定にはPyMC3を利用

- 事前分布が正規分布であるものに関して、平均0、精度1/9とした
- 事前分布がガンマ分布であるものに関して、 $\alpha = 1, \beta = 1$ とした



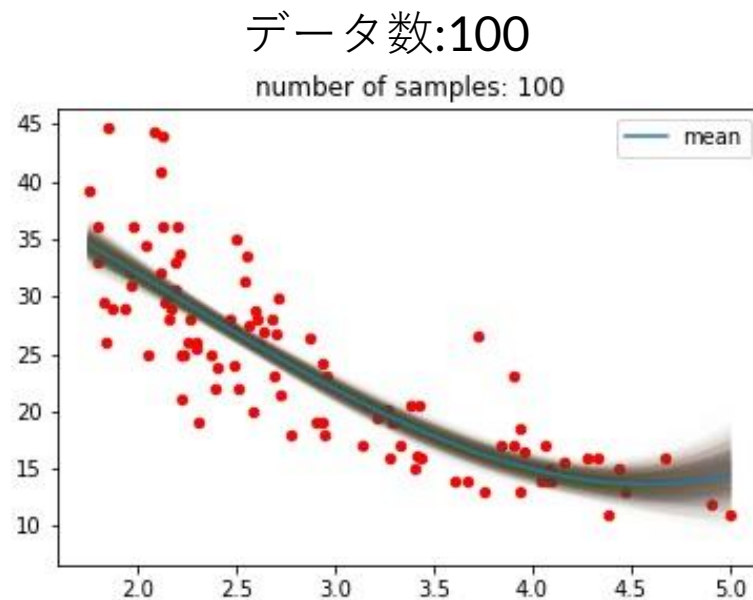
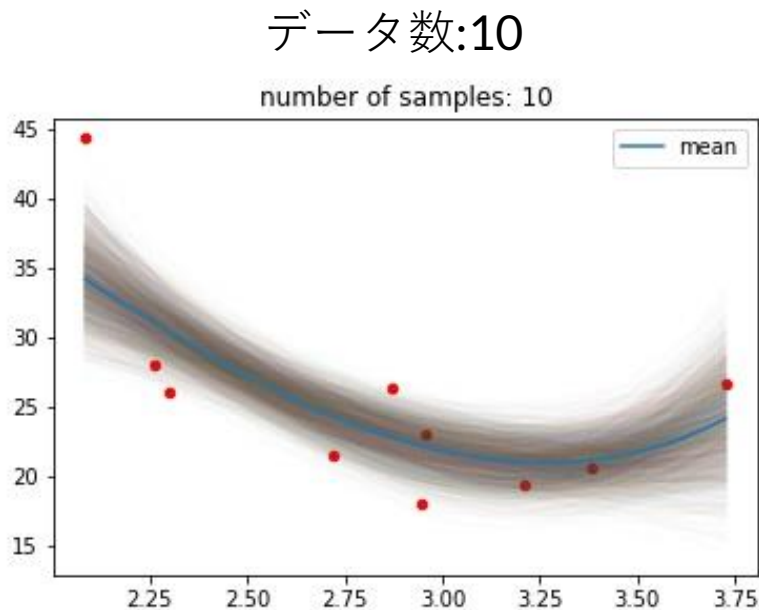
学習データ数を変えたときの予測性能

- データを訓練用:テスト用で8:2に分割
- テストデータに対する平均二乗誤差（MSE）により評価
 - 最小二乗法による回帰との比較
- データ数が少ない場合、OLSと比較してベイズ推定によるものはMSEが小さい



回帰曲線のサンプリング

- サンプリングされた回帰曲線を図示
 - 赤点は観測（学習）データ
- データ数が少ない場合、サンプリングされる曲線のばらつきは大きい



参考

- 須山敦志, 杉山将(監修). (2017).
ベイズ推論による機械学習入門
機械学習スタートアップシリーズ. 講談社.
- 東京大学教養学部統計学教室. (1991).
統計学入門(基礎統計学I). 東京大学出版会.