0.1　AIC

次の回帰モデルを考える．

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u},$$

$$\boldsymbol{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix},\ \boldsymbol{X} = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{2n} & \cdots & X_{kn} \end{bmatrix},\ \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix},\ \boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}.$$

ここで，

$$\boldsymbol{u} \sim \mathrm{N}_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$$

を仮定する．ただし，$\boldsymbol{I}_n$ は n 次単位行列．

このとき，

$$\boldsymbol{y} \sim \mathrm{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$$

であるから，尤度関数は，

$$L = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\}$$

となり，対数尤度関数は次式で与えられる．

$$\log L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{1}$$

よって

$$\begin{aligned} \frac{\partial \log L}{\beta} =& \frac{1}{2}(\boldsymbol{X}^\top\boldsymbol{y} - \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}) = 0, \\ \frac{\partial \log L}{\sigma^2} =& -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = 0 \end{aligned}$$

より，$(\boldsymbol{\beta}, \sigma^2)$ の最尤推定量 $(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$ は，次で与えられる．

$$\begin{aligned} \hat{\boldsymbol{\beta}} =& (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}, \\ \hat{\sigma}^2 =& \frac{1}{n}\boldsymbol{e}^\top\boldsymbol{e}, \quad \boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}. \end{aligned} \tag{2}$$

(2) を (1) に代入した最大尤度の符号を逆転させた

$$\ell = -\log L = \frac{n}{2}(\log 2\pi + 1) + \frac{n}{2}\log\hat{\sigma}^2$$

を損失関数という．

$$\mathrm{AIC} = \ell + k$$

を **AIC**（**Akaike Information Criteria**）という．

説明変数の個数 k が増えると，残差平方和 $\boldsymbol{e}^\top\boldsymbol{e}$ は小さくなり，分散の最尤推定量 $\hat{\sigma}^2$ が小さくなるので，損失関数も小さくなる．

したがって，AIC は，損失関数 ℓ に説明変数の個数 k をペナルティとして加えたものと言える．