0.1 LASSO for lineære modeller

Betragt n observationer $\{x_i, y_i\}_{i=1}^n$, hvor $x_i = (x_{i1}, \ldots, x_{ip})$ er en p dimension vektor af forklarende variable eller prediktorer og $y_i \in \mathbb{R}$ er den tilhørende respons variabel.

Det velkendte estimat for mindste kvadraters metode for (β_0, β) findes udfra

$$\min_{\beta_0,\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2\right\}$$

Dette estimat har ofte lav bias men høj varians, hvilket betyder

Lasso finder løsningen til optimerings problemet

$$\min_{\beta_0,\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2\right\}, \quad s.t. \sum_{j=1}^{p}|\beta_j| \leq t \tag{1}$$

Betingelsen $\sum_{j=1}^{p}|\beta_j| \leq t$ kan skrives mere kompakt ved $\|\beta\|_1 \leq t$. Dette kan udtrykkes på matrix-vektor notation. Lad $\mathbf{y} = (y_1, \ldots, y_n)$ være en n dimensional vektor med responsvariable og $\mathbf{X}$ være en $n \times p$ matrix med $x_i \in \mathbb{R}^p$ som den i'te række, da kan (1) omskrives til

$$\min_{\beta_0,\beta}\left\{\frac{1}{2n}\|\mathbf{y} - \beta_0\mathbf{1} - \mathbf{X}\beta\|_2^2\right\}, \quad s.t.\|\beta\|_1 \leq t,$$

hvor $\mathbf{1}$ er en n dimensionel vektor bestående af 1 og $\|\cdot\|_2$ betegner den Euklidiske norm af vektorer.

Grænsen t begrænser summen af de absolutte værdier af parameter estimaterne. Denne skal specificeres ved en ekstern procedure kaldet *kryds validering*, som vil blive diskuteret i kap –.

Ofte standardiseres prediktorerne $\mathbf{X}$ således at kolonnerne er centeret og har varians 1. Dvs $\frac{1}{n}\sum_{i=1}^{n} x_{ij} = 0$ og $\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 = 1$. Hvis ikke prediktorerne standardiseres da vil lasso estimaterne afhænge af enhederne. Hvis prediktorerne er målt i samme enhed, da vil vi typisk ikke standardisere. For fuldstændigheden, antager vi også at responsvariablen y_i er centeret, dvs $\frac{1}{n}\sum_{i=1}^{n} y_i = 0$. Når data er centreret da kan vi se bort fra skæringen β_0 i lasso optimeringen. Given en optimal lasso løsning $\hat{\beta}$ på det centreret data, kan vi finde løsningen for det ikke-centreret data. Der gælder at

$$\hat{\beta}^{\text{ikke-centreret}} = \hat{\beta}^{\text{centreret}}$$

$$\hat{\beta}_0^{\text{ikke-centreret}} = \bar{y} - \sum_{j=1}^{p}\bar{x}_j\hat{\beta}_j$$

Derfor ser vi bort fra skæringen resten af kapitlet.

Vi kan omskrive lasso problemet til Lagrange form

$$\min_{\beta}\left\{\frac{1}{2n}\|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_1\right\} \tag{2}$$

for $\lambda \geq 0$. Af Lagrange dualiteten er der en bijektion mellem (1) og (2): for hver værdi af t hvor $\|\beta\|_1 \leq t$ er opfyldt, da findes en tilhørende værdi af λ som giver den samme løsning for (2). Mens løsningen $\hat{\beta}_\lambda$ til (2) løser grænse problemet med $t = \|\hat{\beta}_\lambda\|_1$