3. $\min_{\beta} \|\beta\|^2 + \left(\sum_{i=1}^{m} s_i\right)$

s.t. $y_i(\beta^T x_i) \geq 1 - s_i$ for $i = 1, \ldots, m$

$s_i \geq 0$ for $i = 1, \ldots, m$

4. Ridge

$$R = \sum_{j=1}^{p} (y_j - \beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

$$R = (y_1 - \beta_1)^2 + \ldots + (y_p - \beta_p)^2 + \lambda(\beta_1^2 + \ldots \beta_p^2)$$

$$\frac{\partial R}{\partial \beta_j} = 2(y_j - \beta_j)(-1) + \lambda(2\beta_j)$$

$$= 2(\beta_j - y_j) + 2\lambda\beta_j$$

$$= 2(\beta_j - y_j + \lambda\beta_j)$$

$$= 2((1+\lambda)\beta_j - y_j)$$

Setting $\frac{\partial R}{\partial \beta_j} = 0$:

$$0 = 2((1+\lambda)\beta_j - y_j)$$

$$0 = (1+\lambda)\beta_j - y_j$$

$$(1+\lambda)\beta_j = y_j$$

$$\beta_j = \frac{y_j}{(1+\lambda)}$$

$$\therefore \boxed{\hat{\beta}_j^R = \frac{y_j}{(1+\lambda)}}$$

4. Lasso

$$L = \sum_{j=1}^{p} (y_j - \beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

$$L = (y_1 - \beta_1)^2 + \ldots + (y_p - \beta_p)^2 + \lambda(|\beta_1| + \ldots + |\beta_p|)$$

$$\frac{\partial L}{\partial \beta_j} = -2(y_j - \beta_j) + \lambda\left(\frac{\beta_j}{|\beta_j|}\right)$$

$$= 2\beta_j - 2y_j + \lambda\left(\frac{\beta_j}{|\beta_j|}\right)$$

setting $\frac{\partial L}{\partial \beta_j} = 0$

$$0 = 2\beta_j - 2y_j + \lambda\left(\frac{\beta_j}{|\beta_j|}\right)$$

$$0 = \beta_j - y_j + \frac{\lambda}{2}\left(\frac{\beta_j}{|\beta_j|}\right)$$

$$\beta_j = y_j - \frac{\lambda}{2}\left(\frac{\beta_j}{|\beta_j|}\right)$$

if $\beta_j > 0$, then $\frac{\beta_j}{|\beta_j|} = 1$
if $\beta_j < 0$, then $\frac{\beta_j}{|\beta_j|} = -1$
therefore

$$\beta_j = y_j - \frac{\lambda}{2} \quad \text{or} \quad \beta_j = y_j + \frac{\lambda}{2}$$

it can be seen that if $y_j > \frac{\lambda}{2}$, then $\beta_j$ must be positive and $\frac{\beta_j}{|\beta_j|} = 1$

$\therefore$ if $y_j > \frac{\lambda}{2}$ then $\hat{\beta}_j = y_j - \frac{\lambda}{2}$

and if $y_j < -\frac{\lambda}{2}$, then $\beta_j$ must be negative and $\frac{\beta_j}{|\beta_j|} = -1$

$\therefore$ if $y_j < -\frac{\lambda}{2}$, then $\hat{\beta}_j = y_j + \frac{\lambda}{2}$

in the case where $|y_j| \leq \frac{\lambda}{2}$,

$$\beta_j = y_j - \frac{\lambda}{2} < 0$$

so $L$ is decreasing for all $\beta_j < 0$

and

$$\beta_j = y_j + \frac{\lambda}{2} > 0$$

so $L$ is increasing for all $\beta_j > 0$

therefore

$$\therefore \text{ if } |y_j| \leq \frac{\lambda}{2}, \text{ then } \hat{\beta}_j = 0$$

$$\boxed{\therefore \hat{\beta}_j^L = \begin{cases} y_j - \frac{\lambda}{2} & \text{if } y_j > \frac{\lambda}{2} \\ y_j + \frac{\lambda}{2} & \text{if } y_j < -\frac{\lambda}{2} \\ 0 & \text{if } |y_j| \leq \frac{\lambda}{2} \end{cases}}$$