penalty	function	smooth version	optimizer	reference
ridge	$p(x_j) = \lambda x_j^2$	λx_j^2	bfgs, glmnet, ista	(Hoerl & Kennard, 1970)
lasso	$p(x_j) = \lambda \lvert x_j \rvert$	$\lambda\sqrt{x_j^2+\varepsilon}; \varepsilon > 0$	bfgs, glmnet, ista	(Tibshirani, 1996)
adaptiveLasso	$p(x_j) = \frac{1}{w_j}\lambda \lvert x_j \rvert$	$\frac{1}{w_j}\lambda\sqrt{x_j^2+\varepsilon}; \varepsilon > 0$	bfgs, glmnet, ista	(Zou, 2006)
elasticNet	$p(x_j) = \alpha\lambda \lvert x_j \rvert + (1-\alpha)\lambda x_j^2$	$\alpha\lambda\sqrt{x_j^2+\varepsilon} + (1-\alpha)\lambda x_j^2; \varepsilon > 0$	bfgs, glmnet, ista	(Zou & Hastie, 2005)
cappedL1	$p(x_j) = \lambda \min(\lvert x_j \rvert, \theta); \theta > 0$	$--$	glmnet,ista	(Zhang, 2010)
lsp	$p(x_j) = \lambda \log(1 + \lvert x_j \rvert/\theta); \theta > 0$	$--$	glmnet,ista	(Candès et al., 2008)
scad	$p(x_j) = \begin{cases} \lambda \lvert x_j \rvert & \text{if } \lvert x_j \rvert \leq \lambda \\ \frac{-x_j^2 + 2\theta\lambda\lvert x_j \rvert - \lambda^2}{2(\theta-1)} & \text{if } \lambda < \lvert x_j \rvert \leq \lambda\theta \\ (\theta+1)\lambda^2/2 & \text{if } \lvert x_j \rvert \geq \theta\lambda \end{cases}; \theta > 2$	$--$	glmnet,ista	(Fan & Li, 2001)
mcp	$p(x_j) = \begin{cases} \lambda \lvert x_j \rvert - x_j^2/(2\theta) & \text{if } \lvert x_j \rvert \leq \theta\lambda \\ \theta\lambda^2/2 & \text{if } \lvert x_j \rvert > \lambda\theta \end{cases}; \theta > 0$	$--$	glmnet,ista	(Zhang, 2010)