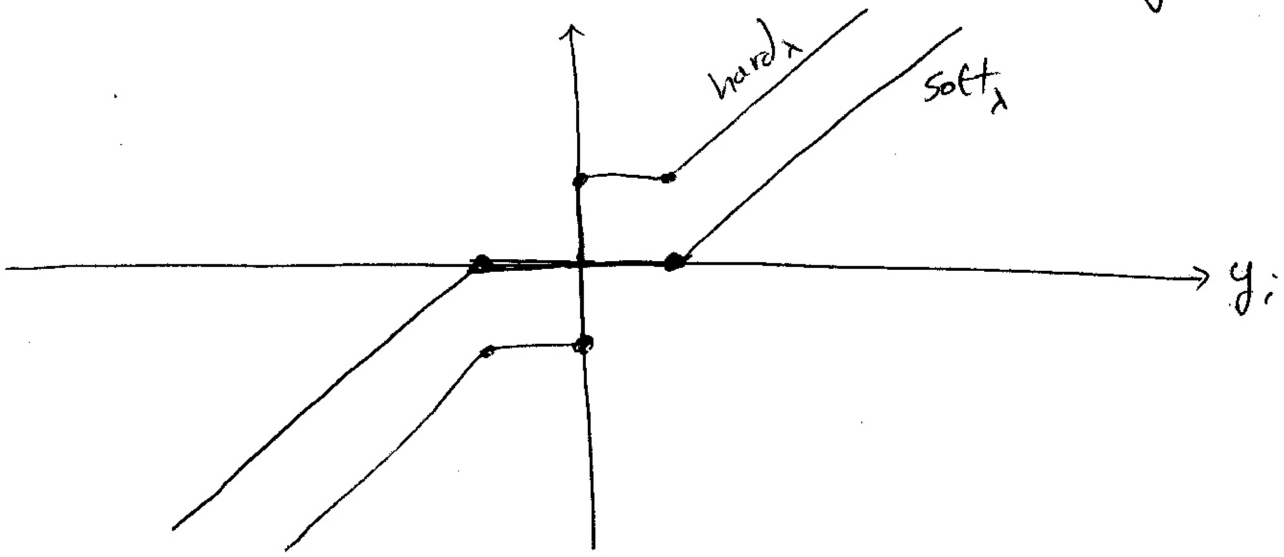
▷ $\text{soft}_\lambda(y_i) = \text{sign}(y_i)\cdot(|y_i|-\lambda)_+$ is called soft-thresholding.

▷ $\text{hard}_\lambda(y_i) = y_i \cdot \mathbb{1}\{|y_i|>\lambda\}$ is hard-thresholding



Excess Risk and BV Tradeoff

Recall $R_n(g_\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell_i(g_\theta)$ and $R(g) = \mathbb{E}\,\ell_1(g)$

(ERM) $\min_{\theta\in\Theta} R_n(g_\theta) \longrightarrow g_{\hat\theta}$ Let $\theta^* = \underset{\theta\in\Theta}{\text{argmin}}\, R(g_\theta)$

and

$g^* = \underset{g}{\text{argmin}}\, R(g)$

$$
\begin{aligned}
R(g_{\hat\theta}) &= \left[R(g_{\hat\theta}) - R_n(g_{\hat\theta})\right] + R_n(g_{\hat\theta}) \\
&\leq \left[R(g_{\hat\theta}) - R_n(g_{\hat\theta})\right] + R_n(g_{\theta^*}) \\
&= \left[R(g_{\hat\theta}) - R_n(g_{\hat\theta})\right] + \left[R_n(g_{\theta^*}) - R(g_{\theta^*})\right] + R(g_{\theta^*}) \\
&\leq \underbrace{2\left[\sup_{\theta\in\Theta} |R(g_\theta) - R_n(g_{\hat\theta})|\right]}_{\text{"Variability"}} + \underbrace{\left[\sup_{\theta\in\Theta} |R(g_\theta) - R(g^*)|\right]}_{\text{"Bias"}} + \underbrace{R(g^*)}_{\text{Bayes risk}}
\end{aligned}
$$