3.

$$\Rightarrow \min_{\vec{w}, b, \vec{\xi}} \frac{1}{2}\|\vec{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n} \xi_i^2$$

$$\text{s.t. } \xi_i = \max\{0, 1 - y_i(\vec{w}^T\vec{x}_i + b)\}.$$

It assocrates with ERM with squared ~~hindge~~ hinge loss

(b). if $\xi_i \geq 0$ dropped,

obj.fun reduces to: $\min_{\vec{w}, b, \vec{\xi}} \frac{1}{2}\|\vec{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n} \xi_i^2$

$\text{s.t. } \xi_i \geq 1 - y_i(\vec{w}^T\vec{x}_i + b).$

suppose if $1 - y_i(\vec{w}^T\vec{x}_i + b) < 0$. so that ξ_i can be ~~choosen~~ chosen as less than zero satisfying the constraints above.

$(\xi_i \geq 1 - y_i(\vec{w}\vec{x}_i + b))$

Then it's obvious that for $\xi_i < 0$. $\xi_i^2 > 0^2$.

for we choose these ξ_i to be zero still satisfies constraints above but with smaller objective value

Thus there is no need to specify $\xi_i \geq 0$ because we will choose $\xi_i = 0$ for those ξ_i can be less than zero.

(C) advantage: Since this loss uses squared slack variables, to minimize the objective function, there will be more penlty to data points which violate the margin.

disvantage: But alse more penlty means there will be less robustness than hinge loss.