2.

4) Subgradient methods for the optimal soft margin hyperplane

since $J(\vec{w},b) = \frac{1}{n}\sum_{i=1}^{n} L(y_i, \vec{w}^T\vec{x}_i + b) + \frac{\lambda}{2}\|\vec{w}\|^2$

Then for $J_i(\vec{w},b)$ satisfying $J(\vec{w},b) = \sum_{i=1}^{n} J_i(\vec{w},b)$

$\Rightarrow$ $\underline{J_i(\vec{w},b) = \frac{1}{n} L(y_i, \vec{w}^T\vec{x}_i + b) + \frac{\lambda}{2n}\|\vec{w}\|^2}$ where $L(y,t) = \max\{0, 1-yt\}$

Now determine $\vec{u}_i$: $\vec{u}_i = \nabla J_i(\vec{w},b) = \nabla J_i(\vec{\theta})$

suppose $g(z) = \max\{0, z\}$ $h_i(\vec{z}) = 1 - y_i(\vec{z}\cdot\tilde{x}_i)$

where $\tilde{x}_i = [1 \ \ \vec{x}_i]^T$

Then
$$\begin{aligned}
\vec{u}_i &= \nabla J_i(\vec{\theta}) \\
&= \frac{1}{n}\nabla[g(h_i(\vec{\theta}))] + \frac{\lambda}{n}[0 \ \ \vec{w}^T]^T \\
&= \frac{1}{n}[\nabla h_i(\vec{\theta})\cdot g'(h_i(\vec{\theta}))] + \frac{\lambda}{n}[0, \vec{w}^T]^T \\
&= \frac{1}{n}(-y_i\cdot\tilde{x}_i)\cdot \text{scale} + \frac{\lambda}{n}[0 \ \ \vec{w}^T]^T
\end{aligned}$$

where
$$\text{Scale} = \begin{cases} 1 & \text{if } 1 - y_i(\vec{\theta}^T\hat{x}_i) > 0 \\ 0 & \text{if } 1 - y_i(\vec{\theta}^T\tilde{x}_i) < 0 \\ \text{random number in } (0,1) & \text{if } 1 - y_i(\vec{\theta}^T\tilde{x}_i) = 0. \end{cases}$$

$\hat{x}_i = [1 \ \ \vec{x}_i]^T$

5) ERM Losses

(a) merge the condition:

$$\left.\begin{aligned} y_i(\vec{w}^T\vec{x}_i + b) &\geq 1 - \xi_i \quad \forall i \\ \xi_i &\geq 0 \quad \forall i \end{aligned}\right\} \Leftrightarrow \xi_i \geq \max\{0, 1 - y_i(\vec{w}^T\vec{x}_i + b)\}$$

Thus obj. fun deduces to:

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

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