A. PROOF OF THEOREM 2.2

Consider a loss function $L(\boldsymbol{y}, \boldsymbol{t}) = \sum_{c=1}^{K} t_c \Psi_c(\boldsymbol{y})$ and $\Psi_c(\boldsymbol{y}) = f(\sum_{k=1}^{K} \phi(y_c - y_k))$. We firstly prove that if $t_i < t_j$, then $y_i < y_j$ in the minimization optimized loss function $L^*(\boldsymbol{t})$.

We can take $i = 1$ and $j = 2$. Consider an optimal prediction $\boldsymbol{y} = [y_1, y_2, y_3, ..., y_K]$ and a new prediction $\boldsymbol{y}' = [y_2, y_1, y_3, ..., y_K]$. Let $\phi(\xi)$ be a non-increasing function and assume $y_1 > y_2$. We have

$$\begin{aligned}
L(\boldsymbol{y}', \boldsymbol{t}) - L(\boldsymbol{y}, \boldsymbol{t}) &= \sum_{c=1}^{K} t_c f(\sum_{k=1}^{K} \phi(y'_c - y'_k)) - \sum_{c=1}^{K} t_c f(\sum_{k=1}^{K} \phi(y_c - y_k)) \\
&= \sum_{c=1}^{K} t_c (f(\sum_{k=1}^{K} \phi(y'_c - y'_k)) - f(\sum_{k=1}^{K} \phi(y_c - y_k))) \\
&= y_1(f(\sum_{k=1}^{K} \phi(y_2 - y_k)) - f(\sum_{k=1}^{K} \phi(y_1 - y_k))) + y_2(f(\sum_{k=1}^{K} \phi(y_1 - y_k)) - f(\sum_{k=1}^{K} \phi(y_2 - y_k))) \\
&= (y_2 - y_1)(f(\sum_{k=1}^{K} \phi(y_1 - y_k)) - f(\sum_{k=1}^{K} \phi(y_2 - y_k))) < 0.
\end{aligned} \tag{1}$$

According to Equation (1), $L(\boldsymbol{y}', \boldsymbol{t}) < L(\boldsymbol{y}, \boldsymbol{t})$. However, this is a contradiction to the optimality $L(\boldsymbol{y}, \boldsymbol{t}) = L^*(\boldsymbol{t})$. Therefore, $y_1 \leq y_2$. Now we assume that $\phi(\xi)$ is differentiable. We know the first order condition $\frac{\partial}{\partial y_c} L(\boldsymbol{y}, \boldsymbol{t}) = 0$ at the optimal solution. Then, We have

$$\begin{aligned}
\frac{\partial}{\partial y_c} L(\boldsymbol{y}, \boldsymbol{t}) &= \frac{\partial}{\partial y_c} \sum_{c'=1}^{K} t_{c'} f(\sum_{k=1}^{K} \phi(y_{c'} - y_k)) \\
&= t_c \frac{\partial}{\partial y_c} f(\sum_{k=1}^{K} \phi(y_c - y_k)) + \sum_{c'=1}^{K} t_{c'} \frac{\partial}{\partial y_c} f(\sum_{k=1}^{K} \phi(y_{c'} - y_c)) \\
&= t_c (\sum_{k=1}^{K} \phi'(y_c - y_k)) f'(\sum_{k=1}^{K} \phi(y_c - y_k)) - \sum_{c'=1}^{K} t_{c'} \phi'(y_{c'} - y_c) f'(\sum_{k=1}^{K} \phi(y_{c'} - y_k)) = 0.
\end{aligned} \tag{2}$$

Therefore,

$$t_c (\sum_{k=1}^{K} \phi'(y_c - y_k)) f'(\sum_{k=1}^{K} \phi(y_c - y_k)) = \sum_{c'=1}^{K} t_{c'} \phi'(y_{c'} - y_c) f'(\sum_{k=1}^{K} \phi(y_{c'} - y_k)). \tag{3}$$

Let us assume $y_1 = y_2 = y$, then the above equality implies that

$$t_1 (\sum_{k=1}^{K} \phi'(y - y_k)) f'(\sum_{k=1}^{K} \phi(y - y_k)) = t_2 (\sum_{k=1}^{K} \phi'(y - y_k)) f'(\sum_{k=1}^{K} \phi(y - y_k)). \tag{4}$$

This is not possible since $\sum_{k=1}^{K} \phi'(y - y_k) \leq 2\phi'(0) < 0$. Therefore, $y_1 < y_2$.

Assume $t_1 < t_2$. We show that

$$\inf\{L(\boldsymbol{y}, \boldsymbol{t}) : y_1 \geq y_2\} > L^*(\boldsymbol{t}). \tag{5}$$

This implies ISC. If the claim is not true, then we can find a new prediction $\boldsymbol{y}^{(m)}$ such that $y_1^{(m)} \geq y_2^{(m)}$ and $\lim_m L(\boldsymbol{y}^{(m)}, \boldsymbol{t}) = L^*(\boldsymbol{t})$. However, we already proved that if $t_1 < t_2$, then $y_1 < y_2$. Now, we can say that the function $\Psi_c(\boldsymbol{y})$ is ISC when $f(\xi)$ is differentiable, $f'(\xi) > 0$, $\phi(\xi)$ is differentiable, $\phi(\xi) \geq 0$, $\phi'(\xi) \leq 0$, and $\phi'(0) < 0$.