Linearly separable sets:

$$W_i x + b_i > w_j x + b_j \qquad \forall x \in A_i,\quad j \neq i$$

or $$W_i x + b_i \geq w_j x + b_j + 1 \qquad \forall x \in A_i,\quad j \neq i$$

$$W_i x + b_i \geq \max_{j \neq i} (w_j x + b_j) + 1$$

or $$\max_{j \neq i} (w_j x + b_j) + 1 - (w_i x + b_i) \leq 0 \qquad \forall x \in A_i$$

or $$\sum_{x \in A_i} \mathrm{ReLU}\left(\max_{j \neq i} (w_j x + b_j) + 1 - (w_i x + b_i)\right) = \min$$

or $$L_1(\theta) = \sum_{i=1}^{k} \sum_{x \in A_i} \mathrm{ReLU}\left(\max_{j \neq i} (w_j x + b_j) + 1 - (w_i x + b_i)\right) =$$

$$\theta = (w, b)$$

Lemma: $Wx + b = \begin{pmatrix} w_1 x + b_1 \\ w_2 x + b_2 \\ \vdots \\ w_k x + b_k \end{pmatrix}$ is a classifier if

$$L_1(\theta) = L_1(\alpha\theta) = \min = 0 \qquad \forall \alpha > 0$$

Lemma: $L_1(\theta)$ is convex

Now we consider the following loss function

$$L_2(\theta) = \sum_{i=1}^{k} \sum_{x \in A_i} \left(\log(\mathbb{1}^T e^{Wx+b}) - (W_{i\cdot} x + b_{i\cdot}) \right)$$

$$f(\theta) = \mathrm{ReLU}\left[\max_{j \neq i} (W_{j\cdot} x + b_{j\cdot}) + 1 - (W_{i\cdot} x + b_{i\cdot}) \right]$$

$$g(\theta) = \log[\mathbb{1}^T e^{Wx+b}] - (W_{i\cdot} x + b_{i\cdot})$$

$$= \log \sum_{j=1}^{k} e^{W_{j\cdot} x + b_{j\cdot}} - (W_{i\cdot} x + b_{i\cdot})$$

Lemma: $\lim_{\alpha \to +\infty} \frac{1}{\alpha} f(\alpha\theta) = \lim_{\alpha \to +\infty} \frac{1}{\alpha} g(\alpha\theta)$

$$= \mathrm{ReLU}\left(\max_{j \neq i} (W_{j\cdot} x + b_{j\cdot}) - (W_{i\cdot} x + b_{i\cdot}) \right)$$

$$\frac{1}{\alpha} f(\alpha\theta) = \mathrm{ReLU}\left[\max_{j \neq i} (W_{j\cdot} x + b_{j\cdot}) + \alpha^{-1} - (W_{i\cdot} x + b_{i\cdot}) \right]$$

$$\xrightarrow{\alpha \to +\infty} \mathrm{ReLU}\left(\max_{j \neq i} (W_{j\cdot} x + b_{j\cdot}) - (W_{i\cdot} x + b_{i\cdot}) \right)$$

$$\frac{1}{\alpha} g(\alpha\theta) = \frac{1}{\alpha} \log \sum_{j=1}^{k} e^{\alpha (W_{j\cdot} x + b_{j\cdot})} - (W_{i\cdot} x + b_{i\cdot})$$

$$= \frac{1}{\alpha} \log \sum_{j=1}^{k} e^{\alpha (W_{j\cdot} x + b_{j\cdot} - (W_{i\cdot} x + b_{i\cdot}))}$$

$$\xrightarrow{?} \mathrm{ReLU}\left[\max_{j \neq i} (W_{j\cdot} x + b_{j\cdot}) - (W_{i\cdot} x + b_{i\cdot}) \right]$$

case 1: $x \in A_i$, $\max_{j \neq i} (W_{j\cdot} x + b_{j\cdot}) < W_{i\cdot} x + b_{i\cdot}$

$$\frac{1}{\alpha} g(\alpha\theta) \to 0 = \mathrm{ReLU}\left(\max (W_{j\cdot} x + b_{j\cdot}) - (W_{i\cdot} x + b_{i\cdot}) \right)$$

case 2, $x \notin A_i$, $x \in A_\ell$ $\quad \ell \neq i$, $W_\ell x + b_\ell > W_j x + b_j$, $j \neq \ell$

$$\frac{1}{\alpha} g(\alpha\theta) \to W_\ell x + b_\ell - W_i x + b_i = \mathrm{ReLU}\left(\max (W_{j\cdot} x + b_{j\cdot}) - (W_{i\cdot} x + \right.$$