Question 2

2.a.
Let $\alpha \in \mathbb{R}_{>0}$
Notice σ is positive-homoegeneous, i.e. $\sigma(\alpha x) = \max\{0, \alpha x\} = \alpha \max\{0, x\} = \alpha\sigma(x)$
Denote $\alpha\theta = \left(\alpha W^{(1)}, ..., \alpha W^{(L)}\right)$.
Let h be defined on θ as in question and $\tilde{h}$ defined on $\alpha\theta$ as in question.

Claim. $\tilde{h}_L(x) = \alpha^L h_L(x)$

Proof. Proof by induction on L.
Base : $\tilde{h}_1(x) = \sigma\left(\alpha W^{(1)^T} x\right) \underbrace{=}_{\alpha\, positive\, homogeneous} \alpha\sigma\left(W^{(1)^T} x\right) = \alpha h_1(x)$
Assume for n<L, $\tilde{h}_n(x) = \alpha^n h_n(x)$
Induction step:
By definition, $\tilde{h_L}(x) = \sigma\left(\alpha W^{(L)^T} \tilde{h}_{L-1}(x)\right) \underbrace{=}_{induction\, hypothesis} \sigma\left(\alpha W^{(L)^T} \alpha^{L-1} h_{L-1}(x)\right) =$

$\underbrace{=}_{\alpha\, positive-homogeneous} \alpha^L \sigma\left(W^{(L)^T} h_{L-1}(x)\right) = \alpha^L h_L(x)$ □

By definition, $F_{\alpha\theta}(x) = W^{(L)^T} \tilde{h}_{L-1}(x) = W^{(L)^T} \alpha^{L-1} h_{L-1}(x) = \alpha^{L-1} F_\theta(x)$
Then $C = \alpha^{L-1}$.

2.b.

$$c_{y_i} = \frac{\exp\left((F_{\alpha\theta}(x))_i\right)}{\sum\limits_{j=1}^{k} \exp\left((F_{\alpha\theta}(x))_j\right)} = \frac{\exp\left(\alpha^L (F_\theta(x))_i\right)}{\sum\limits_{j=1}^{k} \exp\left(\alpha^L (F_\theta(x))_j\right)} \underset{\alpha\to 0}{\to} \frac{1}{k}$$

Then the distribution is uniform.

2.c.
<u>Case 1</u> $(F_\theta(x))_i > (F_\theta(x))_j, \forall j \in \{m_1, ..., m_{k-s}\}$ and $(F_\theta(x))_i = (F_\theta(x))_j, \forall j \in \{m_{k-s+1}, ..., m_k\}$, $(\{m_1, ..., m_k\} = \{1, ..., k\})$

$$c_{y_i} = \frac{\exp\left((F_{\alpha\theta}(x))_i\right)}{\sum\limits_{j=1}^{k} \exp\left((F_{\alpha\theta}(x))_j\right)} = \frac{\exp\left(\alpha^L (F_\theta(x))_i\right)}{\sum\limits_{j=1}^{k} \exp\left(\alpha^L (F_\theta(x))_j\right)} = \frac{1}{\sum\limits_{j=1}^{k} \exp\left(\alpha^L \left((F_\theta(x))_j - (F_\theta(x))_i\right)\right)} = ...$$

$$... = \frac{1}{s + \underbrace{\sum\limits_{j\in\{m_1,...,m_{k-s}\}} \exp\left(\alpha^L \left((F_\theta(x))_j - (F_\theta(x))_i\right)\right)}_{\underset{\alpha\to\infty}{\to} 0}} \underset{\alpha\to\infty}{\to} \frac{1}{s}$$

($s \geq 1$ because exists at least one maximum value then limit is well-defined)

<u>Case 2</u> $\exists 1 \leq j \leq k \; s.t. \quad (F_\theta(x))_i < (F_\theta(x))_j$

$$c_{y_i} = \frac{1}{1 + \underbrace{\sum\limits_{j\neq i} \exp\left(\alpha^L \left((F_\theta(x))_j - (F_\theta(x))_i\right)\right)}_{\underset{\alpha\to\infty}{\to} \infty}} \underset{\alpha\to\infty}{\to} 0$$

Then distribution is uniform on the $s \leq k$ entries with maximum value and zero otherwise.

i.e. $\left(0, ..., \underbrace{\frac{1}{s}}_{m}, ..., 0\right)$ where $m \in argmax_{1 \leq i \leq k} \left(\left(F_{\theta}\left(x\right)\right)_{i} \right)$.