Dry Part

Question 1

We would like to prove that $\epsilon_t = \sum_{i=1}^N D_i^{t+1} \cdot 1_{h_t(x_i) \neq y_i} = \frac{1}{2}$ for h_t the chosen weak classifier.

Recall our definitions from the lecture:

$$D_i^{(t+1)} = \frac{D_i^{(t)} \cdot exp(-w_t y_i h_t(x_i))}{\sum_{j=1}^N D_j^{(t)} \cdot exp(-w_t y_i h_t(x_j))}$$

$$w_t = \frac{1}{2} log(\frac{1}{\epsilon_t} - 1)$$

We will later on use the following statement:

$$exp(-2w_t) = exp(-2 \cdot \frac{1}{2} log(\frac{1}{\epsilon_t} - 1)) = \frac{\epsilon_t}{1-\epsilon_t}$$

For ease of use, we shall denote the set of indices of the examples which where classfied wrongly as A, and A^C for the examples which where classified correctly.
Let's massage the target mathematical definition for the error:

$$\epsilon_t = \sum_{i=1}^N D_i^{t+1} \cdot 1_{h_t(x_i) \neq y_i} =$$

$$\sum_{i \in A} D_i^{t+1} \cdot 1 + \sum_{i \in A^C} D_i^{t+1} \cdot 0 =$$

We can cancel out the correct samples as they contribute 0 to the error. Then we can plug in our expression for D_i^{t+1}:

$$\sum_{i \in A} D_i^{t+1} =$$

$$\sum_{i \in A} \frac{D_i^{(t)} \cdot exp(-w_t y_i h_t(x_i))}{\sum_{j=1}^N D_j^{(t)} \cdot exp(-w_t y_i h_t(x_j))} =$$

Notice how $y_i h_t(x_i) = -1$ for every wrong sample, simplifyig our expression:

$$\sum_{i \in A} \frac{D_i^{(t)} \cdot exp(-w_t \cdot (-1))}{\sum_{j=1}^N D_j^{(t)} \cdot exp(-w_t y_i h_t(x_j))} =$$

$$\sum_{i \in A} \frac{D_i^{(t)} \cdot exp(w_t)}{\sum_{j=1}^N D_j^{(t)} \cdot exp(-w_t y_i h_t(x_j))} =$$

We can now seperate the sum in the denominator to two parts - the wrongly classfied samples and the correct ones, and treat the expressions similarly to before:

$$\frac{\sum_{i \in A} D_i^{(t)} \cdot exp(w_t)}{\sum_{j \in A} D_j^{(t)} \cdot exp(-w_t y_i h_t(x_j)) + \sum_{j \in A^C} D_j^{(t)} \cdot exp(-w_t y_i h_t(x_j))} =$$

$$\frac{\sum_{i \in A} D_i^{(t)} \cdot exp(w_t)}{\sum_{j \in A} D_j^{(t)} \cdot exp(w_t) + \sum_{j \in A^C} D_j^{(t)} \cdot exp(-w_t)} =$$

$$\frac{\sum_{i \in A} D_i^{(t)}}{\sum_{j \in A} D_j^{(t)} + \sum_{j \in A^C} D_j^{(t)} \cdot exp(-2w_t)} =$$

Finally, we managed to get to the given expression we needed, now we just need to plug in our calculation of $exp(-2w_t)$:

$$\frac{\epsilon_t}{\epsilon_t + (1-\epsilon_t)\cdot exp(-2w_t)} =$$
$$\frac{\epsilon_t}{\epsilon_t + (1-\epsilon_t)\cdot \frac{\epsilon_t}{1-\epsilon_t}} =$$
$$\frac{\epsilon_t}{\epsilon_t + \epsilon_t} =$$
$$\frac{1}{2}$$

Question 2

2.1

We shall show that $F_{\alpha\Theta}(x) = c \cdot F_{\Theta}(x)$ where $c = \alpha^L$ for all $L > 1$. We proove this using induction over L.

Base: For $L = 2$ (A nueral network must be with at least 2 layers).

$$F_{\alpha\Theta}(x) = \alpha {W^{(2)}}^T \cdot h^{(1)}_{\alpha\Theta}(x) =$$
$$\alpha W^{(2)} \cdot \sigma(\alpha {W^{(1)}}^T x) =$$
$$\alpha^2 W^{(2)} \cdot \sigma({W^{(1)}}^T x) =$$
$$\alpha^2 W^{(2)} \cdot h^{(1)}_{\Theta}(x) =$$
$$\alpha^2 F_{\Theta}(x)$$

Step: We assume that $F_{\alpha\Theta}(x) = \alpha^L \cdot F_{\Theta}(x)$ for all $L \leq k$ for some $k \in \mathbb{N}$. Therfore for $L = k+1$ we can show that:

$$F_{\alpha\Theta}(x) = \alpha {W^{(k+1)}}^T \cdot h^{(k)}_{\alpha\Theta}(x) =$$
$$\alpha {W^{(k+1)}}^T \cdot \sigma(\alpha {W^{(k)}}^T \cdot h^{(k-1)}_{\alpha\Theta}(x)) =$$
$$\alpha {W^{(k+1)}}^T \cdot \sigma(\alpha^k {W^{(k)}}^T \cdot h^{(k-1)}_{\Theta}(x)) =$$
$$\alpha^{k+1} {W^{(k+1)}}^T \cdot \sigma({W^{(k)}}^T \cdot h^{(k-1)}_{\Theta}(x)) =$$
$$\alpha^{k+1} {W^{(k+1)}}^T \cdot h^{(k)}_{\Theta}(x) =$$
$$\alpha^{k+1} F_{\Theta}(x)$$

2.2