2) Lemma 3.3

$$\left\| \frac{d}{ds} w_r(s) \right\|_2 = \left\| \sum_{i=1}^{n} \underbrace{(y_i - u_i(s))}_{\text{scalar}} \frac{1}{\sqrt{m}} a_r \underbrace{x_i}_{\text{Vector}} \underbrace{\mathbb{1}\{w_r^T(s) x_i \geq 0\}}_{\text{Scalar.}} \right\|_2$$

using the property $\left\| \sum_{i=1}^{n} v_i \right\|_2 \leq \sum_{i=1}^{n} \| v_i \|_2$

$$\leq \sum_{i=1}^{n} \left\| \underbrace{(y_i - u_i(s))}_{\text{scaler}} \underbrace{\frac{a_r}{\sqrt{m}}}_{\text{scalar.}} \cdot x_i \underbrace{\mathbb{1}\{w_r^T(s) x_i \geq 0\}}_{\text{Scalar}} \right\|$$

using Property $\| cv \|_2 = |c| \| v \|_2$ (c is a scalar)

$$\leq \sum_{i=1}^{n} \left| (y_i - u_i(s)) \cdot \frac{a_r}{\sqrt{m}} \mathbb{1}\{w_r^T(s) x_i \geq 0\} \right| \| x_i \|_2$$

$\boxed{\text{using } |abc| = |a||b||c|}$

$\boxed{\begin{array}{c} \| x_i \|_2 = 1 \\ \text{Theorem assumption} \end{array}}$

$$\leq \sum_{i=1}^{n} \frac{1}{\sqrt{m}} | y_i - u_i(s) | \underbrace{| a_r |}_{\{0, 1\}} \underbrace{| \mathbb{1}\{w_r^T(s) x_i \geq 0\} |}_{\in [0, 1]}$$

$$\Rightarrow \leq \sum_{i=1}^{n} \frac{1}{\sqrt{m}} | y_i - u_i(s) |$$

(ii)

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{n} |y_i - u_i(s)| \leq \frac{\sqrt{n}}{\sqrt{m}} \|\underline{y} - \underline{u}(s)\|_2$$

$$\Downarrow$$

$$\frac{1}{\sqrt{m}} |\underline{y} - \underline{u}(s)|_1 \leq \frac{\sqrt{n}}{\sqrt{m}} \|\underline{y} - \underline{u}(s)\|_2 .$$

(to prove)

Consider : $|\underline{x}|_1$ ($\underline{x}$ is any vector)

$= a^T b$ where $a_i = 1$ $b_i = |x_i|$

$\leq \|a\|_2 \|b\|_2$ ($\cos\theta \in [-1, 1]$)

$\leq \sqrt{n} \cdot \|x\|_2$ (where n is dimension of the vector).

from this argument, ~~that~~ treat $\underline{x} = (\underline{y} - \underline{u}(s)) \frac{1}{\sqrt{m}}$.

$$\Rightarrow \frac{1}{\sqrt{m}} |\underline{y} - \underline{u}(s)| \leq \frac{\sqrt{n}}{\sqrt{m}} \|\underline{y} - \underline{u}(s)\|_2$$

3) Theorem 3.3

$$\frac{dw_r(t)}{dt} = -\frac{\partial L(W(t), a(t))}{\partial W_r(t)} \quad — (1)$$

$$\frac{da_r(t)}{dt} = \frac{-\partial L(W(t), a(t))}{\cancel{\partial W_r(t)}\ \partial a_r(t)} \quad — (2)$$

Consider $$\sum_{r=1}^{m} \frac{du_i(t)}{da_r(t)} \cdot \left(\frac{da_r(t)}{dt.}\right) \quad — (3)$$

this can be replaced by Equation (2)

Consider $$\frac{da_r(t)}{dt} = -\frac{\partial L(W(t), a(t))}{\partial a_r(t)} \quad — (4)$$

$$L(W(t), a(t)) = \sum_{i=1}^{n} \frac{1}{2}\left(f(W^{(t)}, a(t), x_i) - y_i\right)^2$$

and $$f(W(t), a(t), x_i) = \frac{1}{\sqrt{m}} \sum_{r=1}^{m} a_r^{(t)} \sigma(w_r^T(t) \cdot x_i)$$

$$\Rightarrow \frac{\partial L(W(t), a(t))}{\partial a_r(t)} = \sum_{i=1}^{n} \left(f(W(t), a(t), x_i) - y_i\right). \frac{\partial}{\partial a_r(t)} \cdot \left(f(W(t), a(t), x_i) - y_i\right)$$

$$\Rightarrow \frac{\partial L(w(t), a(t))}{\partial a_r(t)} = \sum_{i=1}^{n} \left(f(W(t), a(t), x_i) - y_i \right) \cdot \frac{\partial}{\partial a_r(t)} \cdot f\left(w(t), a(t), x_i\right)$$

$$\boxed{\begin{aligned} f\left(w(t), a(t), x_i\right) &= \frac{1}{\sqrt{m}} \sum_{r=1}^{m} a_r(t) \cdot \sigma\left(w_r^T(t) \cdot x_i\right) \\ \frac{\partial f}{\partial a_r(t)} &= \frac{1}{\sqrt{m}} \cdot \sigma\left(w_r^T(t) \cdot x_i\right) \end{aligned}}$$

$$\Rightarrow \frac{\partial L(w(t), a(t))}{\partial a_r(t)} = \sum_{i=1}^{n} f\left(w(t), a(t), x_i) - y_i\right) \cdot \frac{1}{\sqrt{m}} \sigma\left(w_r^T(t), x_i\right)$$

$$\sum_{r=1}^{m} \frac{du_i(t)}{da_r(t)} \cdot \underbrace{\left(\frac{da_r(t)}{dt}\right)}_{= -(*)}$$

$$- \sum_{r=1}^{m} \frac{1}{\sqrt{m}} \sigma\left(w_r^T(t) \cdot x_i\right) \cdot \sum_{j=1}^{n} \left(\left(u_j(t) - y_j\right) \cdot \frac{1}{\sqrt{m}} \sigma\left(w_r^T(t), x_j\right)\right)$$

$$= (-) \frac{1}{m} \sum_{r=1}^{m} \sum_{j=1}^{n} \left(u_j(t) - y_j\right) \cdot \sigma\left(w_r^T(t) \cdot x_i\right) \sigma\left(w_r^T(t), x_j\right)$$

define $G_{ij}(t) = \frac{1}{m} \sigma(w_r^T(t)\, x_i)\, \sigma(w_r^T(t), x_j)$

$$\Rightarrow \sum_{r=1}^{m} \frac{du_i(t)}{da_r} \cdot \frac{da_r(t)}{dt} = \sum_{r=1}^{m} \sum_{j=1}^{n} (y_j - u_j(t)) \cdot G_{ij}(t)$$

2.
$$-2(\underline{y} - \underline{u}(t))^T (H(t) + G(t))(\underline{y} - \underline{u}(t))$$

$$= -2(\underline{y} - \underline{u}(t))^T H(t)(\underline{y} - \underline{u}(t)) + \left(-2(\underline{y} - \underline{u}(t))^T G(t)(\underline{y} - \underline{u}(t))\right)$$

$x^T G(t)\, x \geq 0 \quad \forall x.$ | Since $G(t)$ is Symmetric, and positive Semidefinite)

$$\Rightarrow -2(\underline{y} - \underline{u}(t))^T G(t)(\underline{y} - \underline{u}(t)) \leq 0.$$

$$\Rightarrow \frac{d}{dt} \| y - u(t) \|_2^2 \leq -2(\underline{y} - \underline{u}(t))^T H(t)(\underline{y} - \underline{u}(t))$$