9. Using a new family of sigmoid hypotheses, $\tilde{h}(\vec{x}) = \frac{1}{2}\left(\frac{\vec{w}^T\vec{x}}{\sqrt{1+(\vec{w}^T\vec{x})^2}}+1\right)$

We can write

$$\tilde{h}(\vec{x}) = \theta(\vec{w}^T\vec{x}), \quad \text{where } \theta(s) = \frac{1}{2}\left(\frac{s}{\sqrt{1+s^2}}+1\right)$$

Also, we will have likelihood:

$$P(y|\vec{x}) = \begin{cases} \tilde{h}(\vec{x}) & \text{for } y=+1 \\ 1-\tilde{h}(\vec{x}) & \text{for } y=-1 \end{cases}$$

substitute $\tilde{h}(\vec{x})$ by $\theta(\vec{w}^T\vec{x})$:

$$P(y|\vec{x}) = \begin{cases} \theta(\vec{w}^T\vec{x}) & \text{for } y=+1 \\ 1-\theta(\vec{w}^T\vec{x}) & \text{for } y=-1 \end{cases}$$

We can verify that:

$$\boxed{1-\theta(s)} = 1-\frac{1}{2}\left(\frac{s}{\sqrt{1+s^2}}+1\right) = 1-\frac{s}{2\sqrt{1+s^2}}-\frac{1}{2} = \frac{1}{2}-\frac{s}{2\sqrt{1+s^2}}$$

$$= \frac{1}{2}\left(\frac{-s}{\sqrt{1+(-s)^2}}+1\right) = \boxed{\theta(-s)}$$

Thus,

$$P(y|\vec{x}) = \begin{cases} \theta(\vec{w}^T\vec{x}) & \text{for } y=+1 \\ \theta(-\vec{w}^T\vec{x}) & \text{for } y=-1 \end{cases}$$

$$\Rightarrow P(y|\vec{x}) = \theta(y\cdot\vec{w}^T\vec{x})$$

We can the derive the log likelihood:

$$\ln \prod_{i=1}^{N} P(y_n | \vec{x}_n) = \sum_{i=1}^{N} \ln(P(y_n | \vec{x}_n))$$

$$= \sum_{i=1}^{N} \ln(\theta(y \cdot \vec{w}^T \vec{x}))$$

To find the maximum log likelihood is equivalent to find the minimum cross entropy error;

$$\max_{\vec{w}} \sum_{i=1}^{N} \ln(\theta(y \cdot \vec{w}^T \vec{x})) \Rightarrow \min_{\vec{w}} -\sum_{i=1}^{N} \ln(\theta(y \cdot \vec{w}^T \vec{x}))$$

By our definition of $\theta(s) = \frac{1}{2}\left(\frac{s}{\sqrt{1+s^2}} + 1\right)$,

$$\ln(\theta(y\vec{w}^T\vec{x})) = \ln\left(\frac{1}{2}\left(\frac{y\vec{w}^T\vec{x}}{\sqrt{1+(y\vec{w}^T\vec{x})^2}} + 1\right)\right)$$

$$= \ln \frac{y\vec{w}^T\vec{x} + \sqrt{1+(y\vec{w}^T\vec{x})^2}}{2\sqrt{1+(y\vec{w}^T\vec{x})^2}}$$

$\therefore$ the $\tilde{E}_{in}$ we want to minimize is:

$$\tilde{E}_{in}(\vec{w}) = \frac{1}{N}\left[-\sum_{i=1}^{N} \ln(\theta(y_i \vec{w}^T \vec{x}_i))\right] = \frac{1}{N} \sum_{i=1}^{N} \ln \frac{2\sqrt{1+(y\vec{w}^T\vec{x})^2}}{y\vec{w}^T\vec{x} + \sqrt{1+(y\vec{w}^T\vec{x})^2}}$$

To minimize $\tilde{E}_{in}(\vec{w})$, we need to find the place where $\nabla \tilde{E}_{in}(\vec{w}) = 0$

First we calculate $\nabla \ln(\theta(\ ))$

$$\nabla \ln(\theta(y_i \vec{w}^T \vec{x}_i)) = \frac{1}{\theta(y_i \vec{w}^T \vec{x}_i)} \cdot \nabla \theta(y_i \vec{w}^T \vec{x}_i)$$

calculate $\nabla \theta(s)$:

$$((1+s^2)^{\frac{1}{2}})' \cdot s = \frac{1}{2}(1+s^2)^{-\frac{1}{2}} \cdot 2s \cdot s = \frac{s^2}{\sqrt{1+s^2}}$$

$$\frac{d}{ds}\left(\frac{s}{\sqrt{1+s^2}}\right) = \frac{\sqrt{1+s^2} - \frac{s^2}{\sqrt{1+s^2}}}{1+s^2} = \frac{\frac{(1+s^2)-s^2}{\sqrt{1+s^2}}}{1+s^2} = \frac{\frac{1}{\sqrt{1+s^2}}}{1+s^2} = (1+s^2)^{-\frac{3}{2}}$$

$$\therefore \frac{d}{ds}\theta(s) = \frac{1}{2} \cdot (1+s^2)^{-\frac{3}{2}}$$

$\because y_i^2 = 1$

$$\Rightarrow \nabla \theta(y_i \vec{w}^T \vec{x}_i) = \frac{1}{2} \frac{1}{[1+(y_i \vec{w}^T \vec{x}_i)^2]^{\frac{3}{2}}} y_i \vec{x}_i = \frac{y_i \vec{x}_i}{2[1+(\vec{w}^T \vec{x}_i)^2]^{\frac{3}{2}}}$$

$$\Rightarrow \nabla \ln(\theta(y_i \vec{w}^T \vec{x}_i)) = \frac{1}{\theta(y_i \vec{w}^T \vec{x}_i)} \cdot \boxed{\frac{y_i \vec{x}_i}{2[1+(\vec{w}^T \vec{x}_i)^2]^{\frac{3}{2}}}}$$

$$\Rightarrow \nabla \tilde{E}_{in}(\vec{w}) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2\theta(y_i \vec{w}^T \vec{x}_i)} \frac{y_i \vec{x}_i}{[1+(\vec{w}^T \vec{x}_i)^2]^{\frac{3}{2}}}$$

$$\theta(y_i \vec{w}^T \vec{x}_i) = \frac{1}{2}\left(\frac{y_i \vec{w}^T \vec{x}_i}{\sqrt{1+(y_i \vec{w}^T \vec{x}_i)^2}} + 1\right)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \frac{y_i \vec{x}_i}{\left(\frac{y_i \vec{w}^T \vec{x}_i}{\sqrt{1+(y_i \vec{w}^T \vec{x}_i)^2}} + 1\right)[1+(\vec{w}^T \vec{x}_i)^2]^{\frac{3}{2}}}$$

$\square$