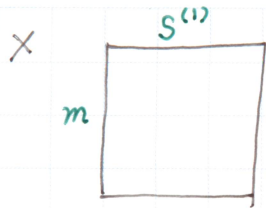
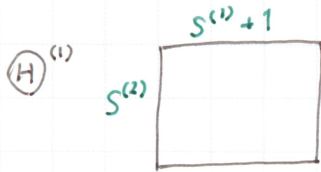


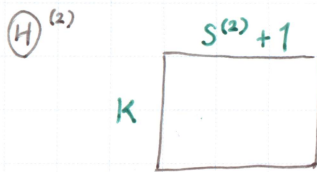
ex4 nn Cost Function (1) 初期状態



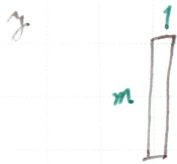
$S^{(1)}$ = input-layer-size = 400
 $S^{(2)}$ = hidden-layer-size = 25
 K = num-labels = 10
 m = datasetの数 = 5000



← Theta2-grad は同じ形

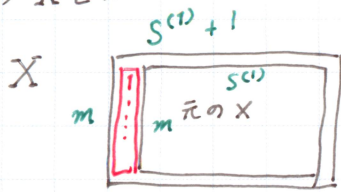


← Theta1-grad は同じ形



ex4 nnCostFunction (2) forward propagation

(2-1) Xを拡張



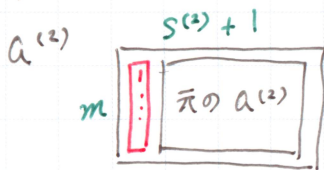
(2-2) z^(2)を求める

$$X \cdot (H^{(1)})^T = m \begin{matrix} S^{(1)}+1 \\ X \end{matrix} \cdot \begin{matrix} S^{(2)} \\ S^{(1)}+1 \\ (H^{(1)})^T \end{matrix} = m \begin{matrix} S^{(2)} \\ z^{(2)} \end{matrix}$$

(2-3) a^(2)を求める

$$\text{sigmoid} \left(m \begin{matrix} S^{(2)} \\ z^{(2)} \end{matrix} \right) = m \begin{matrix} S^{(2)} \\ a^{(2)} \end{matrix}$$

(2-4) a^(2)を拡張



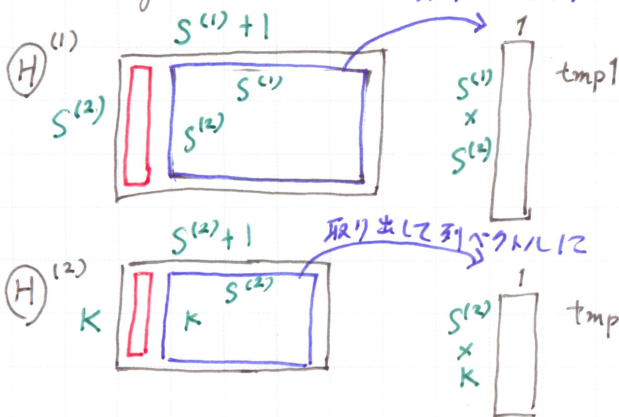
(2-5) z^(3)を求める

$$a^{(2)} \cdot (H^{(2)})^T = m \begin{matrix} S^{(2)}+1 \\ a^{(2)} \end{matrix} \cdot \begin{matrix} K \\ S^{(2)}+1 \\ (H^{(2)})^T \end{matrix} = m \begin{matrix} K \\ z^{(3)} \end{matrix}$$

(2-6) h = a^(3)を求める

$$\text{sigmoid} \left(m \begin{matrix} K \\ z^{(3)} \end{matrix} \right) = m \begin{matrix} K \\ a^{(3)} \end{matrix} = h$$

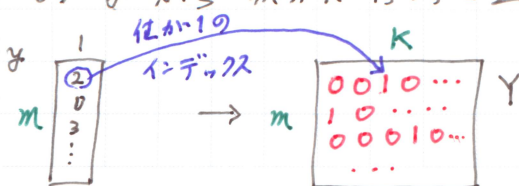
(2-7) regularization



$$\text{tmp1}^T \cdot \text{tmp1} = |\text{tmp1}|^2$$

$$\text{tmp2}^T \cdot \text{tmp2} = |\text{tmp2}|^2$$

(2-8) yから m x K 行列を生成する



ex4 nn Cost Function (3) forward propagation

(2-9) J を計算する

$$m \begin{matrix} \boxed{Y} \\ \text{K} \end{matrix} .* \log(m \begin{matrix} \boxed{h} \\ \text{K} \end{matrix}) + m \begin{matrix} \boxed{1-Y} \\ \text{K} \end{matrix} .* \log(m \begin{matrix} \boxed{1-h} \\ \text{K} \end{matrix}) \rightarrow m \begin{matrix} \boxed{} \\ \text{K} \end{matrix}$$

↑ 各項毎に

$$\text{sum}(\text{sum}(m \begin{matrix} \boxed{} \\ \text{K} \end{matrix})) = \text{sum}(1 \begin{matrix} \boxed{} \\ \text{K} \end{matrix}) = 1 \begin{matrix} \boxed{\text{和}} \\ \text{K} \end{matrix}$$

$$J = 1 \begin{matrix} \boxed{\text{和}} \\ \text{K} \end{matrix} / m + (|\text{tmp1}|^2 + |\text{tmp2}|^2) \times \frac{\lambda}{2m}$$

ex4 nnCostFunction (4) back propagation

(1) $\delta^{(3)}$ を計算する

$$m \begin{matrix} K \\ \hline \end{matrix} \begin{matrix} K \\ \hline \end{matrix} - m \begin{matrix} K \\ \hline \end{matrix} \begin{matrix} K \\ \hline \end{matrix} = m \begin{matrix} K \\ \hline \end{matrix} \begin{matrix} K \\ \hline \end{matrix} \delta^{(3)}$$

(2) $\delta^{(2)}$ を計算する

$$\delta^{(2)} = (\mathbf{H}^{(2)})^T \delta^{(3)} \cdot \mathbf{a}^{(2)} \cdot (1 - \mathbf{a}^{(2)}) \quad \leftarrow \text{元の説明の式}$$

mのループで処理する場合

$$m \begin{matrix} K \\ \hline \end{matrix} \delta^{(3)} \cdot K \begin{matrix} S^{(2)}+1 \\ \hline \end{matrix} \mathbf{H}^{(2)} \cdot m \begin{matrix} S^{(2)}+1 \\ \hline \end{matrix} \mathbf{a}^{(2)} \cdot m \begin{matrix} S^{(2)}+1 \\ \hline \end{matrix} (1 - \mathbf{a}^{(2)})$$

$S^{(2)}+1$

$$m \begin{matrix} S^{(2)}+1 \\ \hline \end{matrix} \delta^{(2)}$$

(注) sigmoidGradient()関数には $z^{(2)}$ を渡す仕様となっている。
 $\text{sigmoid}(z) = \frac{1}{1+e^{-z}} \approx 1$ なので $z^{(2)}$ の先頭に 1 を渡せばよいようだ。本当はもっと大きい値 (100とか) の方がよいと思うが...

(3) $\Delta^{(2)}$ を計算する

$$\Delta^{(2)} = \Delta^{(2)} + \delta^{(3)} \cdot (\mathbf{a}^{(2)})^T \quad \leftarrow \text{元の説明の式}$$

mのループで処理する場合

$$K \begin{matrix} m \\ \hline \end{matrix} \delta^{(3)T} \cdot m \begin{matrix} S^{(2)}+1 \\ \hline \end{matrix} \mathbf{a}^{(2)} \rightarrow K \begin{matrix} S^{(2)}+1 \\ \hline \end{matrix} \Delta^{(2)}$$

$\leftarrow \frac{1}{m}$ を乗算して Theta2-grad

(4) $\Delta^{(1)}$ を計算する

$$\Delta^{(1)} = \Delta^{(1)} + \delta^{(2)} \cdot (\mathbf{a}^{(1)})^T \quad \leftarrow \text{元の説明の式}$$

mのループで処理する場合

$S^{(2)}$ の先頭を取り除くべし

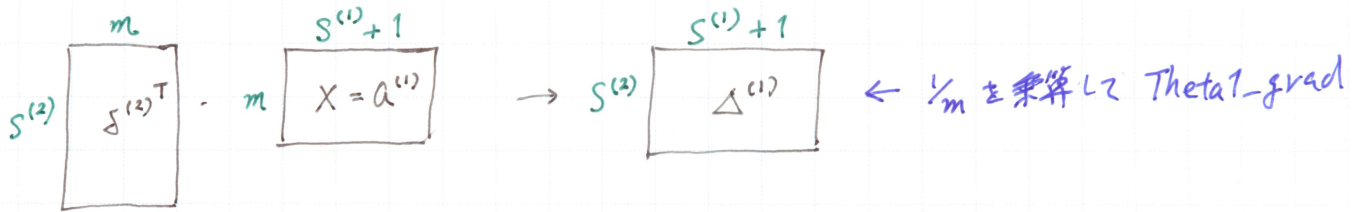
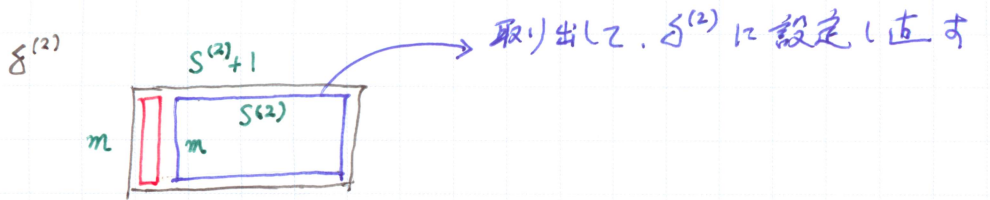
$$m \begin{matrix} S^{(2)} \\ \hline \end{matrix} \delta^{(2)} \cdot m \begin{matrix} S^{(1)}+1 \\ \hline \end{matrix} \mathbf{a}^{(1)} \rightarrow m \begin{matrix} S^{(1)}+1 \\ \hline \end{matrix} \Delta^{(1)}$$

$S^{(1)}+1$

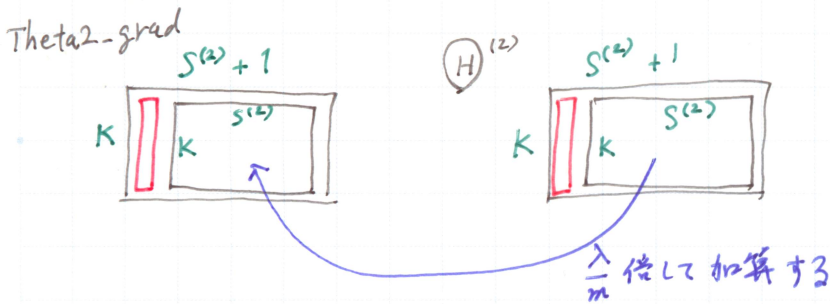
$$m \begin{matrix} S^{(1)}+1 \\ \hline \end{matrix} \mathbf{H}^{(1)}$$

$\Delta^{(1)}$ は $\mathbf{H}^{(1)}$ と同じ次元である必要がある

ex4 nn CostFunction (5) back propagation



(5) regularized $\frac{\partial}{\partial \theta^{(2)}} J(\theta)$ を計算する



(6) regularized $\frac{\partial}{\partial \omega} J(\theta)$ を計算する

