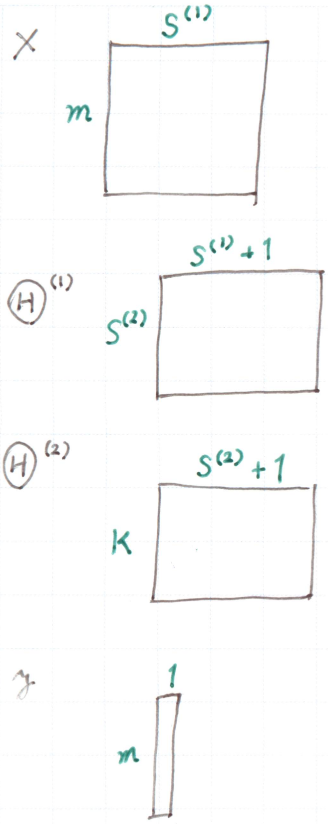
ex4 nnCostFunction (1) 初期状態

$s^{(1)}$ = input_layer_size = 400
$s^{(2)}$ = hidden_layer_size = 25
K = num_labels = 10
m = datasetの数 = 5000

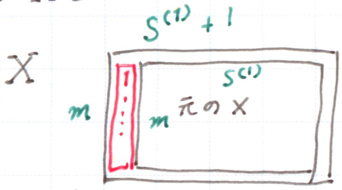


← Theta2_grad は同じ形

← Theta1_grad は同じ形

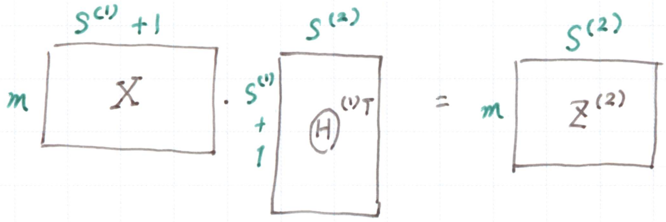
ex4 nnCostFunction (2) forward propergation

(2-1) Xを拡張

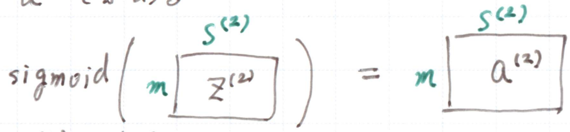


(2-2) $z^{(2)}$を求める

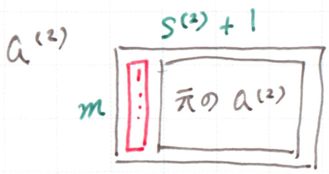
$X \cdot \Theta^{(1)T}$ =



(2-3) $a^{(2)}$を求める

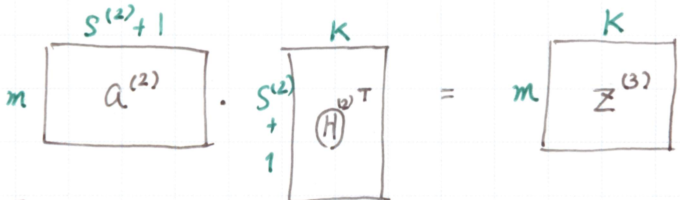


(2-4) $a^{(2)}$を拡張

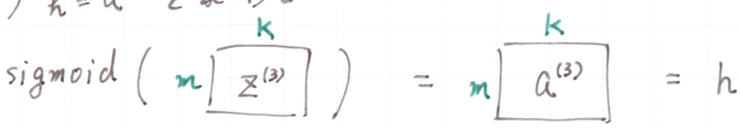


(2-5) $z^{(3)}$を求める

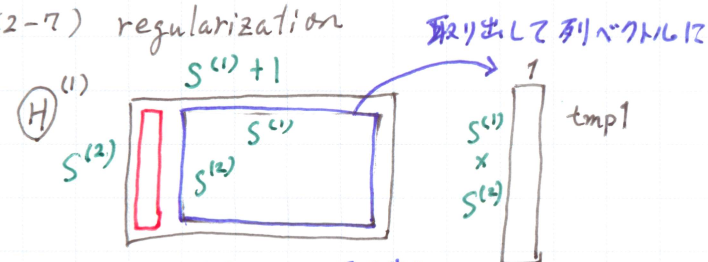
$a^{(2)} \cdot \Theta^{(2)T}$ =



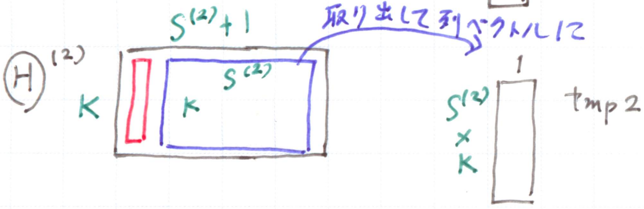
(2-6) $h = a^{(3)}$を求める



(2-7) regularization

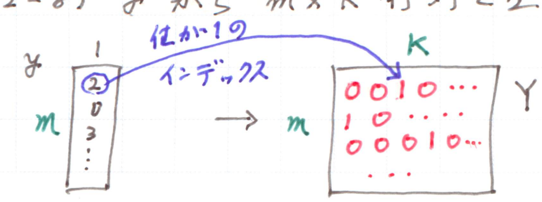


$tmp1^T \cdot tmp1 = |tmp1|^2$



$tmp2^T \cdot tmp2 = |tmp2|^2$

(2-8) y から m×K 行列を生成する



ex4 nnCostFunction (3) forward propergation

(2-9) J を計算する

$$ {}_{m}\overset{K}{\boxed{Y}} \;.*\; \log\left({}_{m}\overset{K}{\boxed{h}}\right) + {}_{m}\overset{K}{\boxed{1-Y}} \;.*\; \log\left({}_{m}\overset{K}{\boxed{1-h}}\right) \rightarrow {}_{m}\overset{K}{\boxed{}} $$

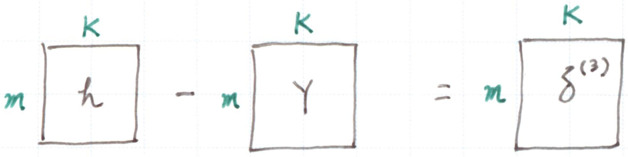
各項毎に

$$ \mathrm{sum}\left(\mathrm{sum}\left({}_{m}\overset{K}{\boxed{}}\right)\right) = \mathrm{sum}\left({}_{1}\overset{K}{\boxed{}}\right) = {}_{1}\overset{1}{\boxed{\text{和}}} $$

$$ J = {}_{1}\overset{1}{\boxed{\text{和}}} / m + \left(|tmp1|^2 + |tmp2|^2\right) \times \frac{\lambda}{2m} $$

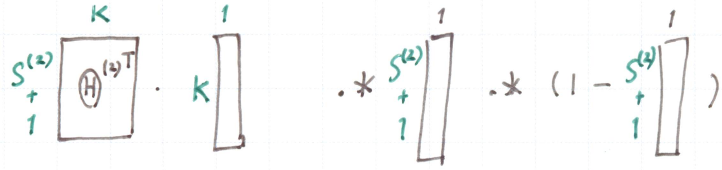
ex4 nnCostFunction (4) back propergation

(1) $\delta^{(3)}$ を計算する

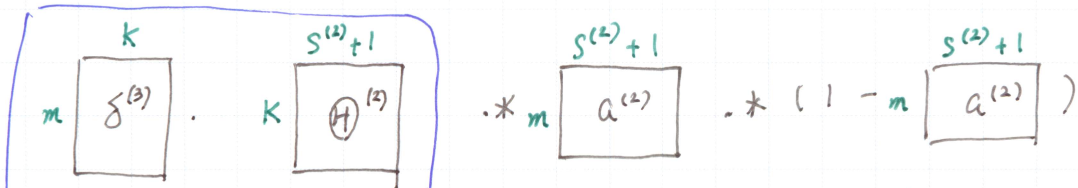


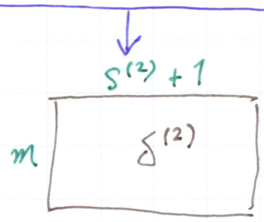
(2) $\delta^{(2)}$ を計算する

$\delta^{(2)} = (\Theta^{(2)})^T \delta^{(3)} .* a^{(2)} .* (1 - a^{(2)})$ ← 元の説明の式

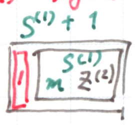


mのループで処理する場合





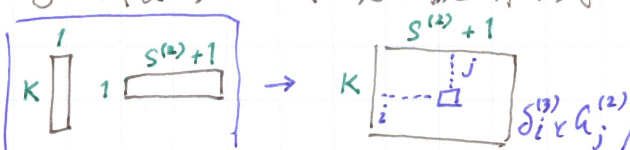
(注) sigmoidGraient( ) 関数には $z^{(2)}$ を渡す仕様となっている。



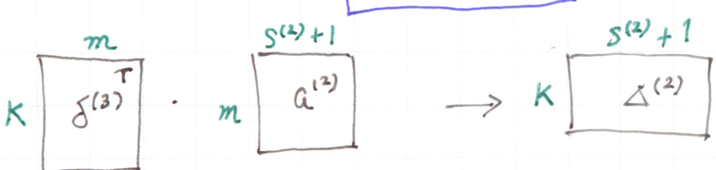
$\mathrm{sigmoid}(1) = \frac{1}{1+e^{-1}} \simeq 1$ なので $z^{(2)}$ の先頭に 1 を渡せばよいようだ。本当はもっと大きい値 (100とか) の方がよいと思うが…

(3) $\Delta^{(2)}$ を計算する

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



mのループで処理する場合

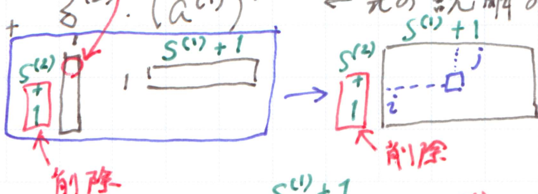


← 1/m を乗算して Theta2_grad

(4) $\Delta^{(1)}$ を計算する

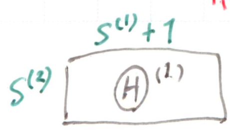
$\delta^{(2)}$の先頭を取り除くべし

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



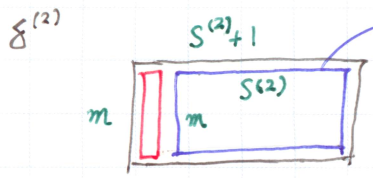
$\delta^{(2)}_i$ と $a^{(1)}_j$

mのループで処理する場合

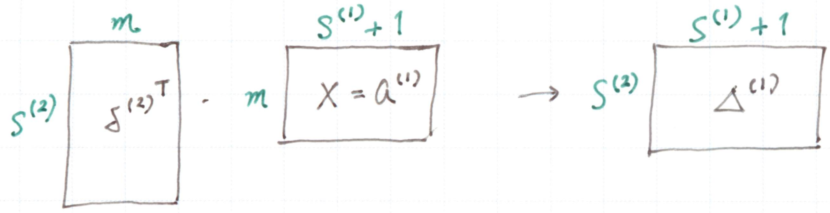


$\Delta^{(1)}$ は $\Theta^{(1)}$ と同じ次元である必要がある

ex4 nnCostFunction (5) back propergation

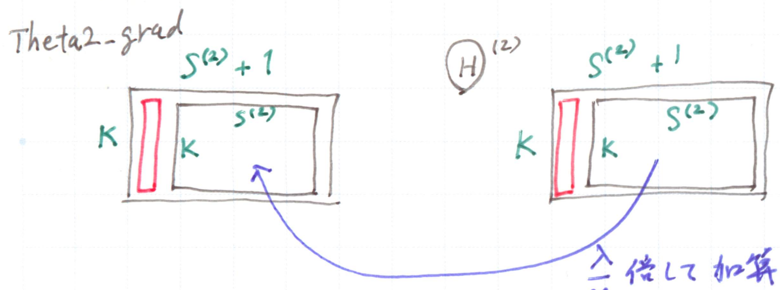


取り出して、$\delta^{(2)}$ に設定し直す。



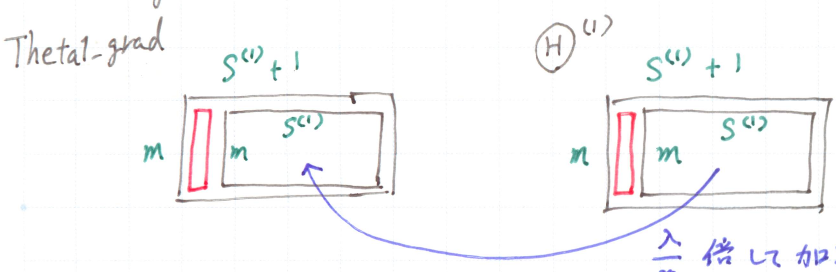
← $1/m$ を乗算して Theta1_grad

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



$\frac{\lambda}{m}$ 倍して加算する

(6) regularized な $\frac{\partial}{\partial \Theta^{(1)}} J(\Theta)$ を計算する



$\frac{\lambda}{m}$ 倍して加算する