

Week 5 Quiz (1)

[1] $\Delta_{ij}^{(2)} := \Delta_{ij}^{(2)} + \delta_i^{(3)} * (a^{(2)})_j$

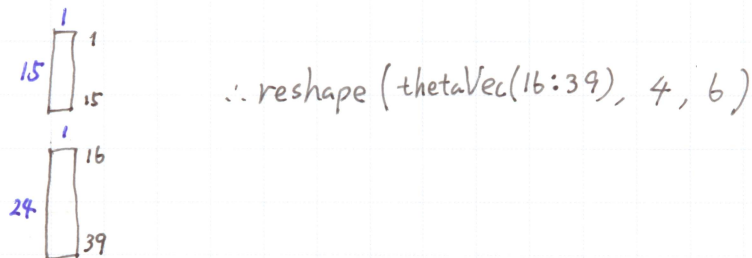


$\therefore \Delta^{(2)} := \Delta^{(2)} + \delta^{(3)} * (a^{(2)})^T$

[2]



$\text{thetaVec} = [\text{Theta1}(:); \text{Theta2}(:)]$



[3] $J(\theta) = 2\theta^3 + 2$ $\begin{cases} x = \theta + \epsilon \\ y = \theta - \epsilon \end{cases}$ $\begin{cases} x - y = 2\epsilon \\ x + y = 2\theta \end{cases}$ $xy = \theta^2 - \epsilon^2$

$$\begin{aligned} J(\theta + \epsilon) - J(\theta - \epsilon) &= 2(x^3 - y^3) \\ &= 2(x - y)(x^2 + xy + y^2) \\ &= 2(x - y)((x + y)^2 - xy) \\ &= 2 \cdot 2\epsilon (4\theta^2 - \theta^2 + \epsilon^2) \\ &= 2 \cdot 2\epsilon (3\theta^2 + \epsilon^2) \end{aligned}$$

$$\begin{aligned} \frac{J(\theta + \epsilon) - J(\theta - \epsilon)}{2\epsilon} &= 2(3\theta^2 + \epsilon^2) \\ &= 6\theta^2 + 2\epsilon^2 \\ &= 6.0002 \quad (\because \theta = 1, \epsilon = 0.01) \end{aligned}$$

Week 5 Quiz (2)

[4] $\left. \begin{array}{l} \lambda \text{ を小さく} \\ \text{MaxIter を大きく} \end{array} \right\} \rightarrow \text{overfitting となる}$

① 大きすぎる λ でも害がない。 λ を大きくすべきなのは単に数値の問題だから $\rightarrow X$

λ を大きくすると $\Delta^{(L)}$ が $(H)^{(L)}$ に比べて小さくなるので、^{1回の}back propagationで $(H)^{(L)}$ が変化しにくくなる

② トレーニングセットに対して overfitting する場合は λ を大きくせよ $\rightarrow O$

λ が小さ過ぎると、細かく $(H)^{(L)}$ を調整し過ぎてしまう

③ gradient-checking を使うと back propagation が正しく実装されていることを確認できる $\rightarrow O$

④ gradient descent を使っている時に gradient-checking は役に立つ。 $\rightarrow X$
しかし minunc のような優れた optimization は必要ない

Week 5 Quiz (3)

[5] ① gradient descent の neural network で, 初期値によって別の最適解 $\rightarrow \bigcirc$
に収束することがある

local minimum に収束することがある.

② gradient descent を用いてトレーニングする時, 有効なバグ手法は $J(H)$ が \llcorner 回しの度に減っていること (少なくとも増えていること) を $\rightarrow \bigcirc$
確認することだ

③ $H^{(1)} = \text{all } 0$ $H^{(2)} = \text{all } 1$ symmetry breaking に十分 $\rightarrow \times$

④ パラメータ全てを 0 ではなく 1 にする $=$ $\rightarrow \times$