DS 598 DEEP LEARNING

Sungjoon Park (BUID: U38522578)

Feb. 1st, 2024

Problem Set 3

1. (Problem 3.1)

   (a) If the activation function in equation 3.1 was linear, there will be only one mapping from input to output like below.

   $$\begin{aligned} y &= \phi_0 + \phi_1(\psi_0 + \psi_1 x) + \phi_2(\psi_0 + \psi_1 x) + +\phi_3(\psi_0 + \psi_1 x) \\ &= (\phi_0 + (\phi_1 + \phi_2 + \phi_3)\psi_0) + (\phi_1 + \phi_2 + \phi_3)\psi_1 x \\ &= a + bx \end{aligned}$$

   This is nothing but a linear equation composed of parameters with a pair of bias and slope where $a = \phi_0 + (\phi_1 + \phi_2 + \phi_3)\psi_0$ and $b = (\phi_1 + \phi_2 + \phi_3)\psi_1$.

   (b) If the activation function were removed, then the shallow network becomes a simple linear function like below.

   $$y = \phi_0 + (\phi_1 + \phi_2 + \phi_3)x$$

   Thus, there exists only one mapping.

2. (Problem 3.2) It can be determined by the figures in g, h and i.

Regions	1st	2nd	3rd	4th
Active	h_3	h_1 and h_3	h_1, h_2 and h_3	h_1 and h_2
Inactive	h_1 and h_2	h_2	None	h_3

* The regions are sorted from the left.

3. (Problem 3.3)
< Joints >
The positions where the three pre-activation functions cross zero in the Figure 3.3 become the three "joints" in the final output. From the left, their horizontal locations are like below.
1st joint: $x = -\frac{\theta_{10}}{\theta_{11}}$
2nd joint: $x = -\frac{\theta_{20}}{\theta_{21}}$
3rd joint: $x = -\frac{\theta_{30}}{\theta_{31}}$
Their vertical locations can be obtained by adding height at each horizontal location.
1st joint: $\phi_0 + (\theta_{30} - \frac{\theta_{10}}{\theta_{11}}\theta_{31})\phi_3$
2nd joint: $\phi_0 + (\theta_{10} - \frac{\theta_{20}}{\theta_{21}}\theta_{11})\phi_1 + (\theta_{30} - \frac{\theta_{20}}{\theta_{21}}\theta_{31})\phi_3$
3rd joint: $\phi_0 + (\theta_{10} - \frac{\theta_{30}}{\theta_{31}}\theta_{11})\phi_1 + (\theta_{20} - \frac{\theta_{30}}{\theta_{31}}\theta_{21})\phi_2$

< Linear Regressions >
The slopes of the four regressions are as follows.
1st slope: $\phi_3\theta_{31}$
2nd slope: $\phi_1\theta_{11} + \phi_3\theta_{31}$
3rd slope: $\phi_1\theta_{11} + \phi_2\theta_{21} + \phi_3\theta_{31}$
4th slope: $\phi_1\theta_{11} + \phi_2\theta_{21}$

By drawing on the positions of the three joints and four slopes above, the four regressions are derived as follows.
1st regression: $y = (\phi_0 + \phi_3\theta_{30}) + \phi_3\theta_{31}x$, where $x \in [0, -\frac{\theta_{10}}{\theta_{11}})$
2nd regression: $y = (\phi_0 + \phi_3\theta_{30} + \phi_1\theta_{10}) + (\phi_1\theta_{11} + \phi_3\theta_{31})x$, where $x \in [-\frac{\theta_{10}}{\theta_{11}}, -\frac{\theta_{20}}{\theta_{21}})$
3rd regression: $y = (\phi_0+\phi_2\theta_{20}+\phi_1\theta_{10}+\phi_3\theta_{30})+(\phi_1\theta_{11}+\phi_2\theta_{21}+\phi_3\theta_{31})x$, where $x \in [-\frac{\theta_{20}}{\theta_{21}}, -\frac{\theta_{30}}{\theta_{31}})$
4th regression: $y = (\phi_0 + \phi_1\theta_{10} + \phi_2\theta_{20}) + (\phi_1\theta_{11} + \phi_2\theta_{21})x$,where $x \in [-\frac{\theta_{30}}{\theta_{31}}, 1]$

4. (Problem 3.5) The property holds for $\alpha \in \mathbb{R}^+$.
If $z \geq 0$, then $\alpha z \geq 0$. Therefore, $\mathrm{ReLU}[\alpha z] = \alpha z = \alpha\mathrm{ReLU}[z]$.
Otherwise, that is, if $z < 0$, $\alpha z < 0$. Thus, $\mathrm{ReLU}[\alpha z] = 0 = \alpha\mathrm{ReLU}[z]$.
For this reason, $\mathrm{ReLU}[\alpha z] = \alpha\mathrm{ReLU}[z]$

5. (Problem 3.6)

   (a) Multiplying θ_{10} and θ_{11} by a positive real number, α, respectively does not impact the horizontal intercept of pre-activation function, that is, the x value at the first joint. This is because the original position $-\frac{\theta_{10}}{\theta_{11}}$ is the same as $-\frac{\alpha\theta_{10}}{\alpha\theta_{11}} = -\frac{\theta_{10}}{\theta_{11}}$. However, the pre-activation graph will be changed such that its slope and y-intercept are $\alpha\theta_{11}$

and $\alpha\theta_{10}$ respectively, while maintaining their signs. This change will be offset in the output layer by dividing ϕ_1 by a positive real number, α, because $\frac{\phi_1}{\alpha} \times (\alpha(\theta_{10} + \theta_{11}))x$ is the same as $\phi_1 \times (\theta_{10} + \theta_{11})x$ and the active area (support) is not changed.

(b) On the other hand, if the parameters θ_{10} and θ_{11} is multiplied by a negative real number, α, the pre-activation graph rotates around the horizontal axis maintaining the same horizontal intercept as before. The final output, however, is not changed because of the same reason in the case (a).
The pre-activation and post-activation lines (not weighted lines) in both cases (a) and (b) will have steeper slopes and farther y-intercepts from the origin if the absolute value of α is larger than 1. Otherwise, their slopes will be smaller and their y-intercepts will be closer to the origin.

6. (Problem 3.7) As is the case of the above problem, there can exist multiple sets of parameters that result in the same mapping from inputs to outputs. Therefore, the loss function of the equation 3.1 can have more than one sets of parameters that attain the same minimum loss.

7. (Problem 3.14)

   (a) $<$ From the input layer to the hidden layer $>$
   $h_1 = a(\theta_{10} + \theta_{11}x_1 + \theta_{12}x_2 + \theta_{13}x_3)$
   $h_2 = a(\theta_{20} + \theta_{21}x_1 + \theta_{22}x_2 + \theta_{23}x_3)$
   $h_3 = a(\theta_{30} + \theta_{31}x_1 + \theta_{32}x_2 + \theta_{33}x_3)$

   (b) $<$ From the hidden layer to the output layer $>$
   $y_1 = \phi_{10} + \phi_{11}h_1 + \phi_{12}h_2 + \phi_{13}h_3$
   $y_2 = \phi_{20} + \phi_{21}h_1 + \phi_{22}h_2 + \phi_{23}h_3$

8. (Problem 3.17)
   $<$ From the input layer to the hidden layer $>$
   $(D_i + 1) \times D$
   $<$ From the hidden layer to the output layer $>$
   $(D + 1) \times D_o$
   $<$ Total number of parameters $>$
   $((D_i + 1) \times D) + ((D + 1) \times D_o) = (D_i + D_o) \times D + (D + D_o)$

9. (Problem 3. 18)

$<$ The maximum number of regions in figure 3.8j $>$

$$\begin{aligned}\sum_{j=0}^{2}\binom{3}{j} &= \binom{3}{0}+\binom{3}{1}+\binom{3}{2}\\ &= 1+3+3\\ &= 7\end{aligned}$$

$<$ The maximum number of regions when $D = 5$ $>$

$$\begin{aligned}\sum_{j=0}^{2}\binom{5}{j} &= \binom{5}{0}+\binom{5}{1}+\binom{5}{2}\\ &= 1+5+10\\ &= 16\end{aligned}$$