



۱. (۵۰ نمره)

(آ) این جمله از بزرگ شدن بیش از حد ضرایب جلوگیری می کند و به دلیل کوچک کردن فضای ضرایب، احتمال overfit را کم می کند.

(ب)

$$\begin{aligned}\frac{\partial F(W)}{\partial W} &= 0 \\ 2\lambda W + 2X^T(XW - y) &= 0 \\ 2\lambda W + 2X^T XW &= 2X^T y \\ (\lambda I + X^T X)W &= X^T y \\ W &= (\lambda I + X^T X)^{-1} X^T y\end{aligned}$$

(ج)

$$\begin{aligned}w_2 &= \arg \min_w L(w) + \lambda \|w\|^2 \Rightarrow L(w_2) + \lambda \|w_2\|^2 \leq L(w_1) + \lambda \|w_1\|^2 \\ w_1 &= \arg \min_w L(w) \Rightarrow L(w_1) \leq L(w_2)\end{aligned}$$

از دو رابطه بالا می توان نتیجه گرفت:

$$\begin{aligned}L(w_1) + \lambda \|w_2\|^2 &\leq L(w_2) + \lambda \|w_1\|^2 \Rightarrow \lambda \|w_2\|^2 \leq \lambda \|w_1\|^2 \\ \|w_2\|^2 &\leq \|w_1\|^2 \Rightarrow \|w_2\| \leq \|w_1\|\end{aligned}$$

۲. (۴۰ نمره) اگر لایه k ام دارای h_k نورون و لایه $k+1$ دارای h_{k+1} نورون باشد آنگاه مقادیر لایه $k+1$ را می توان به صورت زیر نوشت:

$$x^{(k+1)} = W^{(k+1)} \begin{bmatrix} x_1^{(k)} \\ \vdots \\ x_{h_k}^{(k)} \\ 1 \end{bmatrix}$$

که در آن ماتریس $W^{(k+1)}$ یک ماتریس $(h_{k+1} + 1) \times (h_k + 1)$ که نشان دهنده ماتریس ضرایب بین نود های این دو لایه مخفی است (سطر آخر به صورت $[0 \ \dots \ 0 \ 1]$ است تا نود bias برای لایه بعدی تولید شود). پس می توان خروجی را به شکل زیر نوشت:

$$y = W^{(d+1)} W^{(d)} \dots W^{(1)} \tilde{x} = Z \tilde{x}$$

که $Z = W^{(d+1)} W^{(d)} \dots W^{(1)}$ و d تعداد لایه های مخفی و $\tilde{x} = [x_1 \ \dots \ x_n \ 1]$ است.

$$Z = Wx + b$$

$$\frac{\partial E}{\partial W} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial Z} \frac{\partial Z}{\partial w} = \frac{\partial E}{\partial y} f'(Z)x$$

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial Z} \frac{\partial Z}{\partial b} = \frac{\partial E}{\partial y} f'(Z)$$