Image Denoising - Homework 5

Ex 1.1 NAND logic

A	B	
0	0	1
0	1	1
1	0	1
1	1	0

Given a perceptron $f(A,B) = \begin{cases} 1 & \text{if } w_1 A + w_2 B + b > 0 \\ 0 & \text{o.w} \end{cases}$

We can take $\boxed{w_1 = w_2 = -1}$ and $\boxed{b = 2}$

In that case
$f(0,0): 2 > 0 \Rightarrow f(0,0) = 1$
$f(0,1): -1+2 = 1 > 0 \Rightarrow f(0,1) = 1$
$f(1,0): -1+2 = 1 > 0 \Rightarrow f(1,0) = 1$
$f(1,1): -1-1+2 = 0 \Rightarrow f(1,1) = 0$

Ex 1.2 We can represent a 4x4 DCT transform by using one convolutional kernel of size 4x4 (without bias term) for each element of the DCT basis.
Since the input is a grayscale image, we need 16 filters of size 4x4.

Ex 1.3 $f(x) = g(\overbrace{Wx+b}^{z})$ $W \in \mathbb{R}^{4\times 3}$; g sigmoid

$g(z) = \frac{1}{1+e^{-z}}$ $\quad g'(z) = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{(1+e^{-z})} \cdot \frac{e^{-z}}{(1+e^{-z})}$

$= g(z)(1-g(z))$, In the vectorial case $g(z)(\mathbb{1}-g(z))^T$

a) $\frac{\partial f_i}{\partial W_{ij}} = \frac{\partial z_i}{\partial W_{ij}} \frac{\partial g_i}{\partial z_i} = g(Wx+b)_i(1-g(Wx+b)_i)\, x_j$

b) $\frac{\partial f}{\partial b} = \frac{\partial z}{\partial b} \frac{\partial g}{\partial z} = g(Wx+b)(\mathbb{1}-g(Wx+b))^T$

c) $\frac{\partial f}{\partial x} = \frac{\partial z}{\partial x} \frac{\partial g}{\partial z} = g(Wx+b)(\mathbb{1}-g(Wx+b))^T W$

$\hookrightarrow$ since $\frac{\partial z}{\partial x_j} = cols(W)_j \Rightarrow \frac{\partial z}{\partial x} = W$

Image Denoising - Homework 5

Ex 1.4 $f_i(x, \theta_i) \rightarrow$ one layer

Let $\mathcal{F}(x) = f_3(y; \theta_3)$ where $y = f_2(\underbrace{f_1(x; \theta_1)}_{z}, \theta_2)$

$G(x) = y + f_3(y; \theta_3)$

a) $\quad = y + \mathcal{F}(x)$

- $\dfrac{\partial \mathcal{F}}{\partial \theta_3} = \dfrac{\partial f_3}{\partial \theta_3}$

$$\frac{\partial \mathcal{F}}{\partial \theta_2} = \frac{\partial \mathcal{F}}{\partial y} \frac{\partial y}{\partial \theta_2} = \frac{\partial f_3}{\partial y} \cdot \frac{\partial f_2}{\partial \theta_2}$$

$$\frac{\partial \mathcal{F}}{\partial \theta_1} = \frac{\partial \mathcal{F}}{\partial y} \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial \theta_1} = \frac{\partial f_3}{\partial y} \cdot \frac{\partial f_2}{\partial z} \cdot \frac{\partial f_1}{\partial \theta_1}$$

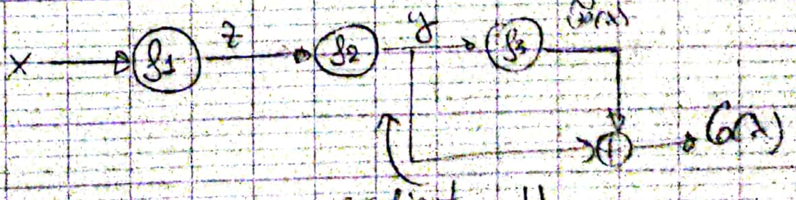
- $\dfrac{\partial G}{\partial \theta_3} = \dfrac{\partial f_3}{\partial \theta_3}$

$$\frac{\partial G}{\partial \theta_2} = \frac{\partial y}{\partial \theta_2} + \frac{\partial \mathcal{F}}{\partial \theta_2} = \frac{\partial f_2}{\partial \theta_2} + \frac{\partial f_3}{\partial y} \frac{\partial f_2}{\partial \theta_2} =$$

$$\frac{\partial G}{\partial \theta_1} = \frac{\partial y}{\partial \theta_1} + \frac{\partial \mathcal{F}}{\partial \theta_1} = \frac{\partial y}{\partial z} \frac{\partial z}{\partial \theta_1} + \frac{\partial f_3}{\partial y} \cdot \frac{\partial f_2}{\partial z} \cdot \frac{\partial f_1}{\partial \theta_1}$$

$$= \frac{\partial f_2}{\partial z} \cdot \frac{\partial f_1}{\partial \theta_1} + \frac{\partial f_3}{\partial y} \cdot \frac{\partial f_2}{\partial z} \cdot \frac{\partial f_1}{\partial \theta_1} =$$

b) For $\dfrac{\partial G}{\partial \theta_1}$



In this case the gradient has a direct path that prevents vanishing

$$\frac{\partial G}{\partial \theta_1} = \frac{\partial f_2}{\partial z} \frac{\partial f_1}{\partial \theta_1} \left(1 + \frac{\partial f_3}{\partial y} \right)$$

this term prevents vanishing if $\dfrac{\partial f_3}{\partial y}$ is small

Image Denoising - Homework 5

Ex 5 Let $x = (x_1, x_2) = (1, \varepsilon) \in \mathbb{R}^2$ $\quad 0 < \varepsilon < 1$ small

Then $\|x\|_1 = 1 + \varepsilon \quad \|x\|_2^2 = 1 + \varepsilon^2$

. Reduce one coordinate of x by $\delta < \varepsilon$ to reduce norm

. For L2 norm we have two possibilities

1) $x' = (1-\delta, \varepsilon) \rightarrow \|x'\|_2^2 = (1-\delta)^2 + \varepsilon^2$
$= 1 + \delta^2 + \varepsilon^2 - 2\delta$

2) $x' = (1, \varepsilon - \delta) \rightarrow \|x'\|_2^2 = 1 + (\varepsilon - \delta)^2$
$= 1 + \varepsilon^2 + \delta^2 - 2\varepsilon\delta$

Then the best case is $x = (1-\delta, \varepsilon)$

. For L1 norm in both cases we obtain the same result $\|x\|_1 = 1 - \delta + \varepsilon$

Escaneado con CamScanner