Elías Masquil

Image Denoising - Homework 5

1.1

NAND :

x_1	x_2	y
0	0	1
0	1	1
1	0	1
1	1	0

$$w_1 x_1 + w_2 x_2 + b \Rightarrow \begin{array}{ll} b & > 0 \\ w_2 + b & > 0 \\ w_1 + b & > 0 \\ w_1 + w_2 + b & < 0 \end{array}$$

$w = (-1, -2)$ $b = 2{,}5$ is one possible set of weights and bias for a NAND perceptron

1.2

2D - DCT can be obtained by applying the following equation successively to the rows and colums of an image

$$Y_k = \alpha_k \, 2 \sum_{j=0}^{N-1} x_j \cos\left[\pi\left(j + \frac{1}{2}\right)\frac{k}{N}\right]$$

with $\alpha_k = \begin{cases} \sqrt{1/4N} & k=0 \\ \sqrt{1/2N} & k=1, \dots, N-1 \end{cases}$ for $k = 0, \dots, N-1$.

x_j pixel j of the row / column.
N patch size

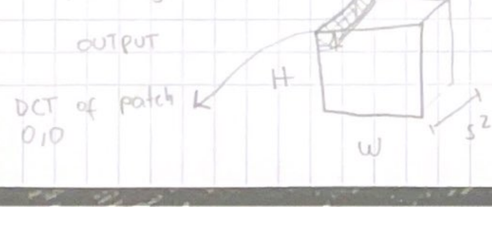
So we need $s^2 = 16$ convolutions of the image patches with the vectors of the DCT basis aranged as 2D images :

$$DCT(k,l)_{ij} = \alpha_k \alpha_l \, 4 \sum_i^3 \cos\left[\pi\left(i + \frac{1}{2}\right)\frac{k}{4}\right] \sum_j^3 \left[\pi\left(j + \frac{1}{2}\right) l\right]$$

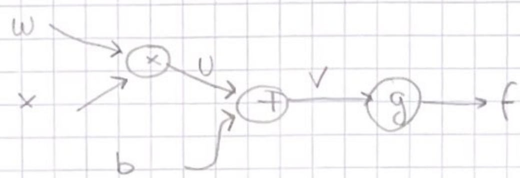
DCT as convolutional layers :

each of the filters correspond to a pair k, l of frequencies.

Finally the DCT of each patch (i, j) is the entry i, j of all the s^2 channels of the output.



1.3 $f(W,x,b) = g(Wx + b)$



$$\frac{df}{dW} = \frac{df}{dV}\frac{dV}{dU}\frac{dU}{dW} \qquad U = Wx$$

$$V = U + b$$

$$\frac{df}{dx} = \frac{df}{dV}\frac{dV}{dU}\frac{dU}{dx} \qquad f = g(V)$$

$$\frac{df}{db} = \frac{df}{dV}\frac{dV}{db}$$

$$\frac{df}{dV} = g(V)(1 - g(V))$$ derivative of sigmod

$$\frac{dV}{dU} = Id_{3\times3} \qquad \frac{dV}{db} = Id_{3\times3}$$

$$U_i = \sum_j W_{ij} x_j$$

$$\frac{dU_i}{dx_j} = W_{ij} \implies \frac{dU}{dx} = W$$

$$\frac{dU_i}{dW_{ij}} = x_j \implies \frac{dU}{dW_{ij}} = \begin{bmatrix} 0 \\ \vdots \\ x_j \\ 0 \\ 0 \end{bmatrix} \rightarrow i\text{th}$$

$$\Rightarrow \boxed{\frac{df}{db} = g(Wx+b)(\vec{1} - g(Wx+b))^T}$$

$$\boxed{\frac{df}{dx} = g(Wx+b)(1 - g(Wx+b))^T W}$$

$$\boxed{\frac{df_i}{\partial W_{ij}} = g(V_i)(1 - g(V_i)) \cdot x_j}$$

Elias Masquil

Image Denoising - Homework 5

1.4

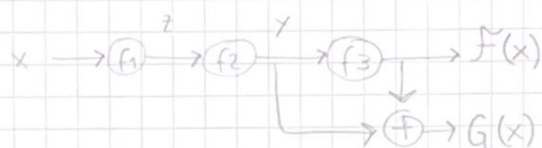
$$F(x) = f_3(y; \theta_3) \qquad z = f_1(x, \theta_1)$$

$$y = f_2(f_1(x; \theta_1); \theta_2) \qquad \frac{dF}{d\theta_3} = \frac{dG}{d\theta_3}$$

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

$$\frac{dF}{d\theta_2} = \frac{df_3}{dy}\frac{df_2}{d\theta_2} \qquad \frac{dF}{d\theta_1} = \frac{df_3}{dy}\frac{df_2}{dz}\frac{df_1}{d\theta_1}$$

$$\frac{dG}{d\theta_2} = \frac{df_3}{dy}\frac{df_2}{d\theta_2} + \frac{df_2}{d\theta_2} \qquad \frac{dG}{d\theta_1} = \frac{df_3}{dy}\frac{df_2}{dz}\frac{df_1}{d\theta_1} + \frac{df_2}{dz}\frac{df_1}{d\theta_1}$$



The gradient has an extra path for G when doing backpropagation, without being scaled by $\frac{df_3}{dy}$

1.5

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

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

We're reducing only one coordinate by a factor of $\delta < \varepsilon$.

$x' = (1-\delta, \varepsilon) \qquad x'' = (1, \varepsilon - \delta)$

$\|x'\|_1 = 1 + \varepsilon - \delta \qquad \|x''\|_1 = 1 + \varepsilon - \delta$

$\|x'\|_2^2 = 1 - 2\delta + \varepsilon^2 + \delta^2 \qquad \|x''\|_2^2 = 1 + \varepsilon^2 + \delta^2 - 2\varepsilon\delta$

For the L1 norm the reduction is the same despite of the selected coordinate.

For the L2 norm, the maximum reduction is obtained by decreasing the first coordinate.

L^2 reg enforces having <u>all</u> entries <u>close</u> to 0, while L^1 reg allows some large weights if <u>most</u> of them <u>are 0</u>.