王芸飞 2021533135

No.
Date. / /

Hw5

1. (a) ① feature map (width size) $= \frac{W-K+2P}{S} + 1 = \frac{64-3+2\times1}{1} + 1$

and ∵ there are 4 output channels $= 64$

∴ output size $= 4 \times 64 \times 64$

② For one unit kernel, there are $3\times3\times4$ parameters

and ∵ there are 4 output channels every layer and 10 kernel (10 layers)

∴ we have $10\times(3\times3\times4\times4 + \underbrace{4\times1}_{bias}) = 1480$ parameters in total

(b) ① we can also use $\left(\frac{W-K+2P}{S}+1\right)$ to calculate the result width

$= \frac{64-2}{2} + 1 = 32$

But pooling layer doesn't influence the depth of input

∴ output depth is also 4

∴ output size $= 4\times32\times32$

② Because max pooling is a rule / formula

∴ it doesn't need other parameters

∴ In pooling layer, we need 0 parameter.

2. (a) Kmeans ++

① The first centroid $x^{(1)} = (2.8)$

		D^2	Normalized
(2.5)	$x^{(2)}$	9	$\frac{9}{194}$
(1.2)	$x^{(3)}$	37	$\frac{37}{194}$
(5.8)	$x^{(4)}$	9	$\frac{9}{194}$
(7.3)	$x^{(5)}$	50	$\frac{50}{194}$
(6.4)	$x^{(6)}$	32	$\frac{32}{194}$ ⇒ [0.541, 0.706)
(8.4)	$x^{(7)}$	52	$\frac{52}{194}$
(4.7)	$x^{(8)}$	5	$\frac{5}{194}$

the first random number is 0.6
∴ we will chose $x^{(6)}$ as the next centroid.

② The second centroid $x^{(6)} = (6.4)$

		D^2	⇒ $\min(D^2)$	Normalized
(2.5)	$x^{(2)}$	17	9	$\frac{9}{58}$ [0, 0.155)
(1.2)	$x^{(3)}$	29	29	$\frac{29}{58}$ [0.155, 0.5)
(5.8)	$x^{(4)}$	17	9	$\frac{9}{58}$
(7.3)	$x^{(5)}$	2	2	$\frac{2}{58}$
(8.4)	$x^{(7)}$	4	4	$\frac{4}{58}$
(4.7)	$x^{(8)}$	13	5	$\frac{5}{58}$

the second random number is 0.2
∴ we will chose $x^{(3)}$ as the third centroid.

∴ Finally three centroids

$$\begin{cases} C_1: x^{(1)} = (2.8) \\ C_2: x^{(6)} = (6.4) \\ C_3: x^{(3)} = (1.2) \end{cases}$$

No.

Date. / /

Iteration 1

(b)

	C_1	C_2	C_3
$x^{(1)}$	0	$\sqrt{32}$	$\sqrt{37}$
$x^{(2)}$	3	$\sqrt{17}$	$\sqrt{10}$
$x^{(3)}$	$\sqrt{37}$	$\sqrt{29}$	0
$x^{(4)}$	3	$\sqrt{17}$	$\sqrt{52}$
$x^{(5)}$	$\sqrt{50}$	$\sqrt{2}$	$\sqrt{37}$
$x^{(6)}$	$\sqrt{32}$	0	$\sqrt{29}$
$x^{(7)}$	$\sqrt{52}$	2	$\sqrt{53}$
$x^{(8)}$	$\sqrt{5}$	$\sqrt{13}$	$\sqrt{34}$

$Q(r,c) = \frac{1}{8}\cdot[0+9+0+9+2+4+0+5]$
$= \frac{29}{8}$

At the same time new $C_1 = (\frac{13}{4}, 7)$
new $C_2 = (7, \frac{11}{3})$
new $C_3 = (1, 2)$

(c) iteration 2

	$C_1\ (\frac{13}{4}, 7)$	$C_2\ (7, \frac{11}{3})$	$C_3\ (1,2)$
$x^{(1)}\ (2,8)$	$\sqrt{\frac{41}{16}}$	$\sqrt{\frac{394}{9}}$	$\sqrt{37}$
$x^{(2)}\ (2,5)$	$\sqrt{\frac{89}{16}}$	$\sqrt{\frac{241}{9}}$	$\sqrt{10}$
$x^{(3)}\ (1,2)$	$\sqrt{\frac{481}{16}}$	$\sqrt{\frac{349}{9}}$	0
$x^{(4)}\ (5,8)$	$\sqrt{\frac{65}{16}}$	$\sqrt{\frac{205}{9}}$	$\sqrt{52}$
$x^{(5)}\ (7,3)$	$\sqrt{\frac{481}{16}}$	$\sqrt{\frac{4}{9}}$	$\sqrt{37}$
$x^{(6)}\ (6,4)$	$\sqrt{\frac{265}{16}}$	$\sqrt{\frac{10}{9}}$	$\sqrt{29}$
$x^{(7)}\ (8,4)$	$\sqrt{\frac{505}{16}}$	$\sqrt{\frac{10}{9}}$	$\sqrt{53}$
$x^{(8)}\ (4,7)$	$\sqrt{\frac{9}{16}}$	$\sqrt{\frac{181}{9}}$	$\sqrt{34}$

∴ Just need two iteration, the algorithm can converge.

2

$Q(r,c) = \frac{1}{8}(\frac{41}{16} + \frac{89}{16} + \frac{65}{16} + 0 + \frac{4}{9} + \frac{10}{9} + \frac{10}{9} + \frac{9}{16})$
$= \frac{185}{96}$

No.
Date. / /

3. ① VAE: Variational Autoencoder
GAN: Generative Adversarial Network ⇒ cycle GAN / stack GAN / conditional GAN

Here, pdata is the distribution of real data, and z is a random noise sampled from prior distribution p(z), and then G(z) is the generated data, and D is the discriminator.

$\boxed{\text{Explanation}}$

② A: minimax objective function

$$\min_{\theta g} \max_{\theta d} \left[E_{x\sim p_{data}} \log D_{\theta d}(x) + E_{z\sim p(z)} \log(1 - D_{\theta d}(G_{\theta g}(z))) \right]$$

B: How to update parameters

a. we update the discriminator at first by using gradient ascent

$$\max_{\theta d} \left[E_{x\sim p_{data}} \log D_{\theta d}(x) + E_{z\sim p(z)} \log(1 - D_{\theta d}(G_{\theta g}(z))) \right]$$

using backprop on the classification objective

b. Then we update the generator by using gradient descent

$$\min_{\theta g} E_{z\sim p(z)} \log(1 - D_{\theta d}(G_{\theta g}(z)))$$

using backprop