CSE616

Assignment 2

Abdullah Aml, 2101398

Question1

- **Without bias:** W of size = (750000 x 100)
- **With bias:** W of size = (750001 x 100)

Question2

Filter sizes = 10 x 5 x 5 = 250

Question3

- First Image: vertical edge filter:

1	0	-1
1	0	-1
1	0	-1

- Second Image: horizontal edge filter:

1	1	1
0	0	0
-1	-1	-1

Question4 (not a clear answer)

The second momentum (exponential moving average) will tend too use old second momentum therefore conserves the old direction.

Qustion5

$$mean = \frac{1}{m}\sum_{k=1}^{m} z^{(m)}$$

$$var = \frac{1}{m}\sum_{k=1}^{m} \left(z^{(m)} - mean\right)^2$$

$$\hat{Z^{(i)}} = \frac{z^{(i)} - mean}{\sqrt{var + \varepsilon}} \qquad \varepsilon \text{¿ avoid dividing by zero}$$

$$y^{(i)} = \gamma \hat{Z^{(i)}} + \beta;\ \gamma, \beta : \text{are learnalble parameters}$$

we use batch-normalization to:

- Improves gradient flow through the network
- Allows higher learning rates
- Reduces the strong dependence on initialization
- Acts as a form of regularization

Qustion6

We have:

input image (256x256x3) → 10 @ (3x3x3) filters ,stride=1 → pooling (3x3), stride=2

256x256x3 → (256-3+1) x (256-3+1) → (254-3+1)/2 x (254-3+1)/2

256x256x3 → (254) x (254) → 126 x 126

receptive field size = 126x126

Question7

Output size = 128 x ceil((128-7+(2x3)+1)/2) x ceil((128-7+(2x3)+1)/2)

= 128 x (64 x 64)

Question9

- Much more number of parameters:
  - more computaion time
- does not have local receptive field if the object position is moved within the image the FCNN will not be able to classify the image

Question10

	4	1	-1	3	
1	-2				
	1	-2			
		1	-2		
			1	-2	
				1	-2

	4x-2	4x1 + 1x-2	1x1 + -1 x -2	-1 x 1 + 3 x -2	1 x 3

output	-8	2	3	-7	3

Question11

Decreasing learning rate

Question12

- It preserves local receptive field
- due to shred weights
- it is translation invariant

Question13

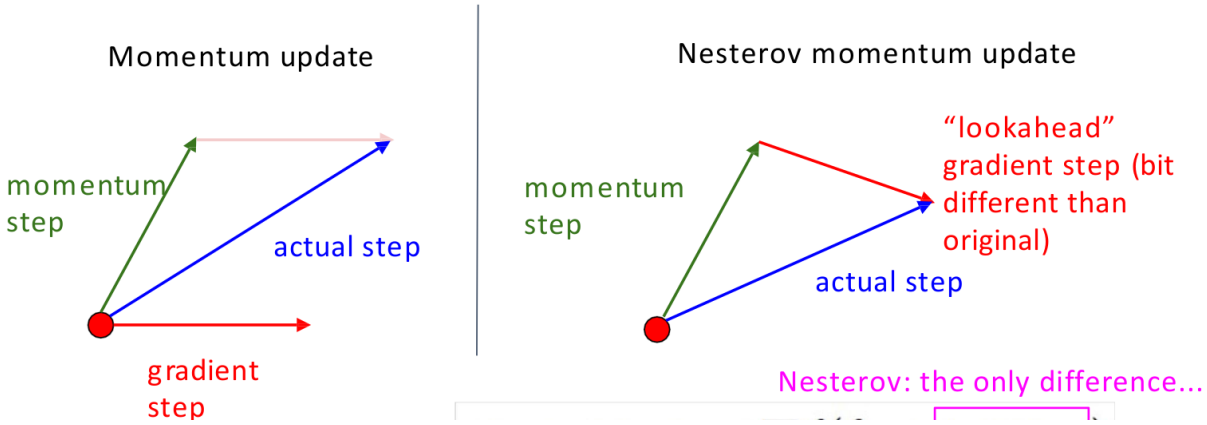
- **In train:** we multiply weights matrix by a matrix with ones and zeros to turnoff some weights with p.

- **In test:** we multiply the output of the weights by p.

Question14

Nesterov conservesk more the direction of the old gradient because it multiplies by learning rate before adding to v

picture form slide:



Question15

$$grad = grad + dx^2; dx \, is \, the \, gradient \, of \, SGD$$

$$lr_{new} = \frac{lr}{\sqrt{grad + \varepsilon}}$$