FIT3181 Deep learning - S2 2022 MUM

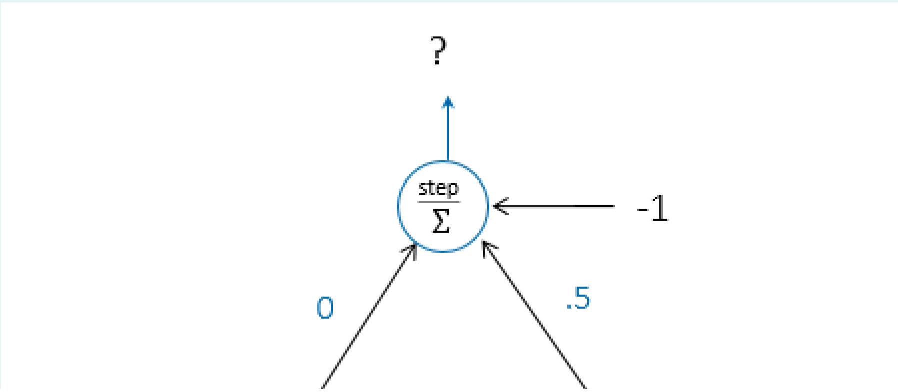
Started on	Saturday, 17 September 2022, 4:44 PM
State	Finished
Completed on	Saturday, 17 September 2022, 5:23 PM
Time taken	39 mins 23 secs
Marks	38.17/45.00
Grade	**8.48** out of 10.00 (**85**%)

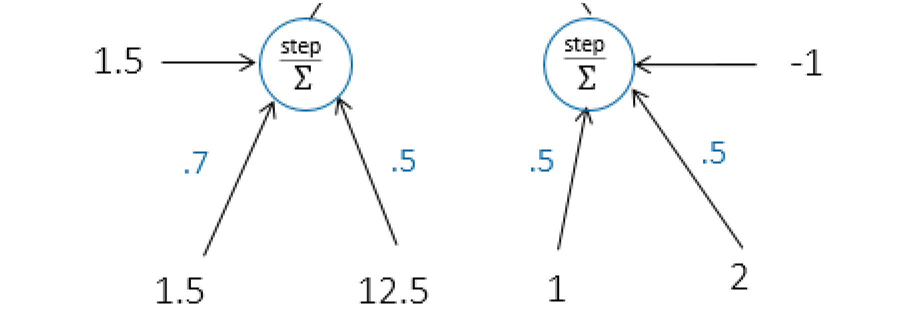
Question **1**

Complete

Mark 2.00 out of 2.00

Let's step(x) = 0 if x <= 0 and step (x) = 1 if x > 0, the output value for the architecture below is:





Select one:

- a. 0
- b. -1
- c. 1
- d. 1/2
- e. -½
- f. None of above

Question 2

Complete

Mark 1.00 out of 1.00

Max pooling to subsample images, in general, reduces the variances between them

Select one:

- True
- False

Question 3

Complete

Which is the case caused by underfitting?

Mark 1.00 out of 1.00

Select one:

- a. Training accuracy: 99%, Testing accuracy: 98%
- b. Training accuracy: 99%, Testing accuracy: 50%
- c. Training accuracy: 90%, Testing accuracy: 40%
- d. Training accuracy: 30%, Testing accuracy: 40%

Question **4**

Complete

Mark 1.00 out of 1.00

When evaluating the performance of a binary classification algorithm, the Precision is the same as the Sensitivity.

Select one:

precision = TP / (TP + FP)
sensitvity = TP / (TP + FN)

- True
- False

Question **5**

Complete

Mark 1.00 out of 1.00

Given a prediction softmax probability p=[0.1, 0.4,0.3,0.2] for a data instance x with the true label y=3 in the label set {1, 2, 3, 4}, what is cross-entropy loss?

Select one:

- a. log 0.4
- b. log 0.3
- c. -log 0.4
- d. -log 0.3

Question 6
Complete
Mark 1.50 out of 1.50

Let $f(x,y) = x\log x + y\log y$ what is the correct formula for the gradient vector ∇f? (hint: $(uv)' = u'v + v'u$ and $(\log h)' = \frac{h'}{h}$)

Select one:

- a. $[\log x + 2, \log y + 2]$
- b. softmax([x,y])
- c. $[\log x, \log y]$
- d. No correct answer(s) listed
- e. $[\log x + 1, \log y + 1]$
- f. $[x + 1, y + 1]$

Question 7
Complete
Mark 0.00 out of 1.00

Assume that over the 4D input tensor with the shape [32, 128, 128, 20], we apply max pooling with kernel size [3,3], strides [2,2], and padding valid. What is the number of parameters for the max pooling layer?

Select one:

- a. 9
- b. 0 floor ((128 - 3)/ 2) + 1 = 63
- c. 36
- d. 180 63 x 3 x 3 x 20

Question **8**

Complete

Mark 0.00 out of 1.00

Given the following 1D input ([15x1]) and filter ([3x1]), assume that padding=same and the stride size = 4, what is the output tensor?

0	1	2	1	1	1	1	1	1	1	1	1	2	1	0

1	1	1

Select one:

- a. [3, 4, 4, 3]
- b. [3, 2, 3, 3]
- c. [3, 3, 3, 2]
- d. [3, 3, 3, 3]

Question **9**

Complete

Mark 1.00 out of 1.00

Given the following 1D input ([15x1]) and filter ([3x1]), assume that padding =same and strides 3, what is the output size?

Select one:

- a. 4
- b. 3
- c. 6
- d. 5

Question **10**

Given an adversarial example x_{adv} of a clean example x w.r.t model f, y∈{1,2,...,M} is the true label. Which

Complete

Mark 1.50 out of 1.50

statements are correct?

Select one or more

- [ ] a. $argmax_{1 \leq m \leq M} f_m(x_{adv}) = y$
- [x] b. $argmax_{1 \leq m \leq M} f_m(x_{adv}) \neq y$
- [ ] c. x_{adv} is a noise image.
- [x] d. x_{adv} and x look very similar under human perspective
- [ ] e. x_{adv} and x look different under human perspective

Question **11**

Complete

Mark 1.50 out of 1.50

With respect to the Pooling layer in a CNN, which of the following statements are true:

Select one or more:

- [ ] a. It enlarges the dimension of the image, hence provides a better resolution
- [x] b. Output tensor of a max-pooling layer will always have the same depth with the input tensor
- [x] c. Max-pooling is locally invariant in the sense that input numbers within a local filter window can be shuffled without changing the final output
- [x] d. It operates at each activation map independently

Question **12**

Complete

Mark 1.00 out of 1.00

Assume that we have 4 classes in {cat = 1,dog = 2,lion = 3, monkey = 4}. Given a data example x with ground-truth label dog, assume that a feed-forward NN gives discriminative scores to this x as h_1=-2, h_2=2, h_3=3, h_4=1 What is the probability to predict x as lion or p(y=lion | x)?

(•) a. exp(3)/[exp(-2) + exp(2) + exp(3) + exp(1)]

a. exp(3)/[exp(-2) + exp(2) + exp(3) + exp(1)]

b. exp(1)/[exp(-2) + exp(2) + exp(3) + exp(1)]

c. [0.25, 0.25, 0.25, 0.25]

d. exp(1)/[exp(-2) + exp(2) + exp(3) + exp(1)]

Question 13

Complete

Mark 1.50 out of 1.50

Which of the following statements are correct regarding the global pooling layer?

Select one or more:

- [ ] a. Unlike a normal pooling layer, the global max pooling layer has lots of parameters to learn
- [ ] b. The global pooling layer usually has kernel or filter size equal to a quarter of the input size
- [x] c. The global pooling layer usually has kernel or filter size equal to input size
- [ ] d. Unlike a normal pooling layer, the global max pooling layer is not invariant to the input
- [x] e. Each feature map will become a single neuron

Question 14

Complete

Mark 2.00 out of 2.00

Assume that the tensor before the last tensor of a CNN has shape [32, 32, 32, 10] and we apply 5 filters each of which has the shape [5,5,10] and strides= [3,3] with padding = 'same' to obtain the last tensor. We flatten this tensor to a fully connected (FC) layer. What is the number of neurons on this FC layer?

a. 32x11x11x5

b. 8x8x5

c. 32x8x8x5

d. 11x11x5 (selected)

- e. 9x9x5

Question **15**

Complete

Mark 0.00 out of 1.00

Which of the following statements are correct for targeted attack?

Select one or more:

- [x] a. In a targeted attack, we minimize the loss of the model on the adversarial example with respect to its ground-truth label.
- [ ] b. In targeted attack, we maximize the loss of the model on the adversarial example with respect to a given label different from ground-truth label.
- [ ] c. In a targeted attack, we minimize the loss of the model on the adversarial example with respect to a given label different from ground-truth label.
- [ ] d. In a targeted attack, we maximize the loss of the model on the adversarial example with respect to its ground-truth label.

Question **16**

Complete

Mark 1.50 out of 1.50

Given an implementation as below. What is the total number of parameters of the model?

```
dnn_model = Sequential()
dnn_model.add(Dense(units=30, input_shape=(10,), activation='relu'))
dnn_model.add(Dense(units=30, activation='relu'))
dnn_model.add(Dense(units=10))
```

Select one:

(30x10 +30) + (30x30 + 30) + (30 x10 + 10) =

- a. 1580
- b. 1570 (selected)
- c. 1500

c. 1500

d. 1560

Question 17
Complete
Mark 3.00 out of 3.00

John was asked to build a prediction algorithm to detect fraud tax claim from a dataset that contains 1,000 genuine tax return claims and 10 fraud claims. John decided to aim for gaining high accuracy by making a deterministic lazy algorithm which simply classifies every claim as genuine. Which of the statements are true:

Select one or more:

- [x] a. Given a newly unseen dataset containing both genuine and fraud claims, John's algorithm will gain 100% recall rate for genuine claims
- [ ] b. John is experiencing an overfitting problem in his approach
- [ ] c. Given a newly unseen dataset containing both genuine and fraud claims, John's algorithm will gain 100% recall rate for fraud claims
- [x] d. John is experiencing an underfitting problem in his approach
- [x] e. With respect to the given dataset, John is making an excellent prediction accuracy
- [x] f. John will run into an imbalanced problem, hence the accuracy is not reliable

Question 18
Complete
Mark 1.00 out of 1.00

Let the softmax prediction probability p(x)=[0.1, 0.4,0.3,0.2] for a data instance x, what is the predicted label y (assuming we are using index starting from 1)?

Select one:

a. $\hat{y} = 4$

b. $\hat{y} = 2$

b. $\hat{y} = 3$

c. $\hat{y} = 2$

d. $\hat{y} = 0.1$

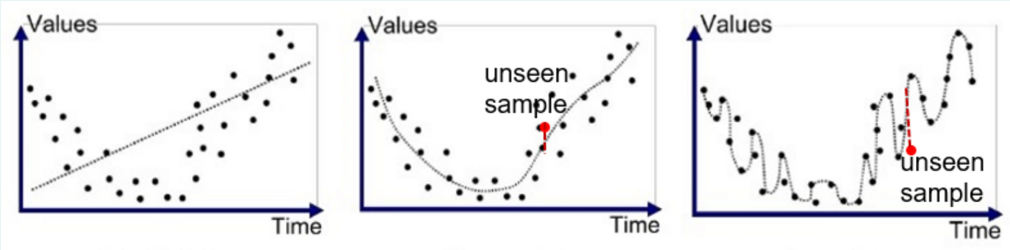
e. $\hat{y} = 1$

Question **19**

Complete

Mark 1.00 out of 1.00

What is the correct statement according to the following figures?



Select one:

a. A is overfitting, B is good-fit, and C is underfitting

b. A is underfitting, B is good-fit, and C is overfitting

c. A is overfitting, B is underfitting, and C is good-fit

d. A is underfitting, B is overfitting, and C is good-fit

Question **20**

Complete

Mark 0.50 out of 1.50

Choose the correct statements for the sigmoid activation function.

Select one or more:

a. The derivative of sigmoid does not approach 0

- [x] b. Derivative of sigmoid becomes very small if the stimulus (input) is too big or too small
- [ ] c. Sigmoid can cause gradient vanishing
- [x] d. Sigmoid is a saturated activation function
- [ ] e. Sigmoid can have negative values

Question **21**

Complete

Mark 1.00 out of 1.00

Which statement is correct for local minima?

Select one:

- a. Local minima is the highest value in the neighborhood around in data point in all directions
- b. Local minima is the lowest value in the neighborhood around in data point in some directions and the highest value in other directions
- (selected) c. Local minima is the lowest value in the neighborhood around in data point in all directions
- d. No correct answer(s) listed

Question **22**

Complete

Mark 0.67 out of 1.00

Which of the following formulas expresses the chain rule for du/dx?

Select one or more:

- [x] a. $\frac{du}{dx} = \frac{du}{dh} \times \frac{dh}{dx}$
- [x] b. $\frac{du}{dx} = \frac{dx}{du} \times \frac{du}{dx} \times \frac{du}{dx}$
- [ ] c. $\frac{du}{dx} = \frac{sin(u)}{cos(x)} \times \frac{dx}{du}$

d. $\frac{du}{dx} = \frac{du}{dv} \times \frac{dv}{dh} \times \frac{dh}{dx}$

e. $\frac{du}{dx} = \frac{d\theta}{dx} \times \frac{du}{d\beta} \times \frac{d\beta}{d\theta}$

Question 23

Complete

Mark 1.00 out of 1.00

Which statement is correct for a saddle point?

Select one:

a. Saddle point is the highest value in the neighborhood around in data point in all directions

b. Saddle point is the lowest value in the neighborhood around in data point in all directions

c. No correct answer(s) listed

d. Saddle point is the lowest value in the neighborhood around in data point in some directions and the highest value in other directions

Question 24

Complete

Mark 0.00 out of 2.00

Regarding different choices for activation functions used in deep NNs, which of the following statements are generally applicable in practice:

Select one or more:

a. The sigmoid and the tanh functions have the same input domain, however tanh has an output range of (-1, 1) hence generally speeds up convergence better than the sigmoid function

b. A very deep NNs with sigmoid function will generally encounter gradient exploding problem as the gradient will grow up very quickly

c. The output range for ReLU function is [0, 1]

- [x] d. ReLU activation function is very fast to compute, hence it works very well in practice
- [ ] e. The sigmoid function is much better than ReLU at dealing with the gradient vanishing problem
- [ ] f. All of above

Question **25**

Complete

Mark 1.50 out of 1.50

Consider a 2D operation on an 36x36 image with a 5x5 filter, stride=2x2, zero padding size = 2, what is the dimension of the output image

Select one:

floor(36 - 1) / 2 + 1 = 18

- ( ) a. 36x36
- ( ) b. 52x52
- ( ) c. No correct answer(s) listed
- (•) d. 18x18
- ( ) e. 5x5

Question **26**

Complete

Mark 2.00 out of 2.00

Which of the following statement are correct regarding Gradient Descent (GD) method when applied to minimize the objective function J(w) (you must select all applicable answers to get the full mark for this question):

Select one or more:

- [x] a. GD updates the parameter in the opposite direction of the current gradient
- [ ] b. GD runs much faster than Stochastic GD for large datasets
- [x] c. GD requires a full gradient calculation over the entire dataset in each iteration

- [ ] d. GD does not guarantee to converge a global minimum even if J(w) is a convex function
- [ ] e. GD guarantees to converge to some local minimum if J(w) is a nonconvex function
- [x] f. GD is a first-order optimization method

Question **27**

Complete

Mark 1.50 out of 1.50

Assume that the last 3D output tensor in a CNN before flattening has the shape [20, 10, 32]. What is the number of neurons on the flattened layer?

Select one:

- (•) a. 6400
- ( ) b. 640
- ( ) c. 320
- ( ) d. 200

Question **28**

Complete

Mark 1.50 out of 1.50

Consider the optimization problem of deep learning:

$$min_\theta J(\theta) = \Omega(\theta) + \frac{1}{N}\sum_{i=1}^{N} l(y_i, f(x_i, \theta))$$

Let $\Omega(\theta)$ be the first term and $\frac{1}{N}\sum_{i=1}^{N} l(y_i, f(x_i, \theta))$ be the second term. Select all correct statements.

Select one or more:

- [ ] a. The second term helps to combat overfitting

- [ ] b. The first term help to combat underfitting
- [x] c. The first term is the regularization term used to encourage simple models
- [x] d. The first term helps to combat overfitting
- [x] e. The second term is the empirical loss term used to guarantee the model working well on the training set

Question **29**
Complete
Mark 2.00 out of 2.00

Which of the following statement are true regarding the gradient vanishing problem when training a Deep NN using BackProp?

Select one or more:

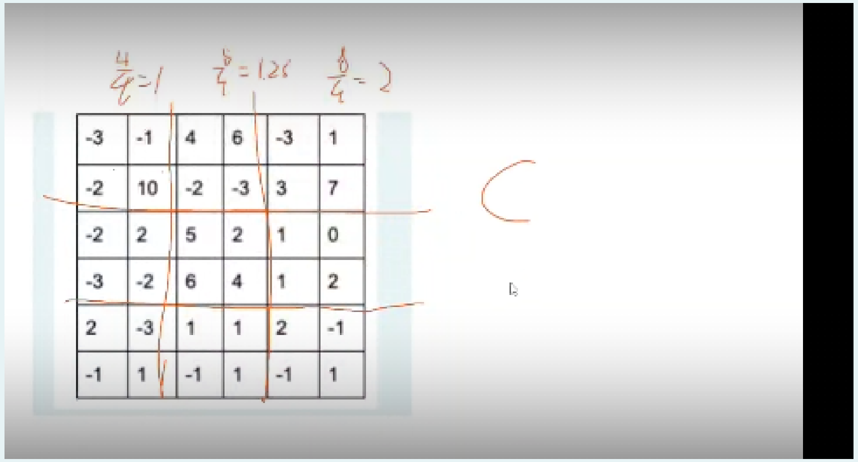
- [ ] a. Gradient clipping is another approach to address gradient vanishing problem in which the direction of the gradient will never be changed
- [ ] b. Gradient vanishing problem can be addressed by using the sigmoid function
- [ ] c. With ReLU activation functions, it caps on the output range hence the gradient signals get smaller and smaller
- [x] d. SGD update leaves the lower connection weights almost unchanged and hence never converges to a good solution
- [x] e. The gradient signals get smaller and smaller as the algorithm progresses down to the lower layers.

Question **30**
Complete
Mark 1.50 out of

Assume that we have a [6,6,1] input tensor as shown below and applying average pooling with kernel size = [2,2], strides = [2,2], and padding option as ‘valid’. What is the value of the first row in the output tensor?

1.50

-3	-1	4	6	-3	1
-2	10	-2	-3	3	7
-2	2	5	2	1	0
-3	-2	6	4	1	2
2	-3	1	1	2	-1
-1	1	-1	1	-1	1



Select one:

- a. [10,6,7]
- b. [1,1,2]
- c. [1,1.25,2]
- d. [-3,-3,-3]

Question **31**

Complete

Mark 0.50 out of 1.00

Consider the gradient descent optimisation method, which of the following statements are correct?

Select one or more:

- a. Very big learning rate helps gradient descent to converge faster
- b. Gradient descent can be gotten stuck in local minima for non-convex functions
- c. The solution is updated by following the direction of the gradient
- d. The solution is updated by following the direction of the negative gradient

Question **32**

Complete

Mark 2.00 out of 2.00

Which of the following statement are true regarding Gradient Descent (GD) method when applied to minimize the objective function J(w):

Select one or more:

- [ ] a. GD guarantees to converge to some local minimum if J(w) is a nonconvex function
- [ ] b. GD is a second-order optimization method
- [x] c. GD updates the parameter in the opposite direction of the current gradient
- [ ] d. GD runs much faster than Stochastic GD for large datasets
- [x] e. GD guarantees to converge a global minimum if J(w) is a convex function

◀ Practice Test 01

Jump to...

Material for Week 0 ▶