FIT 3181/5215 Deep Learning

Quiz for:
Practical Skills in Deep Learning

Tutor Team

Department of Data Science and AI
Faculty of Information Technology, Monash University
Email: **trunglm@monash.edu**

Question 1

What should be if the gradient norm $||g||_2$ (i.e., gradient of the loss w.r.t. all model parameters) of a deep learning model is decreasing to 0?

- ❑ A. The training arrives at local minima
- ❑ B. The training arrives at local maximal
- ❑ C. The training arrives at saddle point
- ❑ D. One of A, B, C
- ❑ E. None of A, B, C

Question 1

What should be if the gradient norm $||g||_2$ (i.e., gradient of the loss w.r.t. all model parameters) of a deep learning model is decreasing to 0?

- ❑A. The training arrives at local minima
- ❑B. The training arrives at local maximal
- ❑C. The training arrives at saddle point
- ❑D. One of A, B, C **[x]**
- ❑E. None of A, B, C

Question 2

Which activation function that Xavier initialization is good for? (MC if needed)

- ❑A. Sigmoid
- ❑B. Tanh
- ❑C. ReLU

Question 2

Which activation function that Xavier initialization is good for? (MC if needed)

- ❑ A. Sigmoid **[x]**
- ❑ B. Tanh **[x]**
- ❑ C. ReLU

Question 3

Which activation function that He initialization good for? (MC if needed)

- ❑A. Sigmoid
- ❑B. Tanh
- ❑C. ReLU

Question 3

Which activation function that He initialization good for? (MC if needed)

- ❑A. Sigmoid
- ❑B. Tanh
- ❑C. ReLU **[x]**

Question 4

What is gradient vanishing problem?

- ❑ A. Too big gradient at the higher layers of model
- ❑ B. Too big gradient at the lower layers of model
- ❑ C. Too small gradient at any layer of model
- ❑ D. Too small gradient at the lower layers of model

Question 4

What is gradient vanishing problem?

- ❑A. Too big gradient at the higher layers of model
- ❑B. Too big gradient at the lower layers of model
- ❑C. Too small gradient at any layer of model **[x]**
- ❑D. Too small gradient at the lower layers of model

Question 5

What is gradient exploding problem?

- ❑ A. Too big gradient at any layer of model
- ❑ B. Too big gradient at the lower layers of model
- ❑ C. Too small gradient at layers of model
- ❑ D. Too small gradient at the lower layers of model

Question 5

What is gradient exploding problem?

- ❑A. Too big gradient at any layer of model **[x]**
- ❑B. Too big gradient at the lower layers of model
- ❑C. Too small gradient at layers of model
- ❑D. Too small gradient at the lower layers of model

Question 6

Which layer can help to reduce gradient vanishing problem? (MC)

- ❑ A. Convolutional layer
- ❑ B. Pooling layer
- ❑ C. Fully Connected layer
- ❑ D. Dropout layer
- ❑ E. Batch normalization layer
- ❑ F. Skip connection layer
- ❑ G. ReLU Activation

Question 6

Which layer can help to reduce gradient vanishing problem? (MC)

- ❑A. Convolutional layer
- ❑B. Pooling layer
- ❑C. Fully Connected layer
- ❑D. Dropout layer
- ❑E. Batch normalization layer **[x]**
- ❑F. Skip connection layer **[x]**
- ❑G. ReLU Activation **[x]**

Question 7

Which is the phenomenon of overfitting problem (given a small dataset such as MNIST and a baseline model can achieve more than 90% training accuracy)?

- ❑A. Training accuracy: 99%, Testing accuracy: 50%
- ❑B. Training accuracy: 99%, Testing accuracy: 90%
- ❑C. Training accuracy: 70%, Testing accuracy: 40%
- ❑D. Training accuracy: 30%, Testing accuracy: 40%

Question 7

Which is the phenomenon of overfitting problem (given a small dataset such as MNIST and a baseline model can achieve more than 90% training accuracy)?

- ❑A. Training accuracy: 99%, Testing accuracy: 50% **[x]**
- ❑B. Training accuracy: 99%, Testing accuracy: 90%
- ❑C. Training accuracy: 70%, Testing accuracy: 40%
- ❑D. Training accuracy: 30%, Testing accuracy: 40%

In fact, there is no a gold threshold to say which case occurs overfitting. But for a small dataset like MNIST, baseline models such as baseline CNNs easily achieve more than 90% training accuracy. So A should be the correct answer.

Question 8

Which is the phenomenon of underfitting problem (given a small dataset such as MNIST and a baseline model can achieve more than 90% training accuracy)?

- ❑ A. Training accuracy: 99%, Testing accuracy: 50%
- ❑ B. Training accuracy: 99%, Testing accuracy: 90%
- ❑ C. Training accuracy: 70%, Testing accuracy: 40%
- ❑ D. Training accuracy: 30%, Testing accuracy: 40%

Question 8

Which is the phenomenon of underfitting problem (given a small dataset such as MNIST and a baseline model can achieve more than 90% training accuracy)?

- ❑A. Training accuracy: 99%, Testing accuracy: 50%
- ❑B. Training accuracy: 99%, Testing accuracy: 90%
- ❑C. Training accuracy: 70%, Testing accuracy: 40% **[x]**
- ❑D. Training accuracy: 30%, Testing accuracy: 40% **[x]**

In fact, there is no a gold threshold to say which case occurs underfitting or overfitting. But for a small dataset like MNIST, baseline models such as baseline CNNs easily achieve more than 90% training accuracy. So C, D should be correct answers.

Question 9

Which likely to be overfitting?

- ❑A. Very big model with few images
- ❑B. Very small model with many images

Question 9

Which likely to be overfitting?

- ❑A. Very big model with few images **[x]**
- ❑B. Very small model with many images

Question 10

Which helps to reduce the overfitting problem? (MC)

- ❑ A. Train DL model as many epochs as possible
- ❑ B. Early stopping
- ❑ C. Adding more data
- ❑ D. Weight decay
- ❑ E. Adding more layer
- ❑ F. Using Dropout
- ❑ G. Adding more neurons

Question 10

Which helps to reduce the overfitting problem? (MC)

- ❑ A. Train DL model as many epochs as possible
- ❑ B. Early stopping **[x]**
- ❑ C. Adding more data **[x]**
- ❑ D. Weight decay **[x]**
- ❑ E. Adding more layer
- ❑ F. Using Dropout **[x]**
- ❑ G. Adding more neurons

Question 11

In batch-normalization algorithm (in training time), which are trainable parameters? (MC)

- ❑A. Minibatch mean μ_B.
- ❑B. Minibatch standard deviation σ_B.
- ❑C. Scaling parameter γ.
- ❑D. Shifting parameter β.

Question 11

In batch-normalization algorithm (in training time), which are trainable parameters? (MC)

- ❑A. Minibatch mean μ_B.
- ❑B. Minibatch standard deviation σ_B.
- ❑C. Scaling parameter γ. **[x]**
- ❑D. Shifting parameter β. **[x]**

1. $\boldsymbol{\mu}_B = \frac{1}{m_B} \sum_{i=1}^{m_B} \mathbf{x}^{(i)}$
2. ${\boldsymbol{\sigma}_B}^2 = \frac{1}{m_B} \sum_{i=1}^{m_B} \left(\mathbf{x}^{(i)} - \boldsymbol{\mu}_B\right)^2$
3. $\widehat{\mathbf{x}}^{(i)} = \frac{\mathbf{x}^{(i)} - \boldsymbol{\mu}_B}{\sqrt{{\boldsymbol{\sigma}_B}^2 + \varepsilon}}$
4. $\mathbf{z}^{(i)} = \boldsymbol{\gamma} \otimes \widehat{\mathbf{x}}^{(i)} + \boldsymbol{\beta}$

Question 12

Given training and testing sets as showing. Which data augmentations should be used?

Training set consists of blue Ford (left) and Chevrolet (right) cars

The two classes in our hypothetical dataset. The one in the left represents Brand A (Ford), and the one in the right represents Brand B (Chevrolet).

Ford

Ford

Ford

Chevrolet

Chevrolet

Typical examples in testing sets

- ❑ A. Horizontally flipping and center crop
- ❑ B. Vertically flipping and center crop
- ❑ C. Horizontally flipping and color shift
- ❑ D. Vertically flipping and color shift

Question 12

Given training and testing sets as showing. Which data augmentations should be used?

Training set consists of blue Ford (left) and Chevrolet (right) cars

The two classes in our hypothetical dataset. The one in the left represents Brand A (Ford), and the one in the right represents Brand B (Chevrolet).

Ford

Ford

Ford

Chevrolet

Chevrolet

Typical examples in testing sets

- ❑A. Horizontally flipping and center crop
- ❑B. Vertically flipping and center crop
- ❑C. Horizontally flipping and color shift **[x]**
- ❑D. Vertically flipping and color shift