Hyperparameter tuning, Batch Normalization, Programming Frameworks

1.
Question 1

Which of the following are true about hyperparameter search?

0 / 1 point

Expand

Incorrect

Incorrect. Since some problems are highly sensitive to the values of some hyperparameters it is best to use a logarithmic or exponential scale.

2.
Question 2

In a project with limited computational resources, which three of the following hyperparameters would you choose to tune? Check all that apply.

1 / 1 point

Expand

Correct

Great, you got all the right answers.

3.
Question 3

During hyperparameter search, whether you try to babysit one model ("Panda" strategy) or train a lot of models in parallel ("Caviar") is largely determined by:

1 / 1 point

Expand

Correct

4.
Question 4

Knowing that the hyperparameter α should be in the range of 0.00001 0.0001 and 1.0 1.0, which of the following is the recommended way to sample a value for α?

0 / 1 point

Expand

Incorrect

No. This gives a random number between 10^{-4} and 10^{0}.

5.
Question 5

Finding good hyperparameter values is very time-consuming. So typically you should do it once at the start of the project, and try to find very good hyperparameters so that you don't ever have to tune them again. True or false?

1 / 1 point

Expand

Correct

6.
Question 6

In batch normalization as presented in the videos, if you apply it on the lth layer of your neural network, what are you normalizing?

1 / 1 point

Expand

Correct

7.
Question 7

When using normalization:

$$z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}$$

In case σ is too small, the normalization of $z^{(i)}$ may fail since division by 0 may be produced due to rounding errors. True/False?

0 / 1 point

Expand

Incorrect

Incorrect. The normalization formula uses a smoothing parameter ϵ so in $z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}}$ use of the ϵ parameter prevents that the denominator be 0.

8.
Question 8

Which of the following is true about batch normalization?

0 / 1 point

Expand

Incorrect

No. Using this value is equivalent to setting the mean again to μ and the variance to $2+\sigma^2+\epsilon$.

9.
Question 9

A neural network is trained with Batch Norm. At test time, to evaluate the neural network we turn off the Batch Norm to avoid random predictions from the network. True/False?

1 / 1 point

Expand

Correct

Correct. During the test, the parameters μ and $2\sigma^2$ are estimated using an exponentially weighted average across mini-batches used during training.

10.
Question 10

Which of the following are some recommended criteria to choose a deep learning framework?

1 / 1 point

Expand

Correct

Correct. The running speed is a major factor, especially when working with large datasets.

1.
Question 1

Which of the following are true about hyperparameter search?

0 / 1 point

Expand

Incorrect

Incorrect. When we have a large number of hyperparameters it is best to use a random search since this will give a higher number of tested values for each hyperparameter.

2.
Question 2

If it is only possible to tune two parameters from the following due to limited computational resources. Which two would you choose?

1 / 1 point

Expand

Correct

Great, you got all the right answers.

3.
Question 3

During hyperparameter search, whether you try to babysit one model (“Panda” strategy) or train a lot of models in parallel (“Caviar”) is largely determined by:

1 / 1 point

Expand

Correct

4.
Question 4

Knowing that the hyperparameter α should be in the range of 0.001 and 1.0. Which of the following is the recommended way to sample a value for α?

0 / 1 point

Expand

Incorrect

No. This picks a value between 1 and 10^4.

5.
Question 5

Finding new values for the hyperparameters, once we have found good ones for a model, should only be done if new hardware or computational power is acquired. True/False?

0 / 1 point

Expand

Incorrect

Incorrect. As the data changes for the model, it might be beneficial to tune some of the hyperparameters again.

6.
Question 6

In batch normalization as presented in the videos, if you apply it on the lth layer of your neural network, what are you normalizing?

1 / 1 point

Expand

Correct

7.
Question 7

When using normalization:

$$\text{?????}(\text{?})=\text{?}(\text{?})-\text{??}2+\text{?}z_{norm}(i)=\sigma 2+\varepsilon z(i)-\mu$$

In case $\text{?}\sigma$ is too small, the normalization of $\text{?}(\text{?})z(i)$ may fail since division by 0 may be produced due to rounding errors. True/False?

1 / 1 point

Expand

Correct

Correct. The normalization formula uses a smoothing parameter $\text{?}\epsilon$ so in $\text{?}\mathrm{norm}(\text{?})=\text{?}(\text{?})-\text{??}2+\text{?}z_{\mathrm{norm}(i)}=\sigma 2+\epsilon z(i)-\mu$ use of the $\text{?}\epsilon$ parameter prevents that the denominator be 0.

8.
Question 8

Which of the following statements about $\text{?}\gamma$ and $\text{?}\beta$ in Batch Norm are true?

1 / 1 point

Expand

Correct

Great, you got all the right answers.

9.
Question 9

A neural network is trained with Batch Norm. At test time, to evaluate the neural network on a new example you should perform the normalization using $\text{?}\mu$ and $\text{?}2\sigma 2$ estimated using an exponentially weighted average across mini-batches seen during training. True/false?

1 / 1 point

Expand

Correct

Correct. This is a good practice to estimate the $\text{?}\mu$ and $\text{?}2\sigma 2$ to use since at test time we might not be predicting over a batch of the same size, or it might even be a single example, thus using the $\text{?}\mu$ and $\text{?}2\sigma 2$ of a single sample doesn't make sense.

10.
Question 10

Which of these statements about deep learning programming frameworks are true? (Check all that apply)

1 / 1 point

Expand

Correct

Great, you got all the right answers.

1.
Question 1

With a relatively small set of hyperparameters, it is OK to use a grid search. True/False?

0 / 1 point

Expand

Incorrect

Incorrect. When the set of hyperparameters is small like a range for ⬚⬚=1,2,3n_l=1,2,3 grid search works fine.

2.
Question 2

If it is only possible to tune two parameters from the following due to limited computational resources. Which two would you choose?

1 / 1 point

Expand

Correct

Great, you got all the right answers.

3.
Question 3

During hyperparameter search, whether you try to babysit one model ("Panda" strategy) or train a lot of models in parallel ("Caviar") is largely determined by:

1 / 1 point

Expand

Correct

4.
Question 4

If you think ⬚β (hyperparameter for momentum) is between 0.9 and 0.99, which of the following is the recommended way to sample a value for beta?

1 / 1 point

Expand

Correct

5.
Question 5

Once good values of hyperparameters have been found, those values should be changed if new data is added or a change in computational power occurs. True/False?

1 / 1 point

Expand

Correct

Correct. The choice of some hyperparameters such as the batch size depends on conditions such as hardware and quantity of data.

6.
Question 6

In batch normalization as presented in the videos, if you apply it on the lth layer of your neural network, what are you normalizing?

1 / 1 point

Expand

Correct

7.
Question 7

Which of the following are true about batch normalization?

0 / 1 point

Expand

Incorrect

No. The ϵ parameter is used to avoid division by 0.

8.
Question 8

Which of the following are true about batch normalization?

0 / 1 point

Expand

Incorrect

You chose the extra incorrect answers.

9.
Question 9

After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example you should:

1 / 1 point

Expand

Correct

10.
Question 10

If a project is open-source, it is a guarantee that it will remain open source in the long run and will never be modified to benefit only one company. True/False?

1 / 1 point

Expand

Correct

Correct. To ensure that a project will remain open source in the long run it must have a good governance body too.

1.
Question 1

If searching among a large number of hyperparameters, you should try values in a grid rather than random values, so that you can carry out the search more systematically and not rely on chance. True or False?

1 / 1 point

Expand

Correct

2.
Question 2

If it is only possible to tune two parameters from the following due to limited computational resources. Which two would you choose?

1 / 1 point

Expand

Correct

Great, you got all the right answers.

3.
Question 3

Using the "Panda" strategy, it is possible to create several models. True/False?

1 / 1 point

Expand

Correct

Correct. Following the "Panda" analogy, it is possible to babysit a model until a certain point and then start again to produce a different one.

4.
Question 4

Knowing that the hyperparameter α should be in the range of 0.001 and 1.0. Which of the following is the recommended way to sample a value for α?

0 / 1 point

Expand

Incorrect

No. This will pick a random value from a uniform scale, which is not the recommended way to choose α.

5.
Question 5

Once good values of hyperparameters have been found, those values should be changed if new data is added or a change in computational power occurs. True/False?

1 / 1 point

Expand

Correct

Correct. The choice of some hyperparameters such as the batch size depends on conditions such as hardware and quantity of data.

6.
Question 6

When using batch normalization it is OK to drop the parameter $W^{[l]}$ from the forward propagation since it will be subtracted out when we compute $z_{normalize}^{[l]}=\gamma z_{normalize}^{[l]}+\beta^{[l]}$. True/False?

1 / 1 point

Expand

Correct

Correct. The parameter $W^{[l]}$ doesn't get subtracted during the batch normalization process, although it gets re-scaled.

7.
Question 7

In the normalization formula $z_{norm}^{(i)}=\frac{z^{(i)}-\mu}{\sqrt{\sigma^2+\varepsilon}}$, why do we use epsilon?

1 / 1 point

Expand

Correct

8.
Question 8

Which of the following is true about batch normalization?

1 / 1 point

Expand

Correct

Correct. When applying the linear transformation $\tilde{z}^{(l)} = \gamma^{[l]} z^{(l)}_{norm} + \beta^{[l]}$ we set the mean and variance of $\tilde{z}^{[l]}$.

9.
Question 9

A neural network is trained with Batch Norm. At test time, to evaluate the neural network on a new example you should perform the normalization using μ and σ^2 estimated using an exponentially weighted average across mini-batches seen during training. True/false?

0 / 1 point

Expand

Incorrect

Incorrect. This is a good practice to estimate the μ and σ^2 to use since at test time we might not be predicting over a batch of the same size, or it might even be a single example, thus using the μ and σ^2 of a single sample doesn't make sense.

10.
Question 10

If a project is open-source, it is a guarantee that it will remain open source in the long run and will never be modified to benefit only one company. True/False?

1 / 1 point

Expand

Correct

Correct. To ensure that a project will remain open source in the long run it must have a good governance body too.