✓ **축하합니다! 통과하셨습니다!**

받은 학점 80% **최신 제출물 학점** 80% **통과 점수:** 80% 이상

다음 항목으로 이동

---

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

1 / 1점

◉ True

○ False

↗ **더 보기**

✓ **맞습니다**
Correct. When the set of hyperparameters is small like a range for $n_l = 1, 2, 3$ grid search works fine.

---

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

0 / 1점

☑ α

> ✓ **Correct**
> Correct. This might be the hyperparameter that most impacts the results of a model.
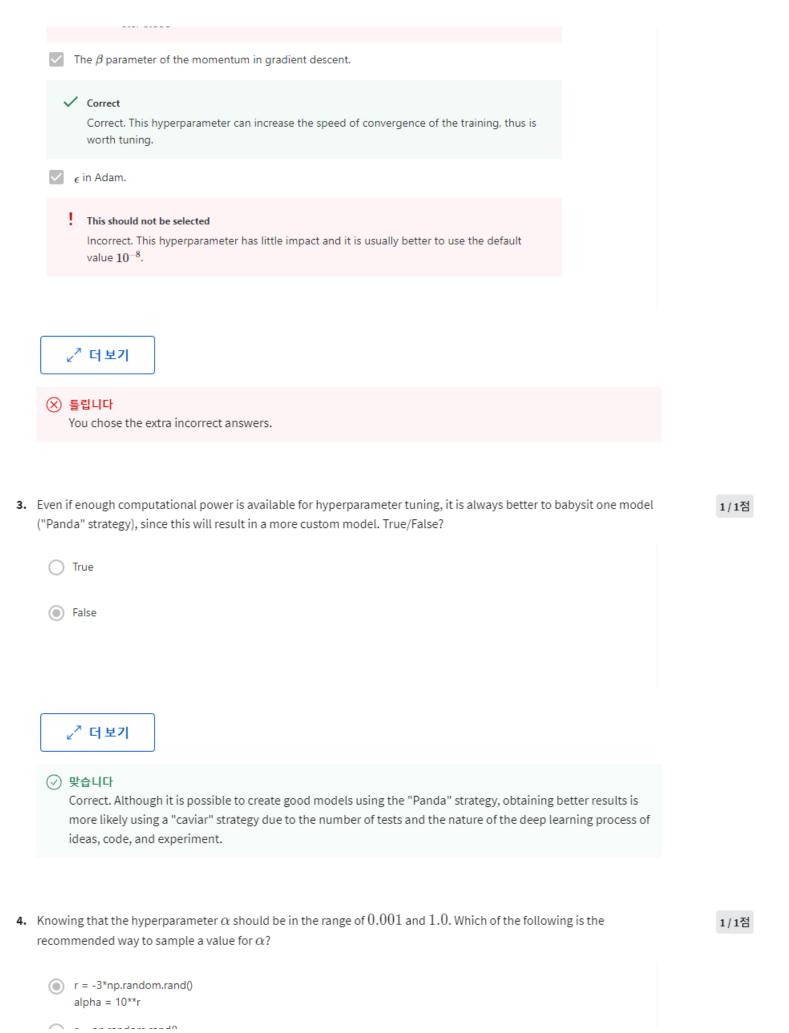
☑ mini-batch size

> ✓ **Correct**
> Correct. This can have a great impact on the results of the cost function, thus it is worth tuning it.

☑ β_1, β_2 in Adam.

> ❗ **This should not be selected**
> Incorrect. This hyperparameter has little impact and it is usually better to use the default values 0.9, 0.999.

The β parameter of the momentum in gradient descent.

✓ **Correct**
Correct. This hyperparameter can increase the speed of convergence of the training, thus is worth tuning.

ϵ in Adam.

! **This should not be selected**
Incorrect. This hyperparameter has little impact and it is usually better to use the default value 10^{-8}.

↗ 더 보기

⊗ **틀립니다**
You chose the extra incorrect answers.

3. Even if enough computational power is available for hyperparameter tuning, it is always better to babysit one model ("Panda" strategy), since this will result in a more custom model. True/False?

   1 / 1점

○ True

◉ False

↗ 더 보기

⊘ **맞습니다**
Correct. Although it is possible to create good models using the "Panda" strategy, obtaining better results is more likely using a "caviar" strategy due to the number of tests and the nature of the deep learning process of ideas, code, and experiment.

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?

   1 / 1점

◉ r = -3*np.random.rand()
   alpha = 10**r

○ r = np.random.rand()
  alpha = 0.001 + r*0.999

○ r = -5*np.random.rand()
  alpha = 10**r

○ r = 4*np.random.rand()
  alpha = 10**r

[↗ 더 보기]

✓ **맞습니다**
   Yes. This gives a random number between $0.001 = 10^{-3}$ and 10^0.

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?

   1/1점

   ○ False

   ● True

[↗ 더 보기]

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

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 $\tilde{z}^{[l]} = \gamma z^{[l]}_{\text{normalize}} + \beta^{[l]}$. True/False?

   0/1점

   ● True

   ○ False

[↗ 더 보기]

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

1/1점

○ In case μ is too small

◉ To avoid division by zero

○ To speed up convergence

○ To have a more accurate normalization

⤢ 더 보기

✓ **맞습니다**

8. Which of the following statements about γ and β in Batch Norm are true?

1/1점

☑ They can be learned using Adam, Gradient descent with momentum, or RMSprop, not just with gradient descent.

✓ **Correct**

☐ The optimal values are $\gamma = \sqrt{\sigma^2 + \varepsilon}$, and $\beta = \mu$.

☑ They set the variance and mean of the linear variable

✓ **Correct**

☐ $$\beta$$ and $$\gamma$$ are hyperparameters of the algorithm, which we tune via random sampling.

☐ There is one global value of $$\gamma \in \Re$$ and one global value of $$\beta \in \Re$$ for each layer, and these apply to all the hidden units in that layer.

Loading [MathJax]/jax/output/CommonHTML/jax.js

⤢ 더 보기

✓ **맞습니다**

Great, you got all the right answers.

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?

- ○ False

- ◉ True

⤢ 더 보기

✓ **맞습니다**
Correct. 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. Which of these statements about deep learning programming frameworks are true? (Check all that apply)

- ☑ Even if a project is currently open source, good governance of the project helps ensure that it remains open even in the long term, rather than become closed or modified to benefit only one company.

  ✓ **Correct**

- ☐ Deep learning programming frameworks require cloud-based machines to run.

- ☑ A programming framework allows you to code up deep learning algorithms with typically fewer lines of code than a lower-level language such as Python.

  ✓ **Correct**

⤢ 더 보기

✓ **맞습니다**
Great, you got all the right answers.