✅ **축하합니다! 통과하셨습니다!**

받은 학점 100%　　최신 제출물 학점 100%　　통과 점수: 80% 이상

다음 항목으로 이동

---

1. Which of the following are true about hyperparameter search?　　`1 / 1점`

- ○ When sampling from a grid, the number of values for each hyperparameter is larger than when using random values.
- ○ When using random values for the hyperparameters they must be always uniformly distributed.
- ○ Choosing values in a grid for the hyperparameters is better when the number of hyperparameters to tune is high since it provides a more ordered way to search.
- ◉ Choosing random values for the hyperparameters is convenient since we might not know in advance which hyperparameters are more important for the problem at hand.

↗ 더 보기

✅ **맞습니다**
Correct. Different problems might be more sensitive to different hyperparameters.

---

2. If it is only possible to tune two parameters from the following due to limited computational resources. Which two would you choose?　　`1 / 1점`

- ☐ β_1, β_2 in Adam.
- ☐ ϵ in Adam.
- ☑ 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.

- ☑ α

  ✓ **Correct**
  Correct. This might be the hyperparameter that most impacts the results of a model.

↗ 더 보기

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

---

3. Using the "Panda" strategy, it is possible to create several models. True/False?　　`1 / 1점`

- ○ False
- ◉ True

↗ 더 보기

✅ **맞습니다**
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. 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 / 1점`

- ◉ r = np.random.rand()
  beta = 1-10**(- r - 1)
- ○ r = np.random.rand()
  beta = 1-10**(- r + 1)
- ○ r = np.random.rand()
  beta = r*0.09 + 0.9
- ○ r = np.random.rand()
  beta = r*0.9 + 0.09

↗ 더 보기

✅ **맞습니다**

---

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`

- ◉ True
- ○ False

↗ 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 $z_{\text{normalize}}^{[l]} = \beta^{[l]} z^{[l]} + \gamma^{[l]}$. True/False?　　`1 / 1점`

- ○ True
- ◉ False

↗ 더 보기

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

7. Which of the following are true about batch normalization?

- ◉ One intuition behind why batch normalization works is that it helps reduce the internal covariance.
- ○ The parameter ϵ in the batch normalization formula is used to accelerate the convergence of the model.
- ○ There is a global value of γ and β that is used for all the hidden layers where batch normalization is used.
- ○ The parameters β and γ of batch normalization can't be trained using Adam or RMS prop.

⤢ 더 보기

⊘ 맞습니다
Yes. Internal covariance is a name to express that there has been a change in the distribution of the activations. Since after each iteration of gradient descent the parameters of a layer change, we might think that the activations suffer from covariance shift.

---

8. Which of the following are true about batch normalization?

- ☑ When using batch normalization we introduce two new parameters $\gamma^{[l]}$, $\beta^{[l]}$ that must be "learned" or trained.

  ✓ **Correct**
  Correct. Batch normalization uses two parameters β and γ to compute $\tilde{z}^{(i)} = \beta z_{norm}^{(i)} + \gamma$.

- ☐ $z_{norm}^{(i)} = \dfrac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.

- ☐ $\beta^{[l]}$ and $\gamma^{[l]}$ are hyperparameters that must be tuned by random sampling in a logarithmic scale.

- ☑ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the mean and variance of $\tilde{z}^{[l]}$.

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

⤢ 더 보기

⊘ 맞습니다
Great, you got all the right answers.

---

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

- ○ Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized.
- ◉ Perform the needed normalizations. use μ and σ^2 estimated using an exponentially weighted average across mini-batches seen during training.
- ○ If you implemented Batch Norm on mini-batches of (say) 256 examples. then to evaluate on one test example. duplicate that example 256 times so that you're working with a mini-batch the same size as during training.
- ○ Use the most recent mini-batch's value of μ and σ^2 to perform the needed normalizations.

⤢ 더 보기

⊘ 맞습니다

---

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?

- ○ True
- ◉ False

⤢ 더 보기

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