Your grade: 95%

Your latest: **95%** • Your highest: **95%** • To pass you need at least 80%. We keep your highest score.

Next item →

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

- ○ When using random values for the hyperparameters they must be always uniformly distributed.
- ○ When sampling from a grid, the number of values for each hyperparameter is larger than when using random values.
- ◉ 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.
- ○ 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.

Correct
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?

- ☐ The β parameter of the momentum in gradient descent.
- ☑ α

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

- ☐ ϵ in Adam.
- ☑ 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 $0.9, 0.999$.

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?

- ◉ False
- ○ True

Correct
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. 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?

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

Correct

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?

- ◉ False
- ○ True

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

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?

- ○ True
- ◉ False

Correct
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.
- ○ The parameters β and γ of batch normalization can't be trained using Adam or RMS prop.
- ○ There is a global value of γ and β that is used for all the hidden layers where batch normalization is used.

Correct
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 is true about batch normalization?

- ○ The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$.
- ○ $z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.
- ◉ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the variance and mean of $\tilde{z}^{[l]}$.
- ○ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ can be learned only using plain gradient descent.

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

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

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

- ○ It must run exclusively on cloud services, to ensure its robustness.
- ◉ Running speed.
- ○ It must be implemented in C to be faster.
- ○ It must use Python as the primary language.

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