Congratulations! You passed!

Grade received 90% | Latest Submission Grade 90% | To pass 80% or higher

Go to next item

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

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

Expand

Correct
Correct. Different problems might be more sensitive to different hyperparameters.

2. Every hyperparameter, if set poorly, can have a huge negative impact on training, and so all hyperparameters are about equally important to tune well. True or False? **1 / 1 point**

- True
- (selected) False

Expand

Correct
Yes. We've seen in the lecture that some hyperparameters, such as the learning rate, are more critical than others.

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

- False
- (selected) True

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. Knowing that the hyperparameter α should be in the range of 0.00001 and 1.0, which of the following is the recommended way to sample a value for α? **0 / 1 point**

- r = np.random.rand()
  alpha = 0.00001 + r*0.99999
- r = np.random.rand()
  alpha = 10**r
- (selected) r = -4*np.random.rand()

alpha = 10**r

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

Expand

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

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**

True

False

Expand

Correct

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 $\tilde{z}^{[l]}_{\text{normalize}} = \beta^{[l]}\,\hat{z}^{[l]} + \gamma^{[l]}$. True/False? **1 / 1 point**

True

False

Expand

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? **1 / 1 point**

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.

Expand

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? **1 / 1 point**

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

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

Expand

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

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**

- True
- False

Expand

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

10. Which of the following are some recommended criteria to choose a deep learning framework? **1 / 1 point**

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

Expand

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