Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

Next item →

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

- (•) False
- ( ) True

Correct

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

- (•) False
- ( ) True

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

- (•) True
- ( ) False

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 0.001 and 1.0. Which of the following is the recommended way to sample a value for α? — 1 / 1 point

- ( )
```
r = -5*np.random.rand()
alpha = 10**r
```
- ( )
```
r = np.random.rand()
alpha = 0.001 + r*0.999
```
- ( )
```
r = 4*np.random.rand()
alpha = 10**r
```
- (•)
```
r = -3*np.random.rand()
alpha = 10**r
```

Correct
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 / 1 point

- (•) True
- ( ) False

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

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

- ( ) $W^{[l]}$
- (•) $z^{[l]}$
- ( ) $b^{[l]}$
- ( ) $a^{[l]}$

Correct

7. Which of the following are true about batch normalization? — 1 / 1 point

- ( ) The parameter ϵ in the batch normalization formula is used to accelerate the convergence of the model.
- (•) One intuition behind why batch normalization works is that it helps reduce the internal covariance.
- ( ) 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

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.

1 / 1 point

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

- [x] 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 $\tilde{z}^{(i)} = \beta z_{norm}^{(i)} + \gamma$.

- [ ] $\beta^{[l]}$ and $\gamma^{[l]}$ are hyperparameters that must be tuned by random sampling in a logarithmic scale.
- [x] The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the variance and mean of $\tilde{z}^{[l]}$.

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

- [ ] $z_{norm}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.

1 / 1 point

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

- ( ) 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.
- ( ) Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized.
- ( ) Use the most recent mini-batch's value of μ and σ^2 to perform the needed normalizations.
- (•) Perform the needed normalizations, use μ and σ^2 estimated using an exponentially weighted average across mini-batches seen during training.

  Correct

1 / 1 point

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

- [x] 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.
- [x] 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