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

- (●) 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 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 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.

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
- ◉ False

Expand

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

3. During hyperparameter search, whether you try to babysit one model ("Panda" strategy) or train a lot of models in parallel ("Caviar") is largely determined by: 1 / 1 point

- ○ The presence of local minima (and saddle points) in your neural network
- ○ The number of hyperparameters you have to tune
- ○ Whether you use batch or mini-batch optimization
- ◉ The amount of computational power you can access

Expand

✓ Correct

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?

0 / 1 point

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

Expand

Incorrect

No. This will pick a random value from a uniform scale, which is not the recommended way to choose α.

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?

1 / 1 point

- ○ True
- ◉ False

Expand

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?

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. In the normalization formula $z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}$, why do we use epsilon? 1 / 1 point

- In case μ is too small
- To avoid division by zero
- To have a more accurate normalization
- To speed up convergence

Expand

Correct

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

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

- [ ] $z^{(i)}_{norm} = \dfrac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.
- [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 $\tilde{z}^{(i)} = \beta z^{(i)}_{norm} + \gamma$.

Expand

✓ **Correct**
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?

1 / 1 point

- (●) True
- ( ) False

Expand

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 these statements about deep learning programming frameworks are true? (Check all that apply) 1 / 1 point

- [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**

- [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.

Expand

Correct
Great, you got all the right answers.