1. With a relatively small set of hyperparameters, it is OK to use a grid search. True/False? 1 / 1 point

- (●) True
- ( ) False

Correct
Correct. When the set of hyperparameters is small like a range for $n_l = 1, 2, 3$ grid search works fine.

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

- [x] α

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

- [x] 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.

- [ ] β_1, β_2 in Adam.
- [ ] ϵ in Adam.

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

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? **1 / 1 point**

○
```
r = 4*np.random.rand()
alpha = 10**r
```

○
```
r = np.random.rand()
alpha = 0.001 + r*0.999
```

◉
```
r = -3*np.random.rand()
alpha = 10**r
```

○
```
r = -5*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

- False
- True

Correct

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 $b^{[l]}$ from the forward propagation because it is effectively canceled out during the normalization step, where we compute $z^{[l]}_{\text{norm}} = \frac{z^{[l]} - \mu}{\sigma}$. True/False? 1 / 1 point

- False
- True

Correct

Yes! The bias $b^{[l]}$ is subtracted out during the computation of the normalized value $z^{[l]}_{\text{norm}}$, making it unnecessary in the context of batch normalization.

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

- ○ 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.
- ◉ 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.

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 statements about γ and β in Batch Norm are true? 1 / 1 point

- ☐ There is one global value of $\gamma \in \Re$ and one global value of $\beta \in \Re$ for each layer, and these apply to all the hidden units in that layer.
- ☑ They can be learned using Adam, Gradient descent with momentum, or RMSprop, not just with gradient descent.

Correct

- ☑ They set the variance and mean of the linear variable $\tilde{z}^{[l]}$ of a given layer.

Correct

- ☐ The optimal values are $\gamma = \sqrt{\sigma^2 + \varepsilon}$, and $\beta = \mu$.
- ☐ β and γ are hyperparameters of the algorithm, which we tune via random sampling.

9. After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example you should: 1 / 1 point

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

✓ **Correct**

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

- ☑ 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**

- ☑ 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**

- ☐ Deep learning programming frameworks require cloud-based machines to run.