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

  1 / 1 point

  ○ False

  ⦿ True

  ⤢ **Expand**

  ✓ **Correct**
  Correct. When the set of hyperparameters is small like a range for $$n\_l = 1, 2, 3$$ grid search works fine.

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

  ⤢ **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 amount of computational power you can access

  ○ Whether you use batch or mini-batch optimization

  ○ The presence of local minima (and saddle points) in your neural network

  ○ The number of hyperparameters you have to tune

  ⤢ **Expand**

---

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?

○ r = np.random.rand()
alpha = 10**r

○ r = np.random.rand()
alpha = 0.00001 + r*0.99999

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

◉ r = -4*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?

◉ False

○ True

⤢ **Expand**

✓ **Correct**

---

6. When using batch normalization it is OK to drop the parameter $b^{[l]}$ 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?

○ False

◉ True

⤢ **Expand**

✓ **Correct**
Correct. Since in the normalization process the values of $$z^{[l]}$$ are re-centered at the origin, it is irrelevant to add the $$b^{[l]}$$ parameter.

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

○ 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.

⊙ One intuition behind why batch normalization works is that it helps reduce the internal covariance.

⤢ **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?

⊙ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the mean and variance of $\tilde{z}^{[l]}$.

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

○ The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$.

○ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ 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 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

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.

○ It must be implemented in C to be faster.

⦿ Running speed.

○ It must run exclusively on cloud services, to ensure its robustness.

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