Your grade: 100%

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

Next item →

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

- ○ When sampling from a grid, the number of values for each hyperparameter is larger than when using random values.
- ◉ 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.
- ○ 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.

✓ **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

- ◉ False
- ○ True

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

3. Even if enough computational power is available for hyperparameter tuning, it is always better to babysit one model ("Panda" strategy), since this will result in a more custom model. True/False? 1 / 1 point

- ○ True
- ◉ False

✓ **Correct**
Correct. Although it is possible to create good models using the "Panda" strategy, obtaining better results is more likely using a "caviar" strategy due to the number of tests and the nature of the deep learning process of ideas, code, and experiment.

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? 1 / 1 point

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

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

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

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

✓ **Correct**
Yes. This will generate a random value between 10^{-5} and 10^{0} chosen randomly in a logarithmic scale.

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

- (•) True
- ( ) False

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

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

Correct

8. Which of the following statements about γ and β in Batch Norm are true? 1 / 1 point

- [ ] The optimal values are $\gamma = \sqrt{\sigma^2 + \varepsilon}$, and $\beta = \mu$.
- [x] They set the variance and mean of the linear variable $\tilde{z}^{[l]}$ of a given layer.

Correct

- [ ] 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.
- [x] They can be learned using Adam, Gradient descent with momentum, or RMSprop, not just with gradient descent.

Correct

- [ ] β and γ are hyperparameters of the algorithm, which we tune via random sampling.

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

- ( ) False
- (•) True

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 the following are some recommended criteria to choose a deep learning framework? 1 / 1 point

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

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