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

- (selected) 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 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.
- When using random values for the hyperparameters they must be always uniformly distributed.

Correct
Correct. Different problems might be more sensitive to different hyperparameters.

2. In a project with limited computational resources, which three of the following hyperparameters would you choose to tune? Check all that apply.

1 / 1 point

- [ ] ϵ in Adam.
- [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.

- [x] α

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

- [x] mini-batch size

Correct
Correct. This can have a great impact on the results of the cost function, thus it is worth tuning it.

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

?

- [x]

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

alpha = 10**r
```

- [ ]

```
r = np.random.rand()

alpha = 0.00001 + r*0.99999
```

- [ ]

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

alpha = 10**r
```

- [ ]

```
r = 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. 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?

1 / 1 point

- [x] False
- [ ] True

Correct

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 $z_{\text{norm}}^{[l]} = \frac{z^{[l]} - \mu}{\sigma}$. True/False?

1 / 1 point

- ◯ False
- ◉ True

Correct
Yes! The bias $b^{[l]}$ is subtracted out during the computation of the normalized value $z_{\text{norm}}^{[l]}$, 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.
- ◉ One intuition behind why batch normalization works is that it helps reduce the internal covariance.
- ◯ There is a global value of γ and β that is used for all the hidden layers where batch normalization is used.
- ◯ 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.

1 / 1 point

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

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

- [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 $\tilde{z}^{[l]}$.

- [ ] $\beta^{[l]}$ and $\gamma^{[l]}$ are hyperparameters that must be tuned by random sampling in a logarithmic scale.

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

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 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