## Congratulations! You passed!

Grade received 98% Latest Submission Grade 98% To pass 80% or higher

Go to next item

**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.** In a project with limited computational resources, which three of the following hyperparameters would you choose to tune? Check all that apply. 0.8 / 1 point

- [ ] The $\beta$ parameter of the momentum in gradient descent.
- [ ] $\epsilon$ in Adam.
- [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.

- [ ] $\beta_1, \beta_2$ in Adam.
- [x] $\alpha$

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

You didn't select all the correct answers

**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

Correct

**4.** If you think $\beta$ (hyperparameter for momentum) is between 0.9 and 0.99, which of the following is the recommended way to sample a value for beta? 1 / 1 point

- ( )
  r = np.random.rand() beta = r*0.9 + 0.09
- (•)
  r = np.random.rand() beta = 1-10**(- r - 1)
- ( )
  r = np.random.rand() beta = 1-10**(- r + 1)
- ( )
  r = np.random.rand() beta = r*0.09 + 0.9

Correct

**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

- ( ) True
- (•) False

Correct

**6.** In batch normalization as presented in the videos, if you apply it on the $l$th layer of your neural network, what are you normalizing? 1 / 1 point

- (•) $z^{[l]}$
- ( ) $b^{[l]}$
- ( ) $W^{[l]}$
- ( ) $a^{[l]}$

Correct

**7.** When using normalization:

$z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}$

In case $\sigma$ is too small, the normalization of $z^{(i)}$ may fail since division by 0 may be produced due to rounding errors. True/False? 1 / 1 point

- ( ) True
- (•) False

Correct
Correct. The normalization formula uses a smoothing parameter $\epsilon$ so in $z^{(i)}_{\text{norm}} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}}$ use of the $\epsilon$ parameter prevents that the denominator be 0.

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

- [ ] The optimal values to use for $\gamma$ and $\beta$ are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$.
- [ ] The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ can be learned only using plain gradient descent.
- [x] The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the mean and variance of $\tilde{z}^{[l]}$.

Correct
Correct. When applying the linear transformation $\tilde{z}^{(i)} = \beta^{[l]} z^{(i)}_{norm} + \gamma^{[l]}$ we set the mean and variance of $\tilde{z}^{[l]}$.

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

**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

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

Correct

**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.