Your grade: 90%

Your latest: **90%** • Your highest: **90%** • To pass you need at least 80%. We keep your highest score.

Next item →

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. In a project with limited computational resources, which three of the following hyperparameters would you choose to tune? Check all that apply. 0 / 1 point

- [x] β_1, β_2 in Adam.

  This should not be selected

  Incorrect. This hyperparameter has little impact and it is usually better to use the default values 0.9, 0.999.

- [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.

- [ ] The β parameter of the momentum in gradient descent.

- [x] α

  Correct

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

- [ ] ϵ in Adam.

Incorrect

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 presence of local minima (and saddle points) in your neural network
- Whether you use batch or mini-batch optimization
- The number of hyperparameters you have to tune
- The amount of computational power you can access

Expand

Correct

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

Expand

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

Expand

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 $W^{[l]}$ from the forward propagation since it will be subtracted out when we compute $\tilde{z}^{[l]} = \gamma z^{[l]}_{\text{normalize}} + \beta^{[l]}$. True/False? 1 / 1 point

- (•) False
- ( ) True

Expand

Correct

Correct. The parameter $W^{[l]}$ doesn't get subtracted during the batch normalization process, although it gets re-scaled.

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

Expand

Correct

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

- [ ] $z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.
- [ ] $\beta^{[l]}$ and $\gamma^{[l]}$ are hyperparameters that must be tuned by random sampling in a logarithmic scale.
- [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 $\tilde{z}^{[l]}$.

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

Expand

Correct
Great, you got all the right answers.

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

- (●) 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.
- ( ) 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.
- ( ) Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized.

Expand

Correct

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

- [ ] 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**

- [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**

Expand

Correct
Great, you got all the right answers.