Congratulations! You passed!

Grade received 90% **Latest Submission Grade** 90% **To pass** 80% or higher

Go to next item

1. If searching among a large number of hyperparameters, you should try values in a grid rather than random values, so that you can carry out the search more systematically and not rely on chance. True or False? **1 / 1 point**

- (•) False
- ( ) True

Expand

Correct

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] α

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

- [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.

- [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.

- [ ] β_1, β_2 in Adam.

Expand

Correct
Great, you got all the right answers.

3. Using the "Panda" strategy, it is possible to create several models. True/False? **0 / 1 point**

- (•) False
- ( ) True

Expand

Incorrect
Incorrect. Following the "Panda" analogy, it is possible to babysit a model until a certain point and then start again to produce a different one.

4. If you think β (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*0.09 + 0.9
- 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)

Expand

Correct

5. Finding new values for the hyperparameters, once we have found good ones for a model, should only be done if new hardware or computational power is acquired. True/False? 1 / 1 point

- False
- True

Expand

Correct
Correct. As the data changes for the model, it might be beneficial to tune some of the hyperparameters again.

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 $\tilde{z}^{[l]} = \gamma z^{[l]}_{\text{normalize}} + \beta^{[l]}$. True/False? 1 / 1 point

- True
- False

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

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

Expand

Correct

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

- There is one global value of $\gamma \in \mathfrak{R}$ and one global value of $\beta \in \mathfrak{R}$ for each layer, and these apply to all the hidden units in that layer.
- They set the variance and mean of the linear variable $\tilde{z}^{[l]}$ of a given layer.

  Correct
- The optimal values are $\gamma = \sqrt{\sigma^2 + \varepsilon}$, and $\beta = \mu$.
- β and γ are hyperparameters of the algorithm, which we tune via random sampling.
- They can be learned using Adam, Gradient descent with momentum, or RMSprop, not just with gradient descent.

  Correct

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

- Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized.
- Use the most recent mini-batch's value of μ and σ^2 to perform the needed normalizations.
- Perform the needed normalizations, use μ and σ^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.

Expand

Correct

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 use Python as the primary language.
- ○ It must run exclusively on cloud services, to ensure its robustness.
- ◉ Running speed.

Expand

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