Hyperparameter tuning, Batch Normalization, Programming Frameworks

10/10 points (100%)

Quiz, 10 questions



Next Item



1/1 points

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?



True



False

Correct



1/1 points

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?



True



False

Correct

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



1/1 points

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:



Whether you use batch or mini-batch optimization

Hyperparameterptuning, Batch Normalization, in Programming Frameworks

10/10 points (100%)

Quiz, 10 questions

The amount of computational power you can access

The number of hyperparameters you have to tune



1/1 points

4.

If you think β (hyperparameter for momentum) is between on 0.9 and 0.99, which of the following is the recommended way to sample a value for beta?

```
1 r = np.random.rand()
2 beta = r*0.09 + 0.9
```



Correct

```
1 r = np.random.rand()
2 beta = 1-10**(- r + 1)
```

```
1 r = np.random.rand()
2 beta = r*0.9 + 0.09
```



1/1 points

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 revisit tuning them again. True or false?





False

Correct

Hyperparameter tuning, Batch Normalization, Programming Frameworks

10/10 points (100%)

Quiz, 10 questions



1/1 points

6.

In batch normalization as presented in the videos, if you apply it on the $\it l$ th layer of your neural network, what are you normalizing?



 $z^{[l]}$

Correct



 $W^{[l]}$

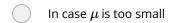




1/1 points

7.

In the normalization formula $z_{norm}^{(i)}=rac{z^{(i)}-\mu}{\sqrt{\sigma^2+arepsilon}}$, why do we use epsilon?

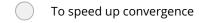




To avoid division by zero

Correct

To have a more accurate normalization





1/1 points

8.

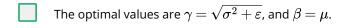
Which of the following statements about γ and β in Batch Norm are true?

There is one global value of $\gamma \in \Re$ and one global value of $\beta \in \Re$ for each layer, and applies to all the hidden units in that layer.

Hyperparameter tuning, Batch Normalization, Programming Frameworks, Selected is correct

10/10 points (100%)

Quiz, 10 questions



Un-selected is correct



Correct

They set the mean and variance of the linear variable $z^{[l]}$ of a given layer.

Correct

eta and γ are hyperparameters of the algorithm, which we tune via random sampling.

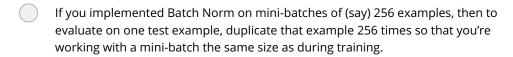
Un-selected is correct

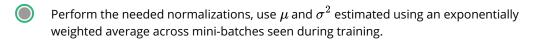


1 / 1 points

9.

After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example you should:





Correct

Use the most recent mini-batch's value of μ and σ^2 to perform the needed
normalizations.

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

Hyperparameter tuning, Batch Normalization, Programming Frameworks

10/10 points (100%)

Quiz, 10 questi Which of these statements about deep learning programming frameworks are true? (Check all that apply)

