

✔ Congratulations! You passed!

Grade
received 80%

Latest Submission
Grade 80%

To pass 80% or
higher

Go to next item

1. Which of the following are true about hyperparameter search?

1 / 1 point

- ☒ 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.
- ☐ 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 sampling from a grid, the number of values for each hyperparameter is larger than when using random values.
- ☐ When using random values for the hyperparameters they must be always uniformly distributed.

↶ ↗ Expand

✔ Correct

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

2. If it is only possible to tune two parameters from the following due to limited computational resources. Which two would you choose?

1 / 1 point

☒ α

✔ Correct

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

☐ ϵ in Adam.

☒ 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.

↶ ↗ 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?

0 / 1 point

- ☐ $r = \text{np.random.rand}()$
 $\text{beta} = r * 0.09 + 0.9$
- ☐ $r = \text{np.random.rand}()$
 $\text{beta} = 1 - 10^{**}(-r - 1)$
- ☐ $r = \text{np.random.rand}()$
 $\text{beta} = r * 0.9 + 0.09$
- ☒ $r = \text{np.random.rand}()$
 $\text{beta} = 1 - 10^{**}(-r + 1)$

[Expand](#)

✗ Incorrect

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

- ☒ True
- ☐ False

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

1 / 1 point

- ☐ True
- ☒ False

[Expand](#)

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

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

1 / 1 point

- ☐ The parameter ϵ in the batch normalization formula is used to accelerate the convergence of the model.
- ☐ There is a global value of γ and β that is used for all the hidden layers where batch normalization is used.
- ☐ 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.

[Expand](#)

✓ 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.

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

1 / 1 point

- ☐ $\beta^{[l]}$ and $\gamma^{[l]}$ are hyperparameters that must be tuned by random sampling in a logarithmic scale.
- ☒ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the variance and mean of $\tilde{z}^{[l]}$.

✓ **Correct**

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

- ☒ 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 $z^{(i)} = \beta z_{norm}^{(i)} + \gamma$.

- ☐ $z_{norm}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.

↶ Expand

✓ **Correct**

Great, you got all the right answers.

9. A neural network is trained with Batch Norm. At test time, to evaluate the neural network we turn off the Batch Norm to avoid random predictions from the network. True/False?

1 / 1 point

- ☒ False
- ☐ True

↶ Expand

✓ **Correct**

Correct. During the test, the parameters μ and σ^2 are estimated using an exponentially weighted average across mini-batches used during training.

10. If a project is open-source, it is a guarantee that it will remain open source in the long run and will never be modified to benefit only one company. True/False?

1 / 1 point

- ☐ True
- ☒ False

↶ Expand

✓ **Correct**

Correct. To ensure that a project will remain open source in the long run it must have a good governance body too.