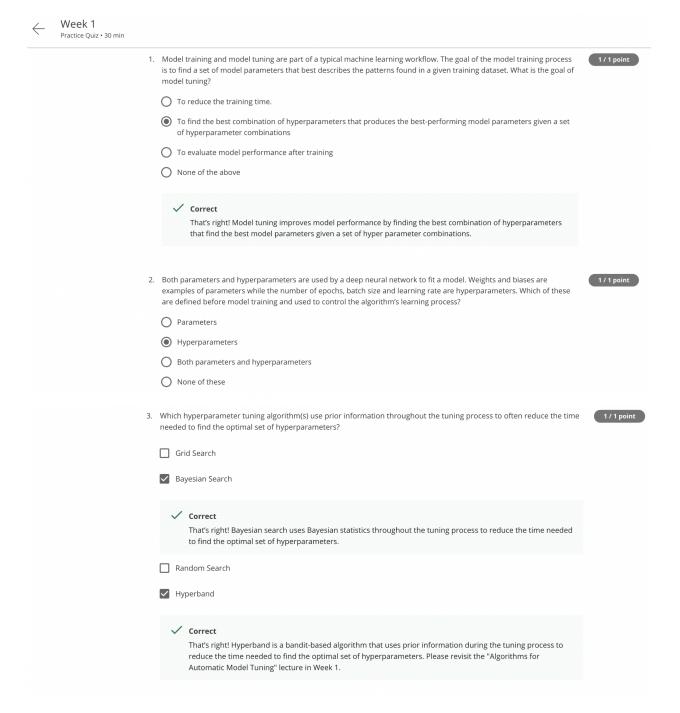
Practical Data Science Specialization



Practical Data Science Specialization

4. Ben needs to deliver the first version of an NLP text classifier with a large number of hyperparameters by the end of the

1 / 1 point

day. Testing all hyperparameter combinations would require many days to compare all combinations. He can either use the Grid Search or Random Search algorithm to perform the hyperparameter tuning on the model.	
Which search algorithm would you recommend to Ben to deliver the model by the end of the day?	
Either, both algorithms will take the same amount of time to produce the results	
Combine the two algorithms	
Grid Search	
Random Search	
Correct Random search is the best option here given the time constraint. In random search, only a limited number of random hyperparameter combinations are tested. Unlike in Grid Search where every combination of hyperparameters is tested. In this case, a full Grid Search would take several days which does not fit within the time constraint.	
5. Which of the following steps are required to perform hyperparameter tuning with a SageMaker hyperparameter Tuning (HPT) Job:	1 / 1 point
Analyze results and select the best model candidate	
Start a SageMaker HPT Job	
Create an Estimator	
O Define the combination of hyperparameters to use for the HPT Job	
All of the above	
Correct Correct! We must first create an Estimator and define the combination of hyperparameters to use for the HPT Job. We then start the SageMaker HPT Job. Finally, we can analyze the results of the tuning job and select the best model candidate.	
 Amazon SageMaker hyperparameter Tuning supports a "warm start" feature that reuses the results from a previous hyperparameter tuning job to speed up subsequent tuning jobs. Warm start helps reduce the overall tuning time. In which of the following scenarios is the warm start feature particularly useful? (Select all that apply) 	1 / 1 point
Adding new hyperparameters	
 Correct Correct! The model can benefit from the knowledge of previous tuning jobs to more quickly find the best hyperparameters in subsequent tuning jobs. 	
Removing all hyperparameters	
✓ Changing hyperparameter tuning ranges	
 Correct Correct! The model can benefit from the knowledge of a previous tuning job to more quickly find the best hyperparameters in subsequent tuning jobs. 	
☐ None of the above	

Practical Data Science Specialization

7.	One week after tuning an NLP text classifier, you collect more data. Luckily, you saved the previous tuning job and its results. You now want to re-run this tuning job with the additional data. Which type of warm start would you use to accomplish this with a SageMaker hyperparameter Tuning (HPT) Job?	1 / 1 point
	TRANSFER_LEARNING	
	O IDENTICAL_DATA_AND_ALGORITHM	
	Can't use warm start in this scenario	
	Correct That's right! With a transfer learning approach you can use updated training data and a different version of the training algorithm with the previous job.	
8.	Amazon SageMaker Managed Spot Training takes advantage of checkpointing which allows you to resume training of your ML model on the next spot instance after an interruption. Which of the following information is included in a checkpoint file? (Select all that apply)	1 / 1 point
	✓ Training configuration	
	✓ Correct Correct! The training configuration is included in the checkpoint file.	
	✓ Model weights	
	 ✓ Correct Correct! The checkpoint file includes the model weights learned up to that point. 	
	Model accuracy	
	✓ Model architecture	
	 Correct Correct! The model architecture is included in the checkpoint file. 	

Hong Hanh

Practical Data Science Specialization

 While training a machine learning model, you realize that the training job is running out of memory, what are some strategies you can use to solve this problem? (Select all that apply) 	1 / 1 point
Reduce the number of epochs	
Reduce the batch size	
 Correct Correct! The model may fit in memory if the batch size is decreased. 	
Reduce the model input size	
Correct Correct! This depends on the data type. For image data, you can reduce the image size or resolution; for NLP, you can reduce the sequence length; for tabular data, you can reduce the dimensions of the embedding vectors.	
☐ Increase the number of model parameters	
10. Distributed training is a commonly used technique to speed up training when dealing with large datasets, complex models, and other scale challenges. Data parallelism is a distributed training strategy where the training data is split up across multiple devices or nodes.	1 / 1 point
True/False: In data parallelism, different models are used on all devices or nodes.	
False	
○ True	