Congratulations! You passed!

Grade received 100% **To pass** 80% or higher

Go to next item

1. Model training and model tuning are part of a typical machine learning workflow. The goal of the model training process is to find a set of model parameters that best describes the patterns found in a given training dataset. What is the goal of model tuning? **1 / 1 point**

- ○ To reduce the training time.
- ◉ To find the best combination of hyperparameters that produces the best-performing model parameters given a set of hyperparameter combinations
- ○ To evaluate model performance after training
- ○ None of the above

Correct
That's right! Model tuning improves model performance by finding the best combination of hyperparameters that find the best model parameters given a set of hyper parameter combinations.

2. Both parameters and hyperparameters are used by a deep neural network to fit a model. Weights and biases are examples of parameters while the number of epochs, batch size and learning rate are hyperparameters. Which of these are defined before model training and used to control the algorithm's learning process? **1 / 1 point**

- ○ Parameters
- ◉ Hyperparameters
- ○ Both parameters and hyperparameters
- ○ None of these

Correct
Correct! hyperparameters are used as inputs to control how the algorithm learns the model parameters during the training process.

3. Which hyperparameter tuning algorithm(s) use prior information throughout the tuning process to often reduce the time needed to find the optimal set of hyperparameters? **1 / 1 point**

- ☑ Bayesian Search

Correct
That's right! Bayesian search uses Bayesian statistics throughout the tuning process to reduce the time needed to find the optimal set of hyperparameters.

- ☑ Hyperband

Correct
That's right! Hyperband is a bandit-based algorithm that uses prior information during the tuning process to reduce the time needed to find the optimal set of hyperparameters. Please revisit the "Algorithms for Automatic Model Tuning" lecture in Week 1.

- ☐ Grid Search
- ☐ Random Search

4. Ben needs to deliver the first version of an NLP text classifier with a large number of hyperparameters by the end of the 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. **1 / 1 point**

Which search algorithm would you recommend to Ben to deliver the model by the end of the day?

- ○ Grid Search
- ○ Either, both algorithms will take the same amount of time to produce the results
- ○ Combine the two algorithms
- ◉ 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
- ○ 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.

6. 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

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**

- ○ IDENTICAL_DATA_AND_ALGORITHM
- ◉ TRANSFER_LEARNING
- ○ 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.

9. 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 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.

- ☐ Reduce the number of epochs
- ☑ Reduce the batch size

Correct
Correct! The model may fit in memory if the batch size is decreased.

- ☐ 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.

- ○ True
- ◉ False

Correct
That's right! In data parallelism the model is replicated on all devices or nodes, then each device or node processes a batch of data. The results from all of the devices or nodes are then combined together. Model parallelism splits the model across different devices or nodes.