Practical Data Science Specialization

1. In order to train an NLP text classifier, you split your data into train, validation, and test datasets. What is the purpose of the validation set?   **1 / 1 point**

   ○ To train the model

   ● To evaluate and tune the model hyper-parameters used during model training

   ○ To evaluate model performance after training

   ○ None of the above

   > ✓ **Correct**
   > That's right! The validation dataset is used to evaluate the model performance during training and can be used to tune the hyper-parameters of the model.

2. You decide to use a BERT pre-trained model to train your text classifier. This model will classify product reviews into positive, neutral, and negative sentiments. Which layers would you typically add to fine tune this type of model to our reviews dataset?   **1 / 1 point**

   ○ Activation Layers

   ○ Pre-trained Layers

   ● Classifier Layers

   ○ Convolutional Layers

3. Transfer learning refers to reusing the knowledge learned from one task (pre-trained model) for another task.   **1 / 1 point**

   Consider this statement: "Fine-tuning is usually faster than pre-training from scratch" and indicate if it is True or False.

   ● True

   ○ False

   > ✓ **Correct**
   > Correct! Unlike pre-training, fine-tuning a model is generally faster than pre-training from scratch since the fine-tuned model is typically trained on a much smaller dataset compared to pre-training.

4. During training from scratch, BERT uses word masking and next sentence prediction in parallel to learn and understand language. In this context, what is the purpose of word masking?   **1 / 1 point**

   ● To learn the surrounding words of each sentence

   ○ To hide words that may decrease model performance

   ○ To predict the next sentence

   ○ To remove "stop" words

   > ✓ **Correct**
   > That's right! Masking forces the model to learn the surrounding words of each sentence

5. **1 / 1 point**

Fine-tuning a pre-trained NLP model has several benefits as mentioned in the course. What is the ideal dataset for fine-tuning?

☑ When they both share a similar vocabulary

> ✓ **Correct**
> Correct! Fine-tuning should be applied to datasets with similar vocabularies.

☑ When they both share a similar language representation for a language (German) or domain (Amazon.com product catalog)

> ✓ **Correct**
> Correct! If 2 models share a similar language representation then fine tuning can work. For example a model trained on that contains German Wikipedia documents can be used to train German product reviews.

☐ Any dataset can be used to fine-tune a model from a pre-trained model

☐ When the pre-trained dataset is exactly the same as the fine-tuning dataset

6. While training an NLP classifier to classify product reviews, you decide to set BERT's maximum sequence length to 100. What is the impact of this particular choice? **1 / 1 point**

☑ Longer reviews with more than 100 tokens will be truncated down to 100.

> ✓ **Correct**
> Correct! The maximum sequence length defines the maximum number of tokens passed to an NLP model per sample.

☐ Shorter reviews with less than 100 tokens will be automatically deleted

☐ Longer reviews with more than 100 tokens will be automatically deleted

☑ Shorter reviews with less than 100 tokens will be padded up to a length of 100.

> ✓ **Correct**
> Correct! The maximum sequence length defines the maximum number of tokens passed to an NLP model per sample.

7. "Amazon SageMaker provides a tool that can send you an SMS if your model starts overfitting during training". Is the statement accurate? **1 / 1 point**

◉ Yes

○ No

8. Amazon SageMaker provides a collection of pre-built solutions and access to pre-trained models for popular machine learning use cases and tasks which can be deployed very easily with Amazon SageMaker Studio. Where can one find these pre-trained models within SageMaker Studio? **1 / 1 point**

◉ Amazon SageMaker JumpStart

○ Amazon S3

○ Amazon Redshift

○ AWS Glue