Questionnaire (2nd chapter)

Total de pontos 0/0

O e-mail do participante (**breno.xavier@ufpe.br**) foi registrado durante o envio deste formulário.

Provide an example of where the bear classification model might work poorly in production, due to structural or style differences in the training data. *

if the model was trained on images of bears in a natural setting, but was then used to classify bears in a more controlled setting such as a zoo, it may not perform as well because the images in the training data and the images in the production setting are structurally or stylistically different

Where do text models currently have a major deficiency? *

I have learned that one major deficiency of text models is their inability to fully understand and capture the context and underlying meaning of words and phrases. Text models often rely on patterns and correlations found in the training data, and are not able to fully grasp the deeper meaning and context of the words and phrases they are processing

What are possible negative societal implications of text generation models? *

one potential negative societal implication of text generation models is the possibility of their use for malicious purposes, such as creating fake news or spreading disinformation

In situations where a model might make mistakes, and those mistakes could be harmful, what is a good alternative to automating a process? *

in situations where a model might make mistakes that could be harmful, a good alternative to automating the process is to have a human review and verify the decisions made by the model. This can help to reduce the risk of mistakes being made and ensure that any potential negative consequences are minimized

What kind of tabular data is deep learning particularly good at? *

tabular data that has a large number of features or columns, this is because deep learning models are able to learn and extract useful patterns and relationships from large amounts of data, and are able to handle a large number of features without requiring a lot of feature engineering or preprocessing

What's a key downside of directly using a deep learning model for recommendation systems?

one key downside of directly using a deep learning model for recommendation systems is that the model may not be able to provide clear explanations for its recommendations

What is DataLoaders? *

dataLoaders is a PyTorch utility that is used to load and prepare data for use in a machine learning model, particularly useful for working with large datasets because it allows the data to be loaded and processed in smaller batches

What four things do we need to tell fastai to create DataLoaders? *

A training dataset and a validation dataset and a test dataset and a batch size

What does the splitter parameter to DataBlock do? *

it is used to specify how the data should be split into training, validation, and test sets

How do we ensure a random split always gives the same validation set? *

in order to ensure that a random split always gives the same validation set, we can set the seed for the random number generator used by the split function

What letters are often used to signify the independent and dependent variables? *

independent: X or x and dependent Y or y

What's the difference between the crop, pad, and squish resize approaches? When might you choose one over the others? *

crop, pad, and squish resize approaches are different ways of resizing images to a target size, crop involves selecting a portion of the image and discarding the rest and the pad involves adding padding around the image to set to the target size the squish is resize the image without preserving the aspect ratio. the crop is often used when the aspect ratio of the target size is the same as original size, the pad is used when the aspect ratio of the images are different and squish is not recomendded to be used cause its distorted the image

What is data augmentation? Why is it needed? *

data augmentation is a technique used to artificially increase the size of a dataset by generating new data points based on existing data and it is needed because machine learning models often perform better when they are trained on larger and more diverse datasets

What is the difference between item_tfms and batch_tfms? *

item_tfms refers to transformations that are applied to individual items in the dataset, while batch_tfms refers to transformations that are applied to a batch of items at once

What is a confusion matrix? *

it is a confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known

What does export save? *

is a feature that allows you to save or transfer information from one place to another.

What is it called when we use a model for getting predictions, instead of training? *

inference.

What are IPython widgets? *

they are interactive tools that allow you to control and display data in your notebooks.

When might you want to use CPU for deployment? When might GPU be better? *

use cpu if if the task requires less computational power and if real-time processing is not a priorit and use gpu if real-time processing is critical.

What are the downsides of deploying your app to a server, instead of to a client (or edge) device such as a phone or PC? *

downsides are: reduced privacy, lower performance, and increased latency.

What are three examples of problems that could occur when rolling out a bear warning system in practice? *

technical malfunctions, low user adoption and incorrect notifications.

What is "out-of-domain data"? *

it is when data is not similar to the data the model was trained.

What is "domain shift"? *

domain shift refers to a change in the distribution of the data that the model is trained on and the distribution of the data it is being applied to.

What are the three steps in the deployment process? *

the steps are: testing, packaging, and deploying.