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

This problem is called when “out of domain data” results in poor results. For example, if all the bear images trained on where hand drawn and in black and white, then real images would not be classified very well.

1. Where do text models currently have a major deficiency?

At generating correct responses. If you give a model a knowledgebase of information, it’s not very good at generating correct responses, although to a layman the responses might seem.

1. What are possible negative societal implications of text generation models?

The creation can seem like a normal response, and spread disinformation, negative sentiment etc on social media.

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

Put a human in the loop to make decisions. For medical imaging, say you have scans for cancer. Let the model make a list of severe images and let a trained human look at it. It can thus make the process so much more efficient because the obvious non cancer cases can be un-prioritised.

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

An example of a popular tabular non dl models is random forests. These models can generally not work well with a large number of columns.

Thus DL is particularly good at data where there are many columns to include from. Particularly, columns containing natural language (book titles, reviews, etc.), and high-cardinality categorical columns (i.e., something that contains a large number of discrete choices, such as zip code or product ID).

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

It can’t easily introduce new relevant recommendations like a human could. For example, at a bookstore, you can ask a human for recommendations and they can recommend similar authors. But an ML recommender would likely just recommend things similar users have bought, instead of entirely different things. A bookstore keeper can use much intuition and experience and tailor a recommendation for a person, while on amazon it does not.

1. What are the steps of the Drivetrain Approach?

The Drivetrain Approach is used in order to make models useful in practice. You can have a model that’s very accurate, but not very useful in practice, as it doesn’t do what users prefer it to.

The basic idea is to start with considering your objective, then think about what actions you can take to meet that objective (aka levers: what actions you can contro) and what data you have (or can acquire) that can help, and then build a model that you can use to determine the best actions to take to get the best results in terms of your objective.

1. How do the steps of the Drivetrain Approach map to a recommendation system?

The objective is to recommend products to users so that they purchase more and/or enjoy their products.

The actions you can take: on amazon you can recommend products in a order of most recommended.

The data you have is other users’ order history and information such as their age, gender, location etc.

1. Create an image recognition model using data you curate, and deploy it on the web.

Done

1. What is DataLoaders?

A DataLoader is a class that provides batches of items at a time to the GPU. Default is 64. DataLoaders is both validation and training DataLoader(s).

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

What kinds of data we are working with

How to get the list of items

How to label these items

How to create the validation set

1. What does the splitter parameter to DataBlock do?

Tells DataBlock how to split the data into training and validation

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

You set the seed

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

X for independent

Y for dependent

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

Crop is simply crops the image from the centre into a certain size.

Pad adds padding to get the image into a certain size.

Squish squishes the image into a certain size.

1. What is data augmentation? Why is it needed?

Data augmentation is when you modify the training images in order to get better results. For example, every epoch of training, randomcrop will randomly select a part of the image to crop down and train on. You can choose to run as many epochs as you want.

This is useful especially when you don’t have much training data to work on, as it generates new data to work with. It can give different angles and parts to work on, just like in real life where you can walk around and see different parts of the same object.

1. What is the difference between item\_tfms and batch\_tfms?

Item\_tmfs does data augmentations one by one to the images using the CPU, we usually use it first to resize the images to the same size.

Batch\_tmfs apply the data augmentations to an entire batch of them using the GPU, we usually do it afterwards as the GPU is better for this.

1. What is a confusion matrix?

A confusion matrix shows how the model preformed on the validation data in terms of how predictions and correct values for the categories.

1. What does export save?

The model.

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

When we use a model for getting predictions, instead of training, we call it inference

1. What are IPython widgets?

IPython widgets are GUI components that bring together JavaScript and Python functionality in a web browser, and can be created and used within a Jupyter notebook.

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

Overall, we'd recommend using a simple CPU-based server approach where possible, for as long as you can get away with it. If you're lucky enough to have a very successful application, then you'll be able to justify the investment in more complex deployment approaches at that time.

1. 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?

It’s hard and expensive to build and maintain your own server, and even rent one, if you’re working with many users. It can end up with long waiting times too.

Because nowadays, most newer iOS devices for example have dedicated ML hardware, it might just be better for them to use their own hardware to do the app’s processing. It would cost less as there are no server fees and realistically could be faster.

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

It’s because in practice the circumstances can easily call for situations the model has not been trained on.

For example, photos at night, at weird angles, pixelated.

Sometimes it is then necessary as a result to collect your own data to remedy this, if there is not data available online that you can find to do so.

1. What is "out-of-domain data"?

The previously described problem. Data in practice the model has not been trained on.

1. What is "domain shift"?

Another problem with the training data.

Over time the training data may become less relevant to the data encountered in practice.

1. What are the three steps in the deployment process?

The manual process: where you run the model in parallel to how normally humans solve the task, and see if the model makes sense, and use it if it’s useful but as an addition.

Limited scope development: where you roll out a small highly supervised model with humans. More ambitiously than in the manual process.

Gradual expansion: where you roll it out slowly and carefully monitor the data with regular checks.

In every stage, compare how and why the model predicts to the other stages.