## Week 1a – What It Means to be AI First

1. What would you use to replace user input by machine learning?
   1. Pre-trained models.
2. What types of data are used in machine learning models?
   1. Both labelled and unlabelled data.
3. What are best practices for data preparation?
   1. Avoid training-serving skew.
   2. Avoid target leakage.
   3. Provide a time signal.
4. What are parts of the machine learning training phase?
   1. Data management.
   2. Create the models.
   3. Evaluating the models.
5. What’s the most efficient way to transcribe speech?
   1. A speech API.

## Week 1b – How Google Does Machine Learning

1. What type of network is used to identify faces, objects and traffic signs?
   1. Convolutional Neural Networks.
2. What is true about machine learning systems?
   1. Almost every single one has a team of people reviewing the algorithm, reviewing their responses and doing random sub-samples and it generates a lot of value for the organisation, for customers and for end users.
3. What are facets that differentiate deep learning networks in multilayer networks?
   1. More complex ways of connecting layers.
   2. Cambrian explosion of computing power to train.
   3. Automatic feature extraction.
4. What’s correct?
   1. Nothing.

## Week 2a – Machine Learning Development with Vertex AI

1. In machine learning development, which phase identifies your use case?
   1. Framing the problem.
2. Typically, machine learning practitioners train models using different architectures, input datasets, hyperparameters and hardware. What architectural type would you use for cyber-security, pattern recognition, self-driving cars and reinforced learning?
   1. GANs or Generative Adversarial Networks.
3. Which Vertex AI service lets you access data, process data in a Dataproc cluster, train a model, share your results and more, all without leaving the JupyterLab interface?
   1. Workbench.
4. Moving from experimentation to production requires packaging, deploying and monitoring your model - which can give you confidence that your model is making useful predictions in production. Monitoring measures key model performance metrics and includes:
   1. Model drift, model performance, model outliers and quality.
5. The way you deploy a TensorFlow model is different from how you deploy a PyTorch model, and even TensorFlow models might differ based on whether they were created using AutoML or by means of code. In the unified set of APIs that Vertex AI provides, you can treat all these models in the same way.
   1. True.
6. Vertex AI is flexible. You choose your training method. \_\_\_\_\_\_\_\_\_\_\_\_\_ lets you create a training application optimized for your targeted outcome. You have complete control over training application functionality; you can target any objective, use any algorithm, develop your own loss functions or metrics, or do any other customization.
   1. Custom training.
7. What is a managed dataset in Vertex AI?
   1. Data loaded into Vertex AI - whether it be from Google Cloud Storage or BigQuery. This means, for example, that it can be linked to a model.

## Week 2b – Machine Learning Development with Vertex Notebooks

1. Fill in the blank: Vertex AI Workbench provides two Jupyter notebook-based options for your data science workflow. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ are Google-managed environments with integrations and features that help you set up and work in an end-to-end notebook-based production environment.
   1. Managed notebook instances.
2. Fill in the blank: Vertex AI Workbench provides two Jupyter notebook-based options for your data science workflow. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ are Deep Learning VM Images instances that are heavily customizable and are therefore ideal for users who need a lot of control over their environment.
   1. User-Managed notebook instances
3. Which statement is correct regarding Vertex AI Workbench Notebooks?
   1. Both options are pre-packaged with JupyterLab and have a pre-installed suite of deep learning packages, including support for the TensorFlow and PyTorch frameworks.
   2. Both options support GPU accelerators and the ability to sync with a GitHub repository.
   3. Both options are protected by Google Cloud authentication and authorization.
4. In a Vertex AI Workbench Jupyter Notebook, you can access your data without leaving the JupyterLab interface.
   1. True.
5. Where can you find the Cloud Storage and Bigquery extension to browse data?
   1. Left side-bar
6. For users who have specific networking and security needs, \_\_\_\_\_\_ can be the best option. You can use VPC Service Controls to set up a \_\_\_\_\_\_ within a service perimeter and implement other built-in networking and security features. You can also configure user-managed notebooks instances manually to satisfy some specific networking and security needs.
   1. User-Managed notebook instances

## Week 2c – Best Practices for Implementing Machine Learning on Vertex AI

1. The data used to train a model can originate from any number of systems, for example, logs from an online service system, images from a local device, or documents scraped from the web. Which of the following is a Best Practice for Preparing and Storing unstructured data such as images, audio, and video?
   1. In Cloud storage
2. Your dataset is considered small, less than 5,000 rows and around 10MB. You are not using AutoML but a Jupyter Notebook instance. Which of the following is a Best Practice for Training a model with a small dataset?
   1. For small datasets, train the model within the notebook instance.
3. Which of the following statement is correct for Explainable AI?
   1. It offers feature attributions to provide insights into why models generate predictions.
4. Use BigQuery to process tabular data and use Dataflow to process unstructured data.
   1. True.

## Week 2d – Responsible AI Development

1. Human biases lead to bias in machine learning models. Unconscious biases exist in our data and exist in two forms. What are the two forms of unconscious biases in data?
   1. There are the human biases that exist in data because data found in “the world” has existing biases with regard to properties like gender, race, and sexual orientation. For example, there may be reporting bias by our subjects because they only choose to reveal certain aspects about themselves or their opinions. We can also run into human biases which arise as part of our data collection and labelling procedures.
2. The impact of biases in collecting data and labeling data affects the entire machine learning pipeline. The biases in the original data are going to be reflected downstream in our models and consequently are going to result in potentially biased outcomes. You need to create a checklist for situations where you should watch out for bias-related issues. What questions should this checklist include?
   1. Does your use case or product specifically use any of the following data: biometrics, race, skin color, religion, sexual orientation, socioeconomic status, income, country, location, health, language, or dialect?
   2. Does your use case or product use data that is likely to be highly correlated with any personal characteristics (for example, zip code or other geospatial data is often correlated with socioeconomic status and/or income; image/video data can reveal information about race, gender, and age)?
   3. Could your use case or product negatively affect individuals’ economic or other important life opportunities?
3. One of the key tools to help in understanding inclusion and how to introduce inclusion across different kinds of groups across your data is by understanding the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
   1. Confusion matrix
4. Which of the following is an example of a “false negative”?
   1. When the label says something exists and the model doesn’t predict it—that’s a false negative. So, in the face detection example in this lesson, the model says that there is no face in the image—when the image’s label says there \*is\* a face.
5. Datasets can contain hundreds of millions of data points, each consisting of hundreds (or even thousands) of features, making it nearly impossible to understand an entire dataset in an intuitive fashion. The key here is to utilize visualizations that help unlock nuances and insights in large datasets. Which tool would be most appropriate?
   1. Facets.
6. Which approach is followed to achieve a better performance across subgroups?
   1. Equality of opportunity
7. The confusion matrix helps which of the following?
   1. Understanding inclusion and how to introduce inclusion across different subgroups within your data.
   2. Evaluating performance in machine learning
8. What is it called when the label says something doesn't exist, but the model says it exists?
   1. False positive.