**Question 1: Cloud Computing for Deep Learning (20 points)**

Cloud computing offers significant advantages for deep learning applications.

(a) Define **elasticity** and **scalability** in the context of cloud computing for deep learning. (10 points)  
(b) Compare **AWS SageMaker**, **Google Vertex AI**, and **Microsoft Azure Machine Learning Studio** regarding their deep learning capabilities. (10 points)

**Expected Output**

Write the definition and comparison for (a) and (b). No code is needed.

**Ans:**

1. **Elasticity and Scalability in Cloud Computing for Deep Learning**

**Elasticity:** Elasticity refers to the ability of a cloud system to dynamically allocate and deallocate resources based on real-time demand. In deep learning, this means provisioning GPUs/TPUs during high computational loads and releasing them when they are no longer needed, optimizing costs and performance.

**Scalability:** Scalability is the capacity of a cloud infrastructure to handle an increasing amount of workload by adding resources (scaling up) or distributing tasks across multiple machines (scaling out). For deep learning, this ensures that training large models or handling multiple concurrent training jobs remains efficient.

1. **Comparison of AWS SageMaker, Google Vertex AI, and Microsoft Azure Machine Learning Studio**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **AWS SageMaker** | **Google Vertex AI** | **Azure Machine Learning Studio** |
| **Hardware Support** | Supports CPUs, GPUs, and AWS Inferential for deep learning acceleration | Supports GPUs and TPUs for large-scale training | Supports CPUs, GPUs, and FPGAs for deep learning workloads |
| **AutoML Capabilities** | Built-in AutoML for  hyperparameter  tuning and model  optimization | Supports GPUs and  TPUs for large-scale  training | Supports CPUs, GPUs, and FPGAs  for deep learning workloads |
| **Prebuilt Models & Services** | Offers pre-trained  models and custom  model deployment | Provides AI APIs and  model training with  deep learning  frameworks | Integrates with OpenAI models and  Azure Cognitive Services |
| **Prebuilt Models & Services** | Seamless  integration with  AWS cloud services  (S3, Lambda, EC2) | Integrates with Google  Cloud services like  BigQuery and Dataflow | Works with Azure cloud ecosystem  and enterprise applications |
| **Pricing & Cost Efficiency** | Pay-as-you-go  model with spot  instance support for  cost savings | Flexible pricing with  TPUs optimized for AI  workloads | Competitive pricing with reserved  instances for enterprise users |