## Question 1

**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** in terms of their deep learning capabilities. (10 points)

**Expected Output**

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

## Answer

**Cloud Computing for Deep Learning**

**(a) Elasticity and Scalability in Cloud Computing for Deep Learning**

1. **Elasticity:** Elasticity refers to the ability of a cloud system to dynamically allocate or deallocate resources based on demand. In deep learning, this means that compute resources such as GPUs or TPUs can be automatically scaled up during model training and scaled down when they are no longer needed, optimizing cost efficiency.
2. **Scalability:** Scalability is the capability of a system to handle increased workloads by adding resources either vertically (upgrading existing instances) or horizontally (adding more instances). In deep learning, scalability ensures that as data and computational demands grow, the system can efficiently scale to accommodate larger training datasets and more complex models.

**(b) Comparison of AWS SageMaker, Google Vertex AI, and Microsoft Azure Machine Learning Studio**

| **Feature** |  | **AWS SageMaker** | **Google Vertex AI** | **Microsoft Azure ML Studio** |
| --- | --- | --- | --- | --- |
| **Ease of Use** |  | Provides built-in Jupyter notebooks and automated ML features | Fully managed with AutoML and integrated notebooks | User-friendly UI with drag-and-drop ML tools |
| **Compute Power** |  | Supports EC2 instances with GPUs/TPUs | Uses Google Cloud TPUs and GPUs for scalable training | Offers GPU and FPGA-based training with Azure VMs |
| **Deep Learning Frameworks** |  | Supports TensorFlow, PyTorch, MXNet, Scikit-learn | Supports TensorFlow, PyTorch, XGBoost, and AutoML | Supports TensorFlow, PyTorch, ONNX, and Scikit-learn |
| **Automation & MLOps** |  | Provides automated model tuning, deployment, and CI/CD integration | Features AutoML and continuous model monitoring | Includes ML pipelines, AutoML, and MLOps integration |
| **Integration & Ecosystem** |  | Deeply integrated with AWS services like S3, Lambda, and DynamoDB | Works seamlessly with Google Cloud services like BigQuery and Dataflow | Integrates well with Azure services like Power BI and Synapse Analytics |
| **Pricing Model** |  | Pay-as-you-go with on-demand, reserved, or spot instances | Pay-as-you-go with cost-efficient TPUs | Pay-as-you-go with per-minute billing for compute instances |