

Student 24 – Answer Sheet

Question 1:

Explain the differences between supervised, unsupervised, and reinforcement learning in machine learning. Provide examples of applications for each approach.

Answer:

Supervised learning is a machine learning approach where algorithms are trained using datasets that contain both input features and their corresponding correct outputs. The model learns to map inputs to outputs, making it effective for tasks where historical data with known results is available. Typical applications include classifying emails as spam or not, diagnosing diseases from medical images, and predicting house prices. Supervised learning excels in both classification (predicting categories) and regression (predicting continuous values).

Unsupervised learning, in contrast, operates on data without explicit output labels. The algorithm must independently identify patterns, clusters, or structures within the data. Common techniques include clustering (such as grouping customers by purchasing behavior) and dimensionality reduction (simplifying complex datasets for visualization). This approach is valuable for customer segmentation in marketing, anomaly detection in cybersecurity, and discovering hidden topics in large text datasets.

Reinforcement learning is a paradigm where an agent interacts with an environment and learns to make decisions by receiving feedback in the form of rewards or penalties. The agent aims to maximize cumulative rewards over time by exploring and exploiting different strategies. Reinforcement learning is especially suited for sequential decision-making problems, such as training AI to play games (like AlphaGo), autonomous vehicle navigation, robotic manipulation, and resource management in logistics.

Question 2:

Describe the architecture and functioning of Convolutional Neural Networks (CNNs) and explain why they are particularly effective for image recognition tasks.

Answer:

Convolutional Neural Networks (CNNs) are deep learning models designed specifically for processing grid-like data, such as images. The architecture of a CNN typically consists of several layers:

- **Convolutional layers:** These apply learnable filters across the input image to extract local features like edges, textures, and shapes.
- **Activation functions:** Functions like ReLU introduce non-linearity, enabling the network to model complex relationships.
- **Pooling layers:** These downsample feature maps, reducing their size while retaining important information, which improves computational efficiency and provides translation invariance.
- **Fully connected layers:** At the end, these layers perform the final classification or regression based on the features extracted by earlier layers.

CNNs are highly effective for image recognition because their local connectivity and parameter sharing reduce the number of learnable parameters, making them efficient and less prone to overfitting. The hierarchical structure of CNNs allows them to learn increasingly complex features at each layer, mirroring the way human vision works. This enables CNNs to recognize objects regardless of their position, scale, or orientation in the image, making them the leading architecture for tasks like image classification, object detection, and facial recognition.

Question 3:

Discuss the ethical considerations and potential societal impacts of implementing artificial intelligence systems in critical decision-making processes.

Answer:

The use of AI in critical decision-making raises several ethical and societal challenges. **Algorithmic bias** is a major concern, as AI systems trained on historical data can inherit and even amplify existing societal prejudices, leading to unfair outcomes in areas such as hiring, lending, and criminal justice. For example, facial recognition systems have shown higher error rates for certain demographic groups, potentially resulting in discrimination.

Transparency and explainability are also crucial issues. Many advanced AI models, especially deep neural networks, operate as "black boxes," making it difficult to understand or explain their decisions. This lack of interpretability can undermine trust, particularly in high-stakes fields like healthcare, where understanding the reasoning behind a recommendation is essential for safety and accountability.

Privacy is another significant concern, as AI systems often require access to large amounts of personal data, raising questions about consent, data security, and potential misuse. Additionally,

the deployment of AI can disrupt labor markets, potentially leading to job displacement and increased economic inequality if the benefits of automation are not widely shared.

To address these challenges, it is essential to combine technical solutions (such as fairness-aware algorithms and explainable AI), regulatory frameworks, and ongoing stakeholder engagement to ensure AI systems are ethical, transparent, and aligned with societal values.

Question 4:

Explain the concept of transfer learning in deep neural networks and discuss its advantages and limitations.

Answer:

Transfer learning is a technique where a neural network trained on one task is adapted for a different, but related, task. In deep learning, this usually involves taking a model pre-trained on a large dataset (such as ImageNet for images) and fine-tuning it for a new task with limited data. Typically, the final layers of the pre-trained model are replaced with new layers suited to the new task, and the network is retrained—sometimes only the new layers, sometimes the entire network with a lower learning rate.

The main advantage of transfer learning is that it leverages knowledge from large, diverse datasets, making it possible to train effective models even when the target dataset is small. This reduces the need for extensive labeled data, speeds up training, and often leads to better performance by providing a strong starting point for learning. Transfer learning also helps prevent overfitting by transferring general feature representations learned from the source task.

However, transfer learning works best when the source and target tasks are similar; if the domains are too different, the transferred knowledge may not be useful or could even hinder performance (a phenomenon known as negative transfer). Additionally, pre-trained models may carry over biases from their original training data, and their architectures may not always be optimal for the new task.

Question 5:

Describe the principles of natural language processing (NLP) and how transformer-based models like BERT have revolutionized language understanding tasks.

Answer:

Natural Language Processing (NLP) is a field focused on enabling computers to analyze, understand, and generate human language. Traditional NLP relied on rule-based systems and statistical models, which often struggled to capture the complexity and context of language.

Transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers), have dramatically advanced NLP. Transformers use self-attention mechanisms, allowing the model to consider the relationships between all words in a sentence simultaneously, rather than processing them sequentially. BERT introduced the concept of bidirectional context, enabling the model to understand the meaning of a word based on both its left and right surroundings.

These models are pre-trained on massive text corpora using tasks like masked language modeling and next sentence prediction, then fine-tuned for specific applications such as sentiment analysis, question answering, and text classification. This approach allows for the development of rich, contextual word representations that capture semantic and syntactic nuances. As a result, transformer-based models have set new performance benchmarks across a wide range of NLP tasks, enabling more accurate and natural human-computer interactions. Despite challenges such as high computational requirements and potential biases, transformers have become the foundation of modern NLP systems.