Intro to Neural Networks + Overview of AI Architectures

AI Architecture for Different Data Modalities

Lesson Objectives

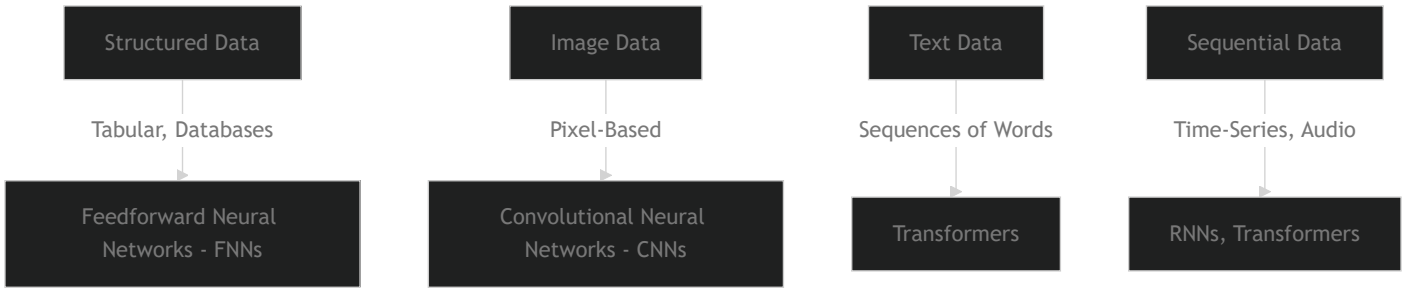
By the end of this lesson, students will be able to:

- Identify different AI architectures suited for structured, image, text, and sequential data.
- Understand the challenges associated with different data modalities.
- Select the most appropriate AI model based on data type and real-world constraints.
- Explore how multimodal AI systems integrate multiple architectures for complex tasks.

Understanding Data Modalities in AI

AI models are designed to process different **types of data**, referred to as **data modalities**. The effectiveness of an AI system depends on selecting the right architecture for the specific data type.

Common Data Modalities and Their Challenges



Key Challenges by Data Type:

- **Structured Data:** Requires feature engineering; deep learning is not always necessary.
- **Image Data:** High-dimensional, requires spatial feature extraction.
- **Text Data:** Requires context understanding, traditional models struggle with long dependencies.
- **Sequential Data:** Capturing temporal dependencies is challenging, especially for long sequences.

Choosing the Right AI Model

Each AI architecture is optimized for specific data types:

Data Type	Best AI Architecture	Example Use Cases
Structured Data	Feedforward Neural Networks (FNNs)	Credit scoring, fraud detection
Image Data	Convolutional Neural Networks (CNNs)	Object recognition, medical imaging
Text Data	Transformers (e.g., BERT, GPT)	Sentiment analysis, chatbots
Sequential Data	RNNs, Transformers	Speech recognition, stock price forecasting

> **Key Takeaway:** The success of an AI model depends not only on the model itself but also on how well it fits the data modality.

Exploring Multimodal AI Systems

What is Multimodal AI?

Multimodal AI systems combine multiple types of input data, such as text, images, and audio, to improve predictions and decision-making. These systems leverage different AI architectures for each modality and fuse their outputs.

Example Applications:

- **Self-driving cars:** Use cameras (CNNs), lidar data (FNNs), and textual navigation inputs (Transformers).
- **Medical diagnosis:** Combines patient history (FNNs) with medical imaging (CNNs) and doctor notes (Transformers).
- **E-commerce recommendations:** Uses purchase history (FNNs), product images (CNNs), and customer reviews (Transformers).

Coding Walkthrough: Implementing a Multimodal AI System

Let's implement a **basic multimodal model** that:

- Uses a **CNN** to process image data.
- Uses a **Transformer-based model** to process text.
- Combines both representations to make a final classification decision.

Python Code:

Copy

```
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from transformers import TFAutoModel

# Define image processing branch (CNN)
image_input = keras.Input(shape=(64, 64, 3))
x = layers.Conv2D(32, (3,3), activation='relu')(image_input)
x = layers.MaxPooling2D((2,2))(x)
```

```
x = layers.Flatten()(x)
image_features = layers.Dense(128, activation='relu')(x)

# Define text processing branch (Transformer)
text_input = keras.Input(shape=(128,))  # Assume we have tokenized text
text_model = TFAutoModel.from_pretrained("distilbert-base-uncased")
text_features = text_model(text_input)[0][:, 0, :]
text_features = layers.Dense(128, activation='relu')(text_features)

# Combine both modalities
combined = layers.concatenate([image_features, text_features])
output = layers.Dense(1, activation='sigmoid')(combined)

# Create and compile model
multimodal_model = keras.Model(inputs=[image_input, text_input], outputs=output)
multimodal_model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['

# Summary of the multimodal model
multimodal_model.summary()
```

Discussion Questions:

- How does combining multiple data types improve model performance?
- What are the biggest challenges in training multimodal AI systems?
- Can you think of other applications where multimodal AI would be useful?

Summary & Key Takeaways

- Different AI architectures are **optimized for different types of data**.
- **Multimodal AI integrates multiple architectures** to process different types of input simultaneously.
- **Combining text, images, and structured data** enables more powerful AI applications.
- **Training multimodal models is more complex** but allows AI systems to make better-informed decisions.