AI Tools and Applications Assignment Report

Part 1: Theoretical Answers

Part 1: Theoretical Understanding

Q1: What is the primary purpose of Scikit-learn in the Python ecosystem?
Scikit-learn is a powerful and widely-used open-source machine learning library for Python.
Its primary purposes include:
- Providing simple and efficient tools for data mining and data analysis
- Offering a consistent interface for various machine learning algorithms
- Supporting both supervised and unsupervised learning
- Including tools for model fitting, data preprocessing, model selection, and evaluation
- Being built on NumPy, SciPy, and matplotlib

Q2: Explain the difference between TensorFlow and PyTorch.

TensorFlow:
- Developed by Google Brain
- Uses static computation graphs by default (though eager execution is now available)
- Better for production deployment
- Strong support for distributed computing
- More comprehensive ecosystem (TFX, TensorBoard, etc.)

PyTorch:
- Developed by Facebook's AI Research lab
- Uses dynamic computation graphs by default
- More pythonic and intuitive for research
- Better for rapid prototyping
- Preferred by researchers for its flexibility
- Easier debugging due to its dynamic nature

Q3: What is spaCy primarily used for in NLP tasks?
spaCy is an open-source software library for advanced natural language processing.
Its primary uses include:
- Tokenization and text processing
- Part-of-speech tagging
- Named Entity Recognition (NER)
- Dependency parsing
- Word vectors and semantic similarity
- Text classification
- Rule-based matching
- Custom pipeline components
- Multi-language support
- Efficient processing of large volumes of text

AI Tools and Applications Assignment Report

Part 2: Model Outputs

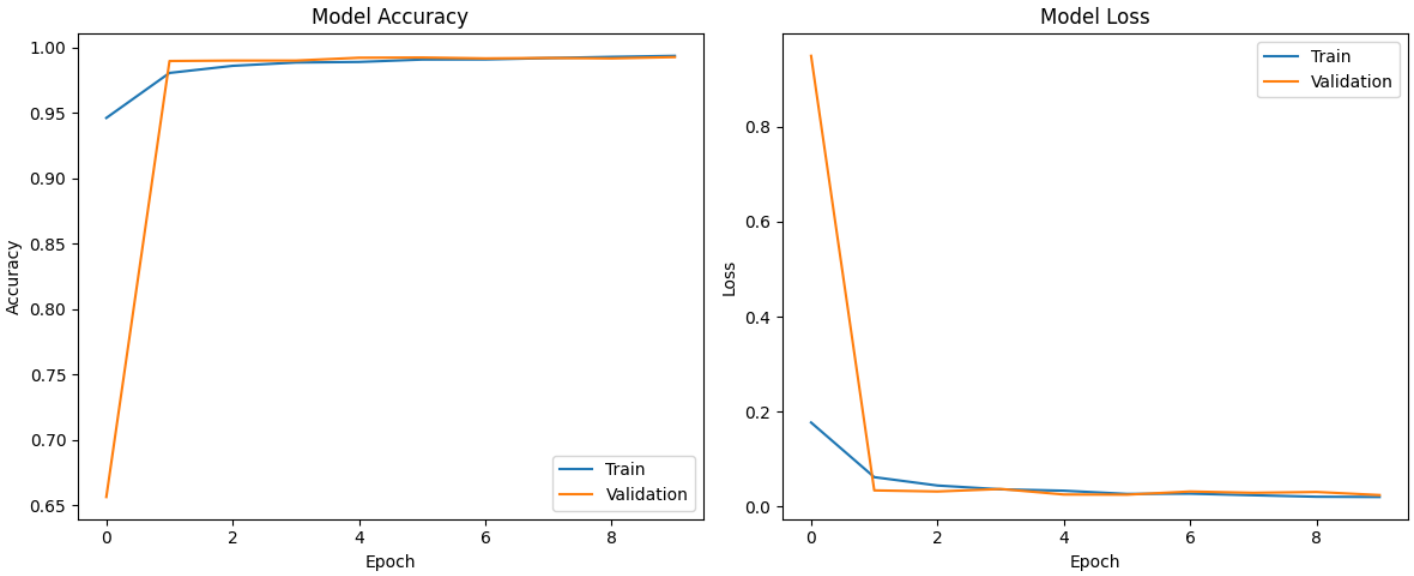
MNIST Model Architecture

```
Model: "sequential"
+---------------------------------------------------------------------------+
| Layer (type)                    | Output Shape           |       Param # |
|---------------------------------+------------------------+---------------|
| conv2d (Conv2D)                 | (None, 28, 28, 32)     |           320 |
|---------------------------------+------------------------+---------------|
| batch_normalization             | (None, 28, 28, 32)     |           128 |
| (BatchNormalization)            |                        |               |
|---------------------------------+------------------------+---------------|
| conv2d_1 (Conv2D)               | (None, 28, 28, 32)     |         9,248 |
|---------------------------------+------------------------+---------------|
| batch_normalization_1           | (None, 28, 28, 32)     |           128 |
| (BatchNormalization)            |                        |               |
|---------------------------------+------------------------+---------------|
| max_pooling2d (MaxPooling2D)    | (None, 14, 14, 32)     |             0 |
|---------------------------------+------------------------+---------------|
| dropout (Dropout)               | (None, 14, 14, 32)     |             0 |
|---------------------------------+------------------------+---------------|
| conv2d_2 (Conv2D)               | (None, 14, 14, 64)     |        18,496 |
|---------------------------------+------------------------+---------------|
| batch_normalization_2           | (None, 14, 14, 64)     |           256 |
| (BatchNormalization)            |                        |               |
|---------------------------------+------------------------+---------------|
| conv2d_3 (Conv2D)               | (None, 14, 14, 64)     |        36,928 |
|---------------------------------+------------------------+---------------|
| batch_normalization_3           | (None, 14, 14, 64)     |           256 |
| (BatchNormalization)            |                        |               |
|---------------------------------+------------------------+---------------|
| max_pooling2d_1 (MaxPooling2D)  | (None, 7, 7, 64)       |             0 |
|---------------------------------+------------------------+---------------|
| dropout_1 (Dropout)             | (None, 7, 7, 64)       |             0 |
|---------------------------------+------------------------+---------------|
| flatten (Flatten)               | (None, 3136)           |             0 |
|---------------------------------+------------------------+---------------|
| dense (Dense)                   | (None, 256)            |       803,072 |
|---------------------------------+------------------------+---------------|
| batch_normalization_4           | (None, 256)            |         1,024 |
| (BatchNormalization)            |                        |               |
|---------------------------------+------------------------+---------------|
| dropout_2 (Dropout)             | (None, 256)            |             0 |
|---------------------------------+------------------------+---------------|
| dense_1 (Dense)                 | (None, 10)             |         2,570 |
+---------------------------------------------------------------------------+
 Total params: 872,428 (3.33 MB)
 Trainable params: 871,530 (3.32 MB)
```

AI Tools and Applications Assignment Report

```
Non-trainable params: 896 (3.50 KB)
Optimizer params: 2 (12.00 B)
```

MNIST Model Training History



Final Training Accuracy: 0.9938

Final Validation Accuracy: 0.9928

Final Training Loss: 0.0204

Final Validation Loss: 0.0245

NLP Analysis: Named Entity Recognition

Text: I bought this product from Amazon and it was delivered by DHL in 2 days.
Entities:
- Amazon: ORG
- DHL: ORG
- 2 days: DATE

Text: The iPhone 13 has a great camera but the battery life could be better.
Entities:

Text: Microsoft Office 365 subscription costs $99 per year.
Entities:

AI Tools and Applications Assignment Report

- Microsoft Office: ORG
- 99: MONEY

AI Tools and Applications Assignment Report

Part 3: Ethical Reflection

Ethical Considerations in AI Model Development

1. Data Bias and Fairness:
- The MNIST dataset primarily contains Latin/Arabic numerals, which may not represent all writing styles
- Potential underrepresentation of certain cultural or regional variations in handwriting
- Need for diverse training data to ensure model generalizability

2. Model Transparency:
- The decision-making process of deep learning models can be opaque
- Importance of model interpretability for critical applications
- Need for clear documentation of model limitations

3. Privacy Concerns:
- When dealing with user data, ensuring proper data anonymization
- Compliance with data protection regulations (GDPR, CCPA, etc.)
- Secure storage and handling of sensitive information

4. Mitigation Strategies:
- Regular bias audits using tools like TensorFlow Fairness Indicators
- Data augmentation to improve model robustness
- Clear communication of model limitations to end-users
- Continuous monitoring of model performance in production
- Implementation of feedback loops for model improvement

5. Responsible AI Practices:
- Regular model retraining with updated data
- Documentation of model development process
- Consideration of potential misuse cases
- Implementation of appropriate safeguards

AI Tools and Applications Assignment Report

Report Details

Report generated on: 2025-10-19 17:30:26