Handwritten Digit Recognition Using Machine Learning and AI

Ayushi, Erum, Sahana

Cohort 5

June 2, 2024

Outline

- Introduction
- Applications
- Datasets
- Recognition Models
  - K-nearest neighbours (KNN)
  - Multi-Layer Perceptron (MLP)
  - Support Vector Machines(SVM)
  - Convolutional Neural Network (CNN): LeNet-5
  - Capsule Networks
  - Random Forest Classifier
- Key Takeaways
- Future Directions

Introduction

Handwritten Digit Recognition System involves reception and interpretation of handwritten digits by a machine.

Applications

- Postal services sorting
- Bank Check Processing
- Human-Computer Interaction
- Digitizing Historical Documents
- Medical Records Management
- Form Processing

Datasets

- MNIST (Modified National Institute of Standards and Technology) Dataset:
  - It consists of 70,000 grayscale images of handwritten digits (0-9).
  - Each image is 28x28 pixels, and the dataset is well-balanced across the 10-digit classes.
- USPS (United States Postal Service) Dataset:
  - Contains 9,298 grayscale images of handwritten digits from postal mail. The images are 16x16 pixels.
  - The smaller image size can make recognition harder.
- SVHN (Street View House Numbers) Dataset:
  - Real-world image dataset for digit recognition obtained from house numbers in Google Street View images. It includes over 600,000 labeled digits.
  - The dataset is in color (RGB) and has a higher resolution.

k-nearest neighbours (KNN) algorithm

- Dataset: MNIST
- Accuracy: 95.00%
- Precision: 95.00%
- Recall: 95.00% (weighted average)
- F1_score: 95.00% (weighted average)
- Training Time: Minimal, as kNN is a lazy learning algorithm
- Hyperparameters:
  - k (Number of Neighbors) - 3
- Special features :
  - Computationally expensive during the prediction phase, especially for large datasets.
  - Sensitive to the scale of the data and irrelevant features, often requiring normalisation and feature selection.

Multi-Layer Perceptron (MLP)

- Dataset: MNIST
- Accuracy: 97.00%
- Precision: [98, 100, 98, 98, 98, 97, 96, 98, 96, 95]
- Recall: [99, 99, 96, 96. 95, 97, 98, 98, 96, 98]
- Log Loss: 0.083
- F1_score: 97.00% (Micro average and Weighted average)
- Training Time: slower due to iterative optimisation process
- Hyperparameters:
  - Hidden Layer Sizes: First hidden layer- 128 neurons Second hidden layer- 64 neurons
  - Training Epochs: 5
  - Solver: Adam
  - Loss function: Sparse categorical cross-entropy
- Special features :
  - Neurons in MLP use activation functions to introduce non-linearity into the network, allowing it to learn complex patterns in the data.
  - They require large amounts of training data to generalise well, and they are prone to overfitting

Support Vector Machine(SVM)

- Dataset: MNIST
- Accuracy: 94.38%
- Precision: 94.46%
- Recall: 94.38%
- F1_score: 94.39%
- Training Time: Faster than CNN's
- Hyperparameters:
  - C (Regularization parameter) - 10
  - Gamma (Kernel coefficient) - 0.001
- Special features :
  - handling non-linear classification tasks without explicitly transforming the input data
  - generally robust to noise in the data and outliers

Convolutional Neural Network (CNN): LeNet-5

- Dataset: MNIST
- Accuracy: 98.52%
- Test Loss: 0.054
- Precision: 99% (micro and weighted average)
- Recall: 99% (micro and weighted average)
- F1_score: 99% (micro and weighted average)
- Training Time: relatively efficient to train compared to larger neural network
- Hyperparameters:
  - Convolutional Layer Parameters: C1 = 6, C2 = 16
  - Pooling Layer Parameters: Pooling Size - 2x2, Stride - 2
  - Activation Functions: 'softmax', 'tanh'
  - Fully Connected Layer Parameters: 120, 84, 10 neurons
- Special features :
  - laid the foundation for subsequent advancements in deep learning for computer vision.

Spatial Invariance in CNNs: Advantages and Disadvantages

Advantages:

- Robustness to Translation: Allows CNNs to recognize objects regardless of position.
- Feature Sharing: Enables CNNs to learn invariant features across spatial locations.

Disadvantages:

- Ambiguity in Localization: Leads to uncertainty in object localization, especially with overlap.
- Difficulty in Separating Objects: Challenge in distinguishing overlapped objects, affecting segmentation.
- Loss of Context: Results in disregarding global context, impacting object recognition.

Capsule Networks

- **Introduction**
  - Capsule Networks address limitations of CNNs in capturing spatial hierarchies.
  - They introduce agreement among kernels for improved object recognition.
- **Key Components**
  - **Primary Capsules**: Group convolutions working together.
  - **Digit Capsules**: Represent classes, capturing detailed class information.
  - **Dynamic Routing**: Mechanism for agreement calculation between capsules.
  - **Loss Function**: Marginal loss for encouraging meaningful representations.
  - **Regularization**: Parallel decoder network to prevent overfitting.

Citation for CapsNet

- Yuxing Tan, Yao Hongge, "Deep Capsule Network Handwritten Digit Recognition", *International Journal of Advanced Network Monitoring and Controls*, vol. 5, no. 4, pp. 1-8, January 2021.
- "Handwritten Indic Character Recognition using Capsule Networks" - Bodhisatwa Mandal,Suvam Dubey, Swarnendu Ghosh, Ritesh Sarkhel, Nibaran Das, Dept. of CSE, Jadavpur University, Kolkata, 700032, WB, India.

Random Forest Classifier

- Dataset: MNIST dataset
- Accuracy: 74.15%
- Precision: 76.95%
- Recall: 73.3%
- F1_score: 72.37%
- Hyperparameters:
  - **n_estimators**
  - **max_depth**
  - **min_samples_leaf**
  - **bootstrap**: True
  - **n_jobs**
  - **random_state**

Key Takeaways

- Familiarised ourselves with machine learning
- Traditional ML models have lesser training time than complex networks
- Complex ML models have higher accuracy than traditional models