MNIST Dataset and Machine Learning Process

1. MNIST Dataset Kya Hai?

   - MNIST (Modified National Institute of Standards and Technology) dataset ek handwritten digit dataset hai.

   - Isme 70,000 images hoti hain:

     - 60,000 images for training.

     - 10,000 images for testing.

   - Har image mein ek digit (0 se 9) likha hota hai aur 28x28 pixels ka grayscale image hota hai.

2. Training aur Testing Data Ka Role

   - Training Data aur Testing Data ke beech fark samajhna zaroori hai.

   Training Data (60,000 images):

   - Purpose: Model ko seekhne ke liye use hota hai.

   - Model ko images aur unke labels diye jaate hain (Example: Ek image hai jo 3 ka digit dikhata hai, toh label bhi 3 hoga).

   - Model Learning: Model in images ko dekh kar patterns seekhta hai ki kaunsa digit kis image mein hai.

   Testing Data (10,000 images):

   - Purpose: Model ki performance evaluate karne ke liye.

   - Testing images woh naye images hote hain, jo model ne training ke dauraan kabhi nahi dekhe.

   - Model Evaluation: Testing set ka use model ki accuracy ko test karne ke liye hota hai.

3. Training aur Testing Data Ka Flow

  1. **Training Phase**:

   - Model ko 60,000 images dikhayi jaati hain aur unke corresponding labels diye jaate hain.

   - Model in data se seekhta hai ki har image ka correct digit kya hai.

  2. **Testing Phase**:

   - Jab model training complete kar leta hai, tab testing data ko use kar ke evaluate kiya jaata hai.

   - Testing images naye hote hain, jo model ne pehle kabhi nahi dekhe hote.

   - Testing data ke images par model apni predictions karta hai aur hum accuracy check karte hain.

4. Training aur Testing ke Beech Ka Fark

  - Training Data mein model ko labels diye jaate hain, taaki wo seekhe.

  - Testing Data mein model ko labels nahi diye jaate, taaki hum uski performance ko evaluate kar sakein.

5. Model Evaluation Example

  - Training data mein, model ko "3" ke 1000 images diye gaye hain.

  - Testing data mein koi naya "3" image diya gaya hai, jise model ne pehle kabhi nahi dekha.

  - Testing phase mein hum check karte hain ki model naye "3" image ko sahi predict kar raha hai ya nahi.

Key Points:

- Training Data: Model ko seekhne ke liye (images aur labels diye jaate hain).

- Testing Data: Model ki performance ko test karne ke liye (naye unseen images).

- MNIST Dataset: 60,000 training images aur 10,000 testing images ka collection hai.