CONVOLUTIONAL NEURAL NETWORKS

CONVOLUTIONAL NEURAL NETWORK

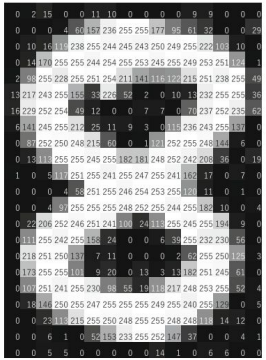
- Convolutional Neural Network (CNN) is a type of **feed-forward** artificial neural network.
- CNN performs a **linear** operation called "*convolution*" between input data and feature extractor **filter** (also known as **kernel**).
- A CNN is minimally composed of four types of layers:
  1. A convolution (CONV) layer
  2. An activation layer (usually ReLU)
  3. A pooling layer
  4. A fully connected layer.

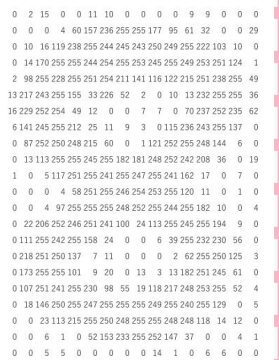
Most successful and widely used application of a CNN is analyzing visual images and videos.
(e.g, object classification, object detection and localization)

Common applications are image recognition, handwriting analysis, analyzing evolution, predicting earthquakes and natural disasters, human genome mapping projects, self-diagnoses of medical problems, automation.

When we see an image

When computer sees an image, it finds number of pixel intensity





Real image of digit 8

Represented in the form of pixel intensity

Represented in the form of array

CNN OPERATION

SureStart

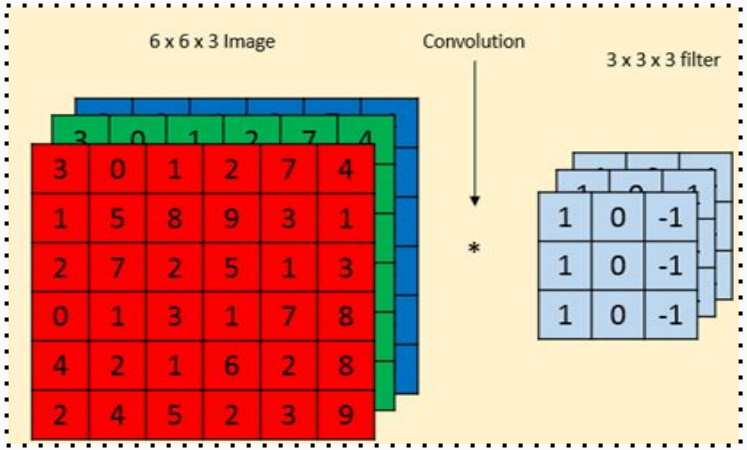
CONV

- An element-wise matrix multiplication (dot product) occurs between the pixel value with filter size.

- The kernel moves left to the right calculating the dot product, then summed resulting in a single value.

- All the single value of dot product create a new feature map.

Images are matrix. RGB image has 3 channels (Red, Blue, Green)



The kernel is square-shaped learnable filter. Usually smaller than the input data (e.g, 3x3, 5x5, 7x7, 9x9). Filter is designed to detect a specific type of feature in the input.

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

4		

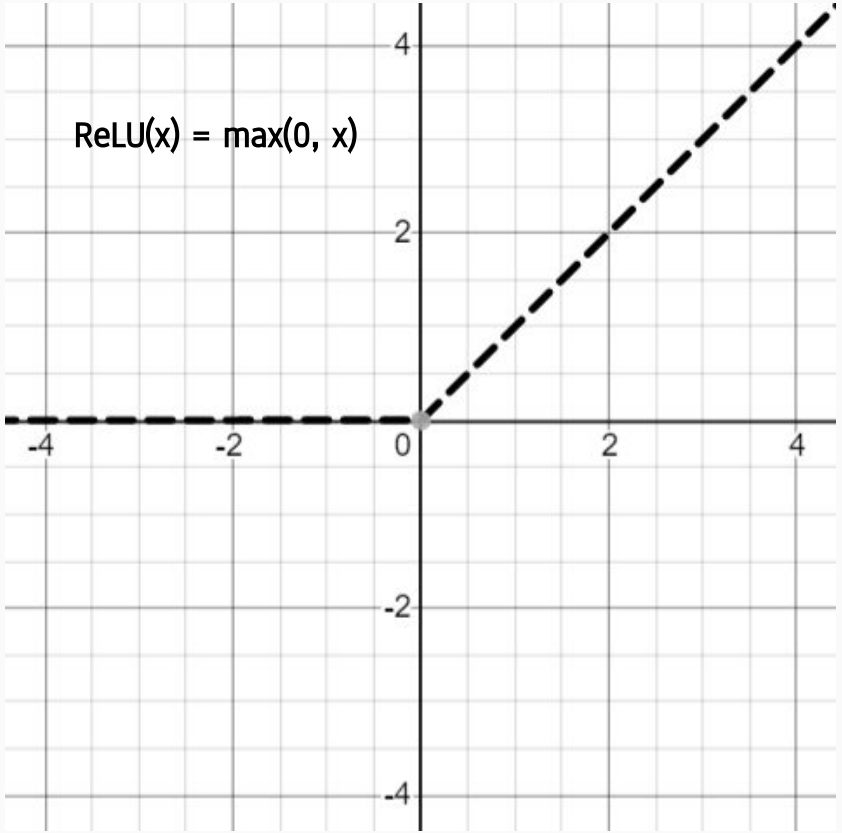
element-wise multiplication

SureStart

ACTIVATION LAYER

Once the feature maps are extracted, the next step is to move them to an activation function: **Rectified Linear Unit (ReLU) layer**.

- ReLU is a nonlinear function.
- No parameters or weights are given to learn from images
- It sets all the negative pixel value to 0 and introduces non-linearity to the network.
- The resulted output now is a rectified feature map.
- Activation functions are used in neural network to learn more complex features.



SureStart

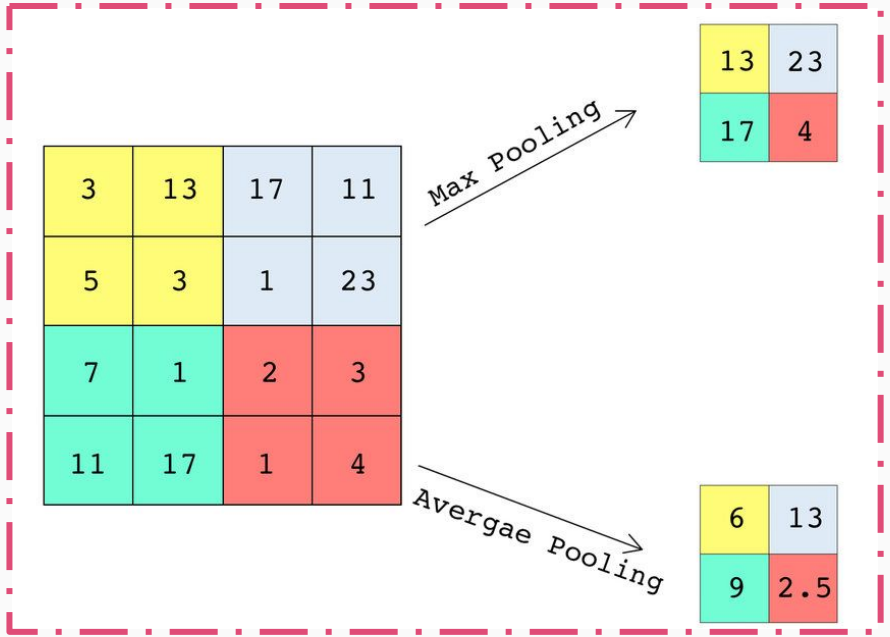
POOLING OPERATION

- The rectified feature map now goes through a pooling layer.

- Pooling layer is a down-sampling (shrinking the image) operation that reduces the dimensionality of feature map.

- The decreased size of the convolved feature map reduces the computational complexity.

- It helps to identify relevant information of the image such as edges, corner and curves.

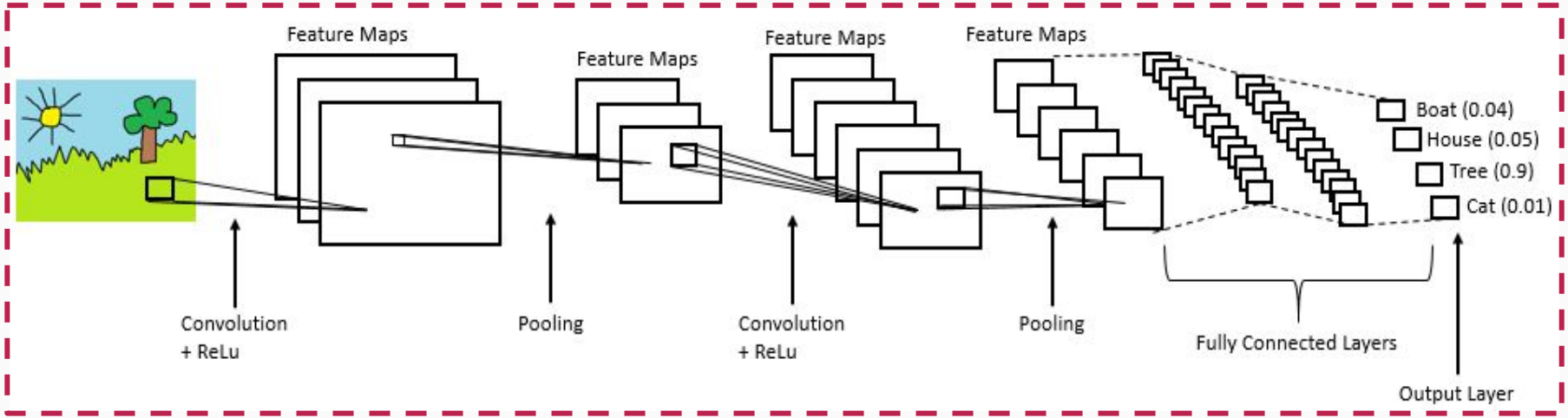


Max Pooling: The largest element is taken from feature map.

Average Pooling: It calculates the average of the elements in a predefined size image section.

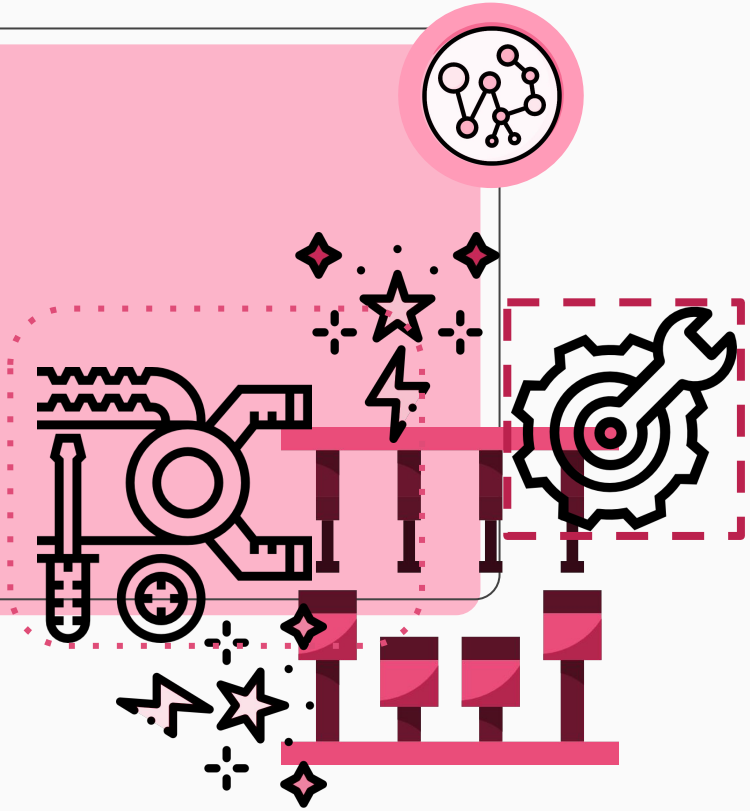
SureStart

FULLY-CONNECTED LAYERS



- The vector flattened matrix of the feature map or vectors are fed into a fully connected layer.
- A fully connected layer mathematically sums the features from all the combinations of the features of the previous layer and gets it ready for classification.

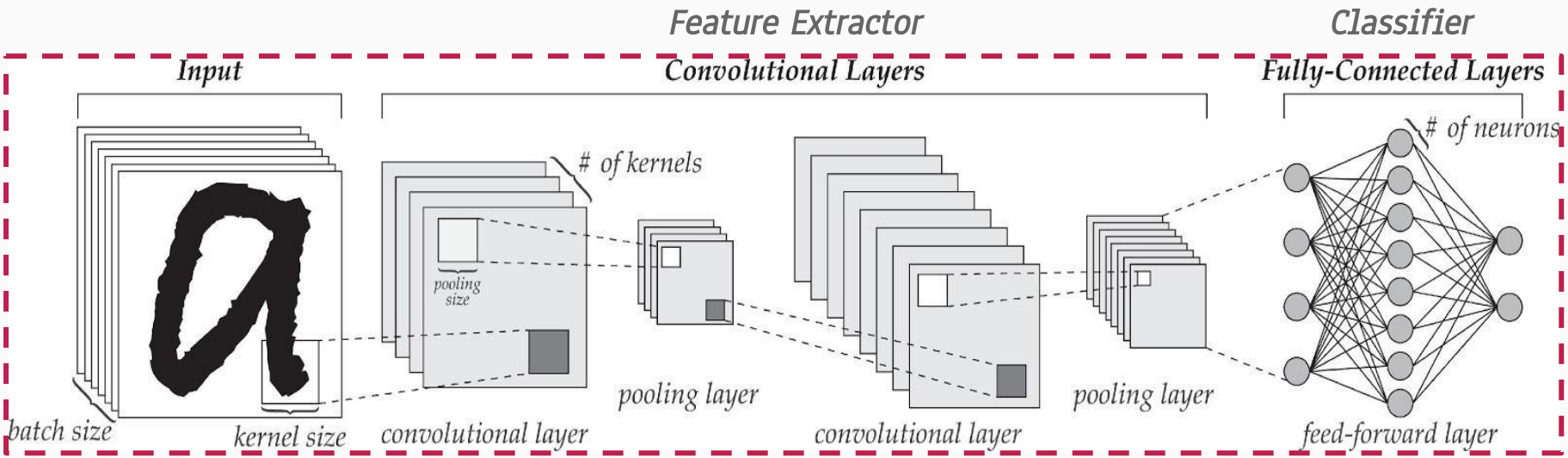
CNN Components



INTUITIVE UNDERSTANDING OF CNN COMPONENTS

- Convolutional Neural Network (CNN) is a type of feed-forward artificial neural network.
- A CNN is minimally composed of four components:
  - A convolution (CONV) layer is a feature extraction layer that extract discriminating feature of given input.
  - An activation layer (usually ReLU) introduces nonlinearity to the network to help model generalize.
  - A pooling layer is a down-sampling layer that select the most salient (relevant) features.
  - A fully connected layer combines the features from previous layers to get it ready for classification.

CNN OPERATION

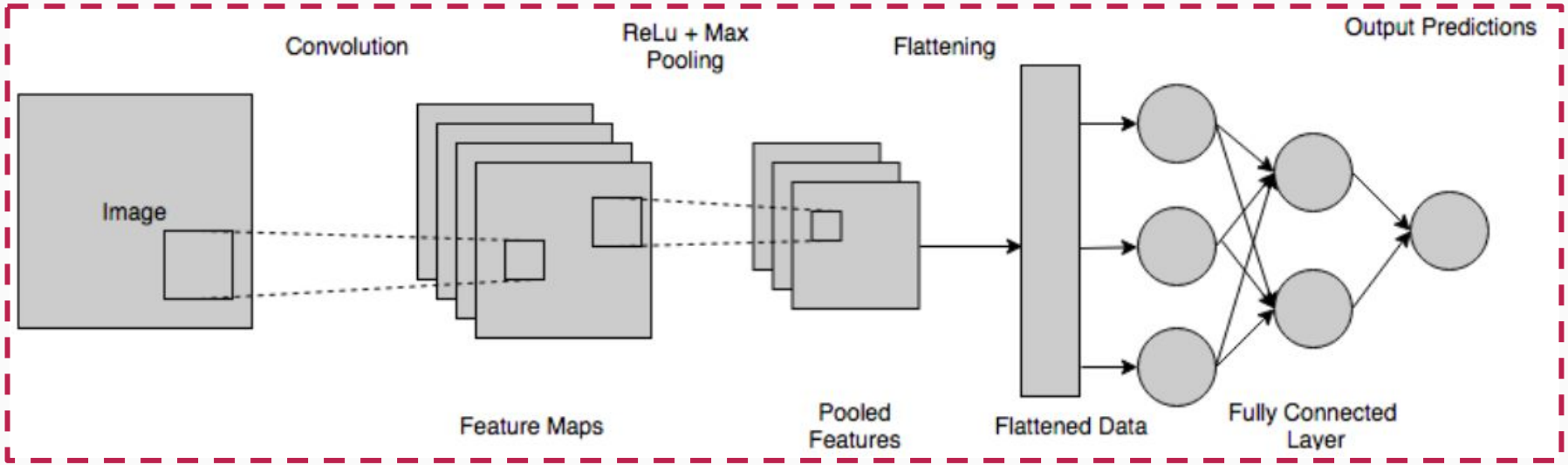


- **Feature Extractor:** CONV layer followed by an activation layer (ReLU) and a pooling layer work as a feature extractor separating and identifying various features of the image.

- **Classifier:** A fully connected layer utilizes the extracted features of previous stages from the convolution and the pooling operation; and predicts the class.

SureStart

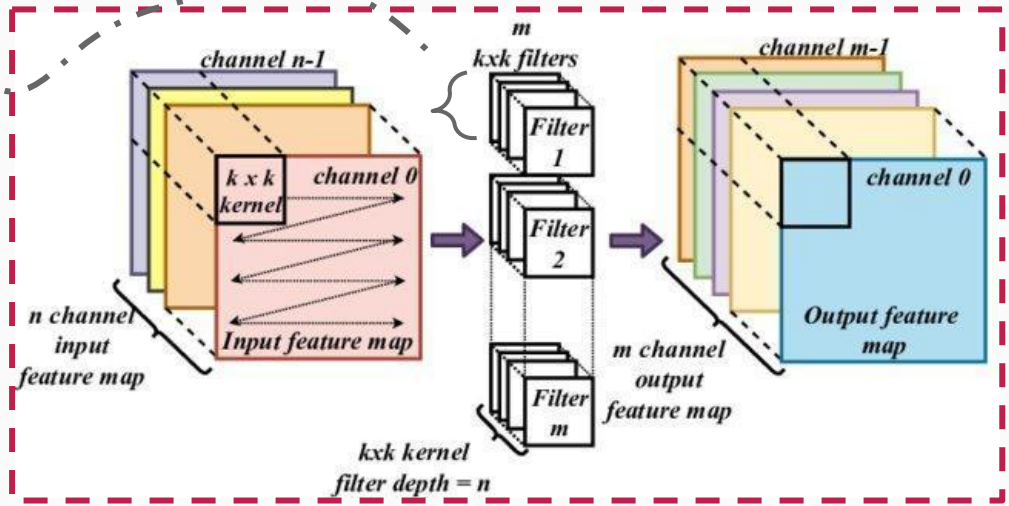
FLATTENING



- Between the pooling and fully connected layers, there is another operation called **flattening**.
- Flattening converts the 2-dimensional output of pooling (i.e., the feature map) into a single linear vector.
- The process of flattening takes the numbers row by row, and concatenate them to form a long feature vector.

CONV LAYER PARAMETERS [PART I]

Depth –
Number of filters used for the convolution operation.



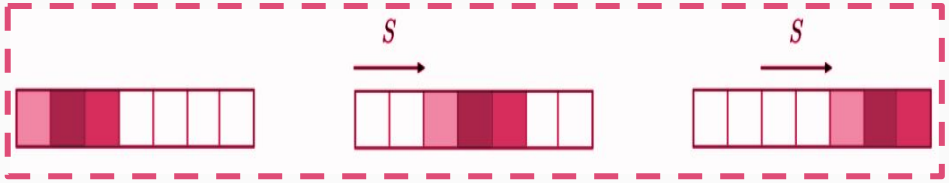
- The output feature map is the result of applying a filter to an image (or in case of an *n* layer CNN, an input feature map).

- The size of the feature map is controlled by three parameters: depth of the filter, padding and stride.

CONV LAYER PARAMETERS [PART 2]

Stride –

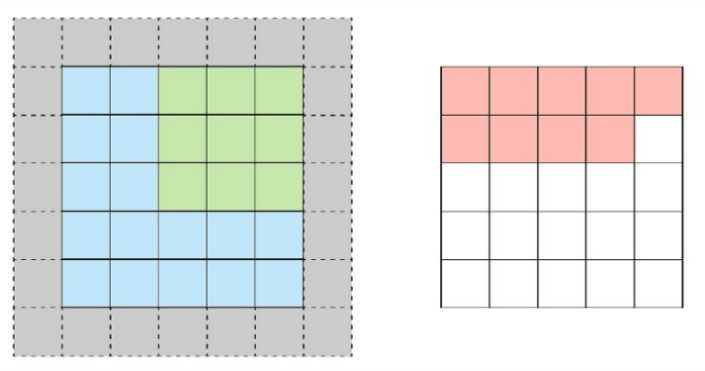
Stride decides by what step the kernel will move throughout the input image.



A movement of filters in the input matrix with a stride size 2.

Padding –

The input matrix is added with zeros around the border matrix. As the size of feature map is always smaller than the input matrix, it is used to maintain the dimension of output as in input. Padding is applied to prevent the feature map from shrinking.



Zero padding added to the image

Feature map

SureStart

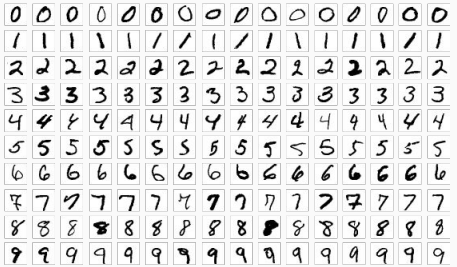
Digit Classification

MNIST HANDWRITTEN DIGIT CLASSIFICATION

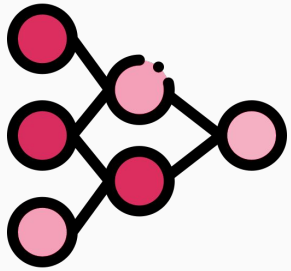
- The MNIST is a standard dataset for handwritten digit classification problem.

- The dataset consists of greyscale images of handwritten digits 0 to 9.

- MNIST database is the subset of NIST (National Institute of Standards and Technology) database.



data and labels



network training

0
1
2
3
4
5
6
7
8
9

predictions

Today's project will help you:

- To get familiar with a standard image dataset.
- To learn how to apply machine learning algorithms to solve an image recognition problem.
- To explore changes in model behavior in relation to changes in model topology and parameters.

SureStart

ABOUT THE MNIST DATASET

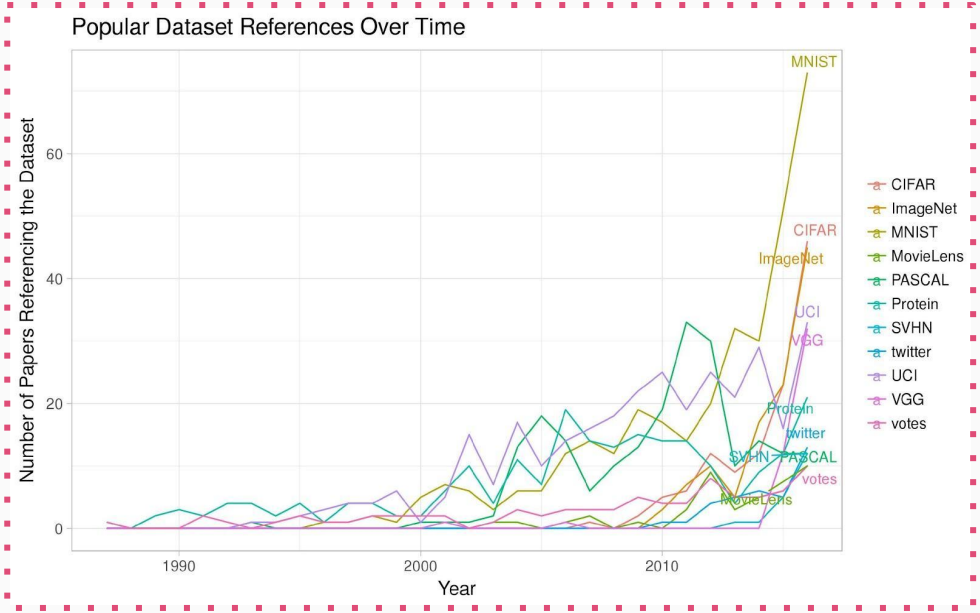
- Training set contains 60,000 examples; test set contains 10,000 examples.
- The sample image size is 28 x 28.
- The dataset contains 10 classes representing integer values from 0 to 9.
- Every sample is marked with its true label.
- Each row of the downloaded excel file contains an image input represented by a 784-sized vector (28×28 pixels), which is the color intensity value between 0 and 255.

MNIST RESEARCH

Over the years, successful **supervised** learning classification have been done with various ML techniques including *linear classifiers, nonlinear classifiers, K-nearest neighbors, SVMs, and convolutional neural nets.*

Multilayer CNN using Keras with Theano and Tensorflow shows the highest accuracy and lowest error rate on this MNIST dataset in comparison with other widely used ML algorithms.

The graph shows the statistics of the popular datasets referenced over time in NIPS papers till 2016. MNIST reigned any other dataset till the date.



Source: https://www.kaggle.com/benhamner/popular-datasets-over-time

SureStart

MNIST CLASSIFICATION STEPS

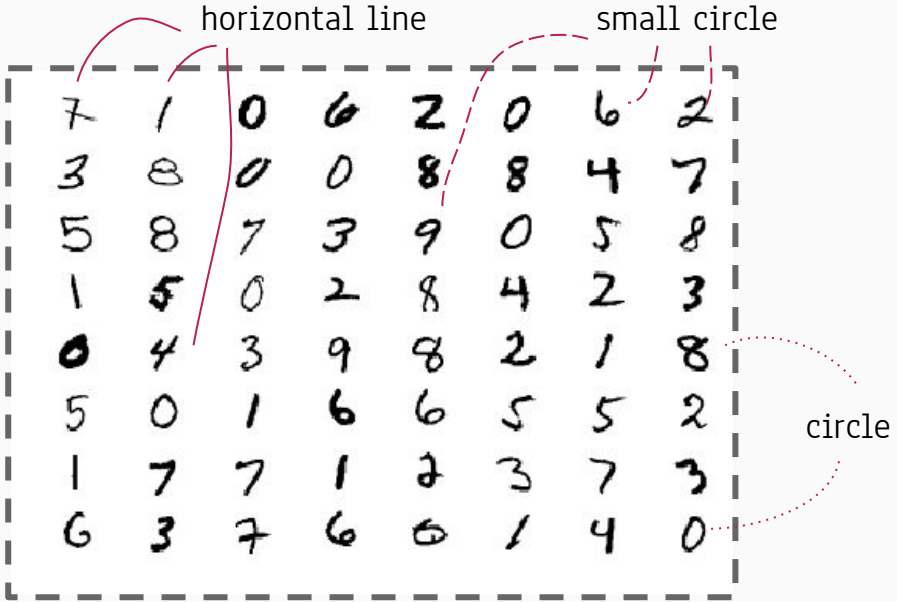


Figure source

1. Prepare and load the dataset.

2. Perform data preprocessing, e.g., scale the dataset.

3. Select a machine learning model (such as a CNN) and define its architecture.

4. Specify loss function and model optimizer.

5. Split the dataset into training, validation, and test sets.

6. Train and validate the model. Repeat this step until you see the accuracy increasing. Here, you are learning the model parameters.

7. Evaluate the final set of model parameters with test dataset.

SureStart