

Classification

MNIST

This chapter uses the *MNIST* dataset, which is a set of 70k small images of digits handwritten by high school students.

Fetching the Data set

```
from sklearn.datasets import fetch_openml
mnist = fetch_openml('mnist_784', as_frame=False)
print(mnist.DESCR)
```

Binary Classifiers

Definition 0.1

Binary Classifiers are capable of distinguishing between just two classes, for example 5 and not 5.

Stochastic Gradient Descent (SDG) classifier

Definition 0.2

SDG Classifiers are a type of machine learning algorithm used for classification tasks, based on the stochastic gradient descent optimization technique.

1. SDG's start with an initial set of parameters, or weights assigned randomly or based on a heuristic
2. In each iteration the classifier computes the gradient of the loss function with respect to the parameters using a small random subset of the training data (stochastic). This is efficient for large datasets.
 - **Loss function:** measures how well the model's predictions match the actual target values (labels) in the training data. It quantifies the error between the predicted output and the true output. The goal is to minimize this value.
 - The gradient represents the direction of the steepest ascent of a function, meaning its the direction in which the loss function increases most rapidly. We go in the opposite direction of this for highest accuracy.
3. The parameters are updated in the opposite direction of the gradient to minimize loss.
 - The parameters or weights of the model are updated by subtracting a fraction of the gradient from the current parameter values.
 - $\text{new_parameter} = \text{old_parameter} - \text{learning_rate} * \text{gradient}$. We subtract because we want the opposite direction of the gradient.
4. 2 and 3 are repeated for a fixed number of iterations or until converge criteria is met
5. After parameters are learned, the classifier can be used to make prediction on new data by computing a score for each class and assigning the class with the highest score as the predicted class.

SDG classifiers are often used for their efficiency and effective especially on large scale datasets.