

LAB 13

Perceptron



You have an email dataset labeled as either spam (1) or not spam (0), and you want to build a perceptron for classification. The dataset has been expanded to include four features for each email:

Features:

1. The number of times the word 'free' appears (Feature 1)
2. The number of dollar signs '\$' in the email (Feature 2)
3. The length of the email text (Feature 3)
4. The presence of the word 'urgent' (Feature 4)

Weights:

Following are initial weights and bias:

- Weight 1: 0.5
- Weight 2: -0.2
- Weight 3: 0.1
- Weight 4: 0.3
- Bias: 0.1

Algorithm:

1. Calculate the weighted sum of inputs (x) for each email.
$$\text{activation}_w(x) = \sum_i w_i \cdot f_i(x) = w \cdot f(x)$$
2. Update the weights and bias for one epoch (one pass through the training dataset).

$$w = w + y^* \cdot f$$

If the activation is:

- Positive, output 1
- Negative, output -1

Training and testing:

Split the dataset into training and testing sets. Use 80% of the data for training and 20% for testing.

Apply the trained perceptron model to the testing data to predict whether the test email is spam or not spam using the testing dataset.

Accuracy Evaluation:

- Calculate the accuracy of the perceptron model on the testing dataset using the following formula:

$$\text{Accuracy} = (\text{Number of Correct Predictions} / \text{Total Number of Predictions}) * 100$$