

# L<sup>A</sup>T<sub>E</sub>X Author Guidelines for CVPR Proceedings

Zeyu Zuo  
The University of Adelaide  
a1894365@adelaide.edu.au

## Abstract

*Predict and diagnose diabetes is a critical challenge in medical field. If diabetes is detected in its early stages, it can be managed and controlled through treatment, which can delay complications or even prevent them altogether. This article aims to using deep learning algorithm known as single layer perceptron to analyze patients data and after several times training, the perceptron could generate a model to predict a patient whether or not has diabetes*

## 1. Introduction

Diabetes is often described as a silent epidemic that affects millions of lives. The significant impact of diabetes on premature health problems, higher mortality rates, decreased life expectancy, underscores its importance as a public health concern. While diabetes is a complex and chronic condition, it often begins silently, without obvious symptoms. The task is to build a machine learning model to accurately predict the patients in the dataset have diabetes or not. The single-layer perceptron is the algorithm will be use to solve this task. A single-layer perceptron can be viewed as an artificial network and use a supervised learning approach. The perceptron functions ,as a linear combiner, quantizes its output into one of two distinct values. Its connection weights and threshold, or bias, can remain constant or be adjusted using gradient descent algorithms. The original procedure for converging the perceptron's weights was first proposed by Rosenblatt [3] , who demonstrated that if inputs from two classes are linearly separable (i.e., they lie on opposite sides of a hyperplane), the perceptron algorithm will converge, effectively positioning the decision boundary between these two classes.

## 2. Dataset

The dataset is collected from women in Pima Indian, provided from the National Institute of Diabetes and Digestive and Kidney Diseases. The datasets used for experimentation and coding is a preprocessed binary classification

dataset . The first column of the data being the class id +1 or -1 and the rest consists of several medical predictor variables include factors:

- Number of times pregnant
- Plasma Glucose Concentration at 2 Hours in an Oral-GlucoseTolerance Test (GTIT)
- Diastolic BloodPressure (mm Hg)
- Triceps SkinFoldThickness (mm)
- 2-Hour Serum Insulin (Uh/ml)
- Body Mss Index (Weight in kg/(Heightin in))
- Diabetes Pedigree Function
- Age (years)

These eight [4] variables were chosen due to their recognized significance as risk factors for diabetes, both in Pima Indians and in other populations.

## 3. Method

Perception can be represented as following:

$$\text{sign}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (1)$$

$$f(x) = \text{sign}(w^T \cdot x + b) \quad (2)$$

where  $w$  is weight,  $x \in \mathbb{R}^n$  is input feature,  $b$  is bias.

The output is either +1 or -1. Therefore, during the training process, we can assign +1 to represent patients have diabetes and -1 to represent patients without diabetes. This way, during the prediction process, the perceptron can classify data into two different class.

### 3.1. Loss function

Similar to conventional neural networks, perceptron loss function can constructed derive gradients of the loss function with respect to the parameters  $w$  and  $b$ . Then utilize gradient descent to update the parameters accordingly. [?] Our goal is to find a hyperplane that can effectively separate the two classes of data as much as possible. A natural choice for the loss function is the total number of misclassified points, but such a function is not continuous and not easily optimized. Another approach is to consider the sum of distances from all misclassified points to the hyperplane as the loss function and minimize this loss function. Drawing an analogy to the formula for the distance from a point in three-dimensional space to a plane, Suppose that the hyperplane is represented as  $w^T \cdot x + b = 0$ , and there are  $m$  misclassified points out of  $n$  total points. Then, the sum of distances from all misclassified points to the hyperplane is given by:

$$Loss(w, b) = \sum_{i=1}^m -y_i(w^T \cdot x_i + b) \quad (3)$$

Gradients of the loss function:

$$\nabla_w Loss(w, b) = - \sum_{i=1}^m y_i x_i \quad (4)$$

$$\nabla_b Loss(w, b) = - \sum_{i=1}^m y_i \quad (5)$$

### 3.2. Training process

In summary, the original form of the Perceptron learning algorithm is as follows:

Input: Training set  $T$ , learning rate  $\eta$

Output:  $w, b$

Steps:

- (1) Randomly Initialize  $w, b$ ;
- (2) Select data from the training set, get  $(x_i, y_i)$ ;
- (3) If misclassified point, update the parameters from equl(4) and equl(5):

$$w = w + \eta y_i x_i \quad (6)$$

$$b = w + \eta y_i \quad (7)$$

where  $\eta$  is the learning rate.

- (4) Go to (2) until there are no misclassified points in the training set.

## 4. Experiment Analysis

The following metrics are normally used to evaluate the performance of machine learning models in binary classification tasks and to adjust model parameters, features,

or algorithms based on the evaluation results to achieve satisfactory outcomes. [1]

Accuracy: Accuracy measures the proportion of correctly classified instances out of the total.

$$Accuracy = (TP+FP) / (TP + TN + FP + FN)$$

Precision: Precision, also known as positive predictive value, measures the proportion of true positive predictions (correctly predicted positive instances) out of all positive predictions. It is calculated as:

$$Precision = TP / (TP + FP)$$

Recall: Recall, also known as sensitivity or true positive rate, measures the proportion of true positive predictions out of all actual positive instances. It quantifies the model's ability to correctly identify positive instances. It is calculated as:

$$Recall = TP / (TP + FN)$$

F1 Score: The F1 Score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is particularly useful when dealing with imbalanced datasets. It is calculated as:

$$F1 \text{ Score} = 2 * (Precision * Recall) / (Precision + Recall)$$

method	accuracy	recall	f1
perceptron	0.747370	0.747610	0.726800
logistic	0.769560	0.722570	0.731980
SVC-RBF	0.770890	0.720200	0.730350
mlp	0.865700	0.817870	0.826410

### 4.1. Effect and Deficiency

After printing each time accuracy , we can find that the Perceptron is effective at learning linear decision boundaries in feature space and can classify data points into two classes by drawing a straight line, plane, or hyperplane that separates them. The Perceptron's algorithm is relatively straightforward to implement. It involves computing weighted sums of input features, applying an activation function, and updating weights based on classification errors. Implemented with a fixed linearly separate data, the Perceptron algorithm converges in a fixed number of iterations to find a solution. However, the perceptron model still has constraints. 1)The input vectors need to be fed to the network one by one or in groups (batches) to allow for adjustments based on outcomes. 2)The output of perceptron is binary result (0,1) because limited by the transfer function. 3) Perceptron can effectively classify input sets that are linearly separable, but it cannot classify non-linear input vectors. [2]

### 4.2. Conclusion

The primary purpose of the study is using a deep learning algorithm to recognize elevated risk of diabetes in patients at an earlier stage. According to result accuracy , we could

find the advantage of each method depends on the specific characteristics of dataset and problem. For linearly separable data with a limited number of features, a simple method like the Perceptron or Logistic Regression may suffice. If the data is non-linear or has complex relationships, SVC with an RBF kernel or MLP could be advantageous. It's essential to choose the method that aligns with the problem's complexity and the available data.

## References

- [1] Gürol Canbek, Seref Sagiroglu, Tugba Taskaya Temizel, and Nazife Baykal. Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. In *2017 International Conference on Computer Science and Engineering (UBMK)*, pages 821–826, 2017. [2](#)
- [2] Yi-Chung Hu. Classification performance evaluation of single-layer perceptron with choquet integral-based topsis. *Applied Intelligence*, 29:204–215, 2008. [2](#)
- [3] Frank Rosenblatt. *The perceptron: a theory of statistical separability in cognitive systems (Project Para)*. Cornell Aeronautical Laboratory, 1958. [1](#)
- [4] Jack W Smith, James E Everhart, WC Dickson, William C Knowler, and Robert Scott Johannes. Using the adap learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the annual symposium on computer application in medical care*, page 261. American Medical Informatics Association, 1988. [1](#)