Examples of Machine Learning

Cristhian Calderon

June 12, 2024

Abstract

I will show examples of machine learning models. We will particularly use Perceptron, GaussianNB, svm.SVC (Support Vector Classifier.Support Vector Classifier)

1 Introduction

What is Machine Learning?

Machine Learning (ML) is a subfield of artificial intelligence (AI) focused on the development of algorithms and statistical models that enable computers to perform specific tasks without explicit instructions. By learning from data, these systems can identify patterns, make decisions, and improve their performance over time. The core idea behind machine learning is to enable computers to learn from experience and adapt their actions accordingly.

Historical Context: The concept of machine learning has its roots in the mid-20th century, with the pioneering work of Alan Turing and his proposal of the Turing Test as a measure of machine intelligence. In the following decades, foundational theories were developed, such as the perceptron model by Frank Rosenblatt in 1957, which laid the groundwork for neural networks. The evolution of machine learning has been closely tied to advancements in computing power and data availability, leading to the significant breakthroughs we witness today.

Types of Machine Learning: Machine learning algorithms are generally categorized into three main types:

Supervised Learning: In this approach, the algorithm is trained on a labeled dataset, meaning the input data is paired with the correct output. The model learns to map inputs to outputs and is evaluated based on its accuracy in predicting the labels of new, unseen data. Common algorithms include linear regression, support vector machines, and neural networks.

Unsupervised Learning: Unlike supervised learning, unsupervised learning deals with unlabeled data. The goal is to infer the natural structure present within a set of data points. Techniques such as clustering (e.g., k-means, hierarchical clustering) and dimensionality reduction (e.g., PCA, t-SNE) are widely used in this category.

Reinforcement Learning: This type of learning is based on the concept of agents interacting with an environment. The agent learns to achieve a goal by receiving rewards or penalties based on its actions. Over time, it develops a strategy to maximize cumulative rewards. Reinforcement learning has been successfully applied in various domains, including robotics, gaming (e.g., AlphaGo), and autonomous vehicles.

Applications of Machine Learning: The versatility of machine learning has led to its adoption in numerous fields:

Healthcare: Predicting disease outbreaks, personalizing treatment plans, and analyzing medical images. Finance: Fraud detection, algorithmic trading, and risk assessment. Marketing: Customer segmentation, sentiment analysis, and recommendation systems. Transportation: Route optimization, autonomous driving, and predictive maintenance. Natural Language Processing (NLP): Language translation, speech recognition, and chatbots. Challenges and Future Directions Despite its remarkable progress, machine learning faces several challenges. These include the need for large amounts of high-quality data, model interpretability, and ethical concerns related to bias and privacy. Addressing these issues is crucial for the responsible and equitable deployment of ML technologies.

Looking ahead, the future of machine learning promises exciting advancements. Areas such as deep learning, which mimics the human brain's neural networks, continue to push the boundaries of what machines can achieve. Additionally, interdisciplinary research is expected to yield innovative solutions, further integrating ML into everyday life and transforming industries worldwide.

2 Some examples to get started

2.1 Conditions for a rainy day

We will construct a model with four parameters: variance, skewness, kurtosis and entropy in order to predict rains.

variance	skewness	kurtosis	entropy	rain
-0.89569	3.0025	-3.6067	-3.4457	1
3.4769	-0.15314	2.53	2.4495	0
3.9102	6.065	-2.4534	-0.68234	0
0.60731	3.9544	-4.772	-4.4853	1
2.3718	7.4908	0.015989	-1.7414	0
-2.2153	11.9625	0.078538	-7.7853	0

Table 1: Conditions for rain. Where "1" represents a rainy day and "0" means that there is no rain.

2.2 Perceptron model

The perceptron is a fundamental machine learning model inspired by biological neurons. Developed by Frank Rosenblatt in 1957, it represents the simplest type of artificial neural network. A perceptron takes several binary inputs, applies corresponding weights, sums them up, and passes the result through an activation function to produce a single binary output. If the weighted sum exceeds a certain threshold, the perceptron activates, outputting one; otherwise, it outputs zero. This model can be used for basic binary classification tasks, where it learns to distinguish between two classes by adjusting the weights based on training data. Despite its simplicity, the perceptron laid the groundwork for more complex neural network architectures and deep learning.

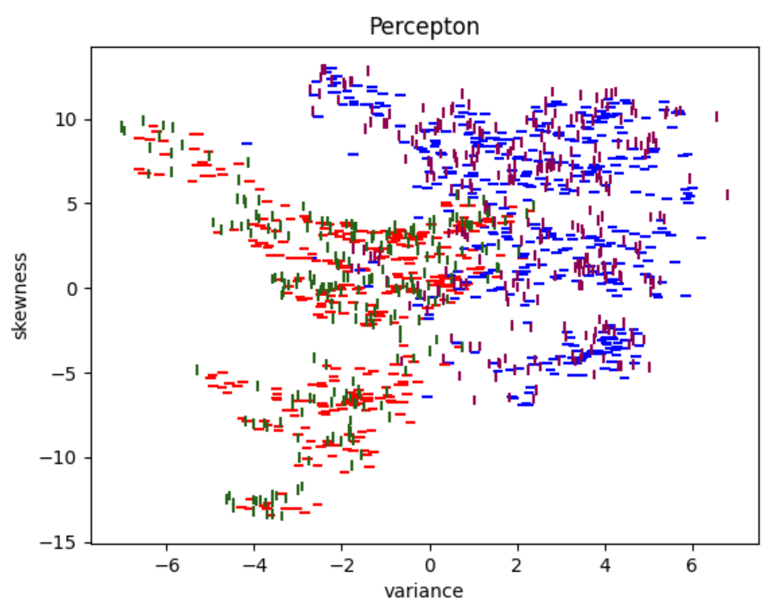

Figure 1: We take the first two conditions: variance and skewness and see how they are correlated and to visualize the predictions. The data red and blue are the training data where red means rainy day and blue means sunny day. The green and purple data are the predictions of the Percepton model where green means rainy and purple, no rains.

We take 40% of the data to train the model and the remaining 60% to test the predictability of the model.

$$\text{Correct} = 535 \tag{1}$$
$$\text{Incorrect} = 13 \tag{2}$$
$$\text{Accuracy} = 97.63\% \tag{3}$$

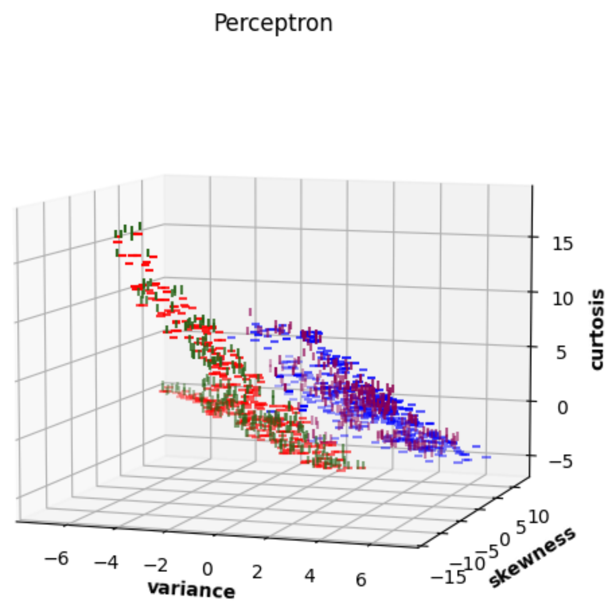

Figure 2: We take the first two conditions: variance and skewness and see how they are correlated and to visualize the predictions. The data red and blue are the training data where red means rainy day and blue means sunny day. The green and purple data are the predictions of the Percepton model where green means rainy and purple, no rains.

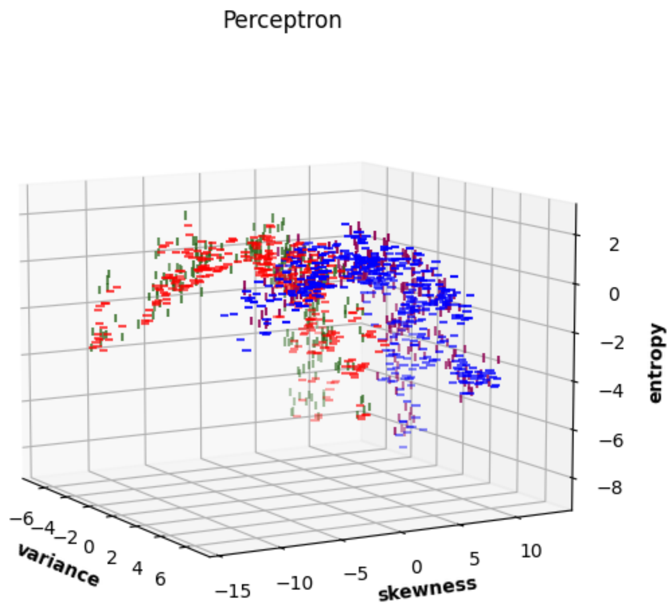

Figure 3: We take the first two conditions: variance and skewness and see how they are correlated and to visualize the predictions. The data red and blue are the training data where red means rainy day and blue means sunny day. The green and purple data are the predictions of the Percepton model where green means rainy and purple, no rains.

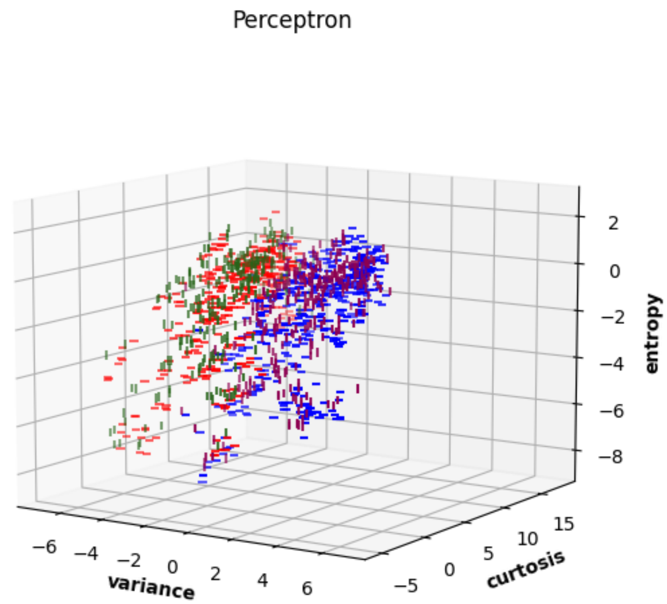

Figure 4: We take the first two conditions: variance and skewness and see how they are correlated and to visualize the predictions. The data red and blue are the training data where red means rainy day and blue means sunny day. The green and purple data are the predictions of the Percepton model where green means rainy and purple, no rains.

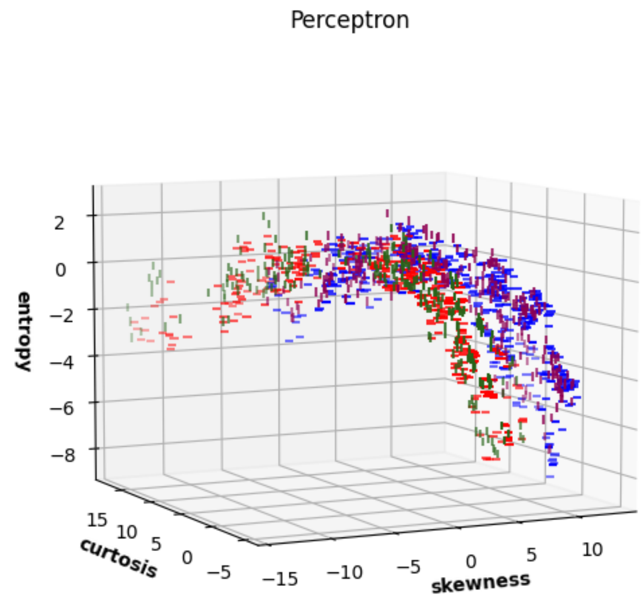

Figure 5: We take the first two conditions: variance and skewness and see how they are correlated and to visualize the predictions. The data red and blue are the training data where red means rainy day and blue means sunny day. The green and purple data are the predictions of the Percepton model where green means rainy and purple, no rains.

2.3 Super Vector Machine with Super Vector Clasiffied

Support Vector Machine (SVM) with the Support Vector Classifier (SVC) is a robust and versatile machine learning model used for classification tasks. SVM aims to find the optimal hyperplane that separates data points of different classes with the maximum margin. The SVC variant enhances this by allowing for some misclassification in the training data to better generalize to new data. SVMs can handle both linear and non-linear classification by using kernel functions, such as polynomial or radial basis function (RBF) kernels, to transform the input data into higher-dimensional spaces. This flexibility makes SVMs effective for a wide range of applications, including image recognition, text categorization, and bioinformatics.

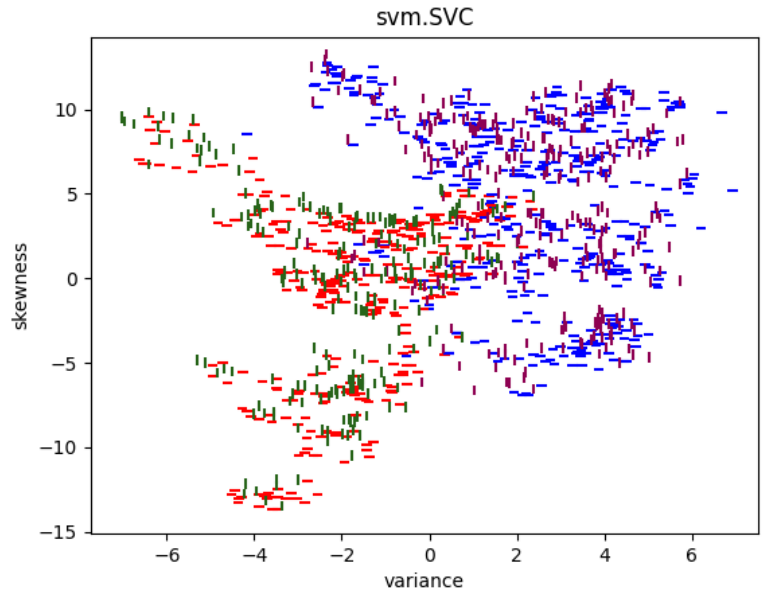

Figure 6: We take the first two conditions: variance and skewness and see how they are correlated and to visualize the predictions. The data red and blue are the training data where red means rainy day and blue means sunny day. The green and purple data are the predictions of the svm SVC model where green means rainy and purple, no rains.

$$\text{Correct} = 545 \tag{4}$$
$$\text{Incorrect} = 3 \tag{5}$$
$$\text{Accuracy} = 99.45\% \tag{6}$$

2.4 KNeighborsClassifier Model

The k-Nearest Neighbors (k-NN) classifier is a simple yet effective machine learning model used for classification tasks. The NeighborClassifier, a specific implementation of this model, operates on the principle of proximity, predicting the class of a given data point based on the majority class among its k closest neighbors in the feature space. The value of k is a crucial hyperparameter, influencing the classifier's performance and generalization ability. The k-NN classifier is non-parametric and instance-based, meaning it makes decisions based on the entire training dataset without making assumptions about the underlying data distribution. This approach is particularly useful for applications like pattern recognition, image classification, and recommendation systems, where the relationship between data points is naturally local and contextual.

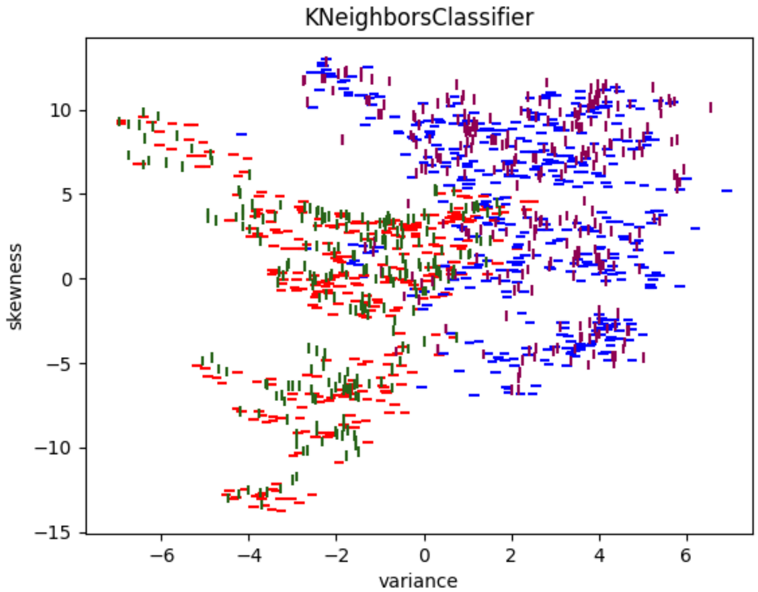

Figure 7: We take the first two conditions: variance and skewness and see how they are correlated and to visualize the predictions. The data red and blue are the training data where red means rainy day and blue means sunny day. The green and purple data are the predictions of the KNeighborsClassifier model where green means rainy and purple, no rains. It is important to note that we only use one neighbor to learn.

$$\text{Correct} = 548 \tag{7}$$

$$\text{Incorrect} = 0 \tag{8}$$

$$\text{Accuracy} = 100\% \tag{9}$$

2.5 KNeighborsClassifier Model

The k-Nearest Neighbors (k-NN) classifier is a simple yet effective machine learning model used for classification tasks. The NeighborClassifier, a specific implementation of this model, operates on the principle of proximity, predicting the class of a given data point based on the majority class among its k closest neighbors in the feature space. The value of k is a crucial hyperparameter, influencing the classifier's performance and generalization ability. The k-NN classifier is non-parametric and instance-based, meaning it makes decisions based on the entire training dataset without making assumptions about the underlying data distribution. This approach is particularly useful for applications like pattern recognition, image classification, and recommendation systems, where the relationship between data points is naturally local and contextual.

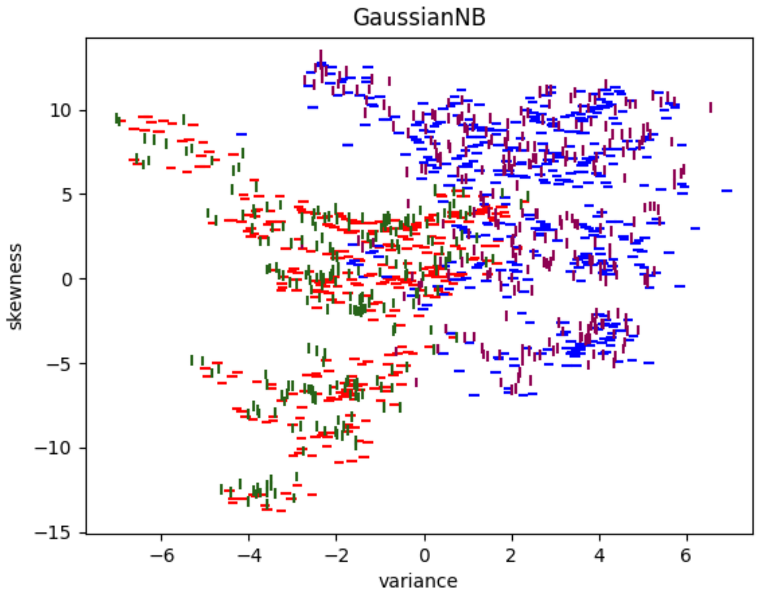

Figure 8: We take the first two conditions: variance and skewness and see how they are correlated and to visualize the predictions. The data red and blue are the training data where red means rainy day and blue means sunny day. The green and purple data are the predictions of the GaussianNB model where green means rainy and purple, no rains. It is important to note that we only use one neighbor to learn.

$$\text{Correct} = 467 \tag{10}$$

$$\text{Incorrect} = 81 \tag{11}$$

$$\text{Accuracy} = 85.22\% \tag{12}$$