Support Vector Machine

January 2023

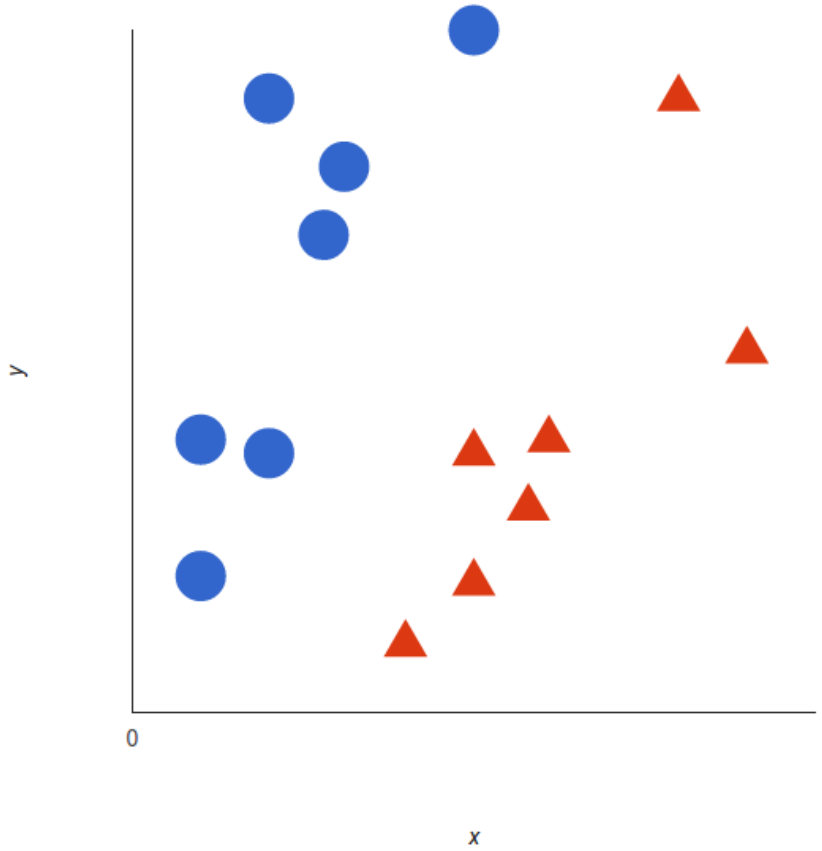
SVM

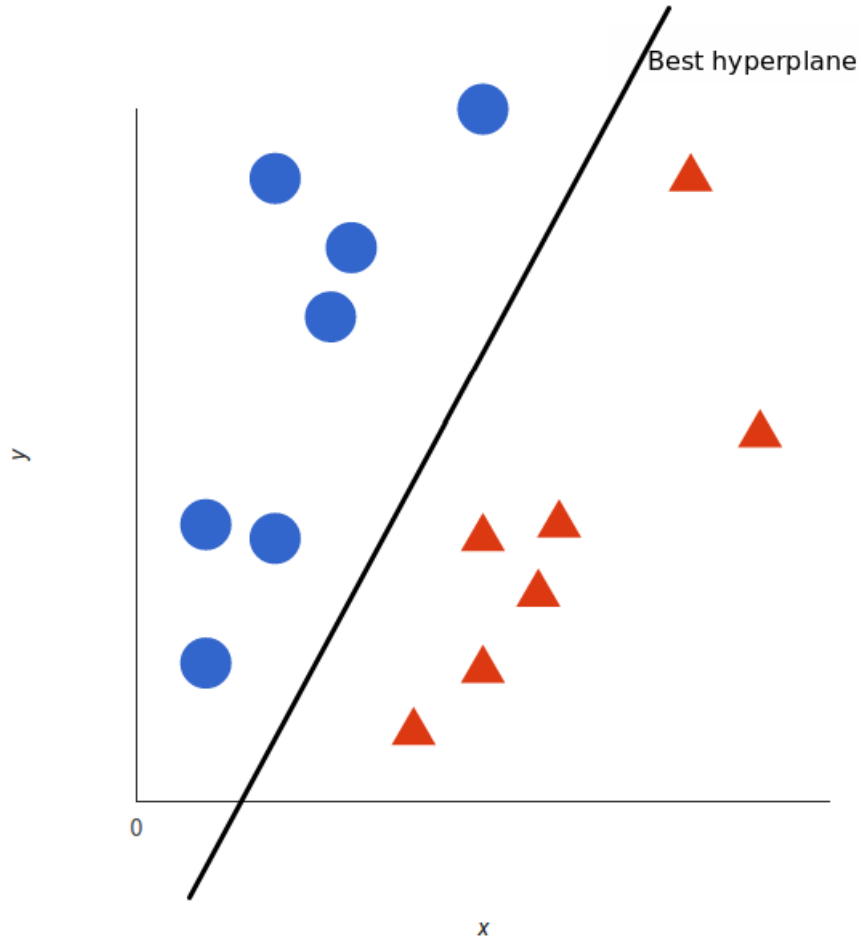
- A Support Vector Machine (SVM) is a type of supervised machine learning algorithm that can be used for classification or regression tasks.
- The goal of an SVM is to find the best boundary (or "hyperplane") that separates different classes of data in a high-dimensional space.

Applications

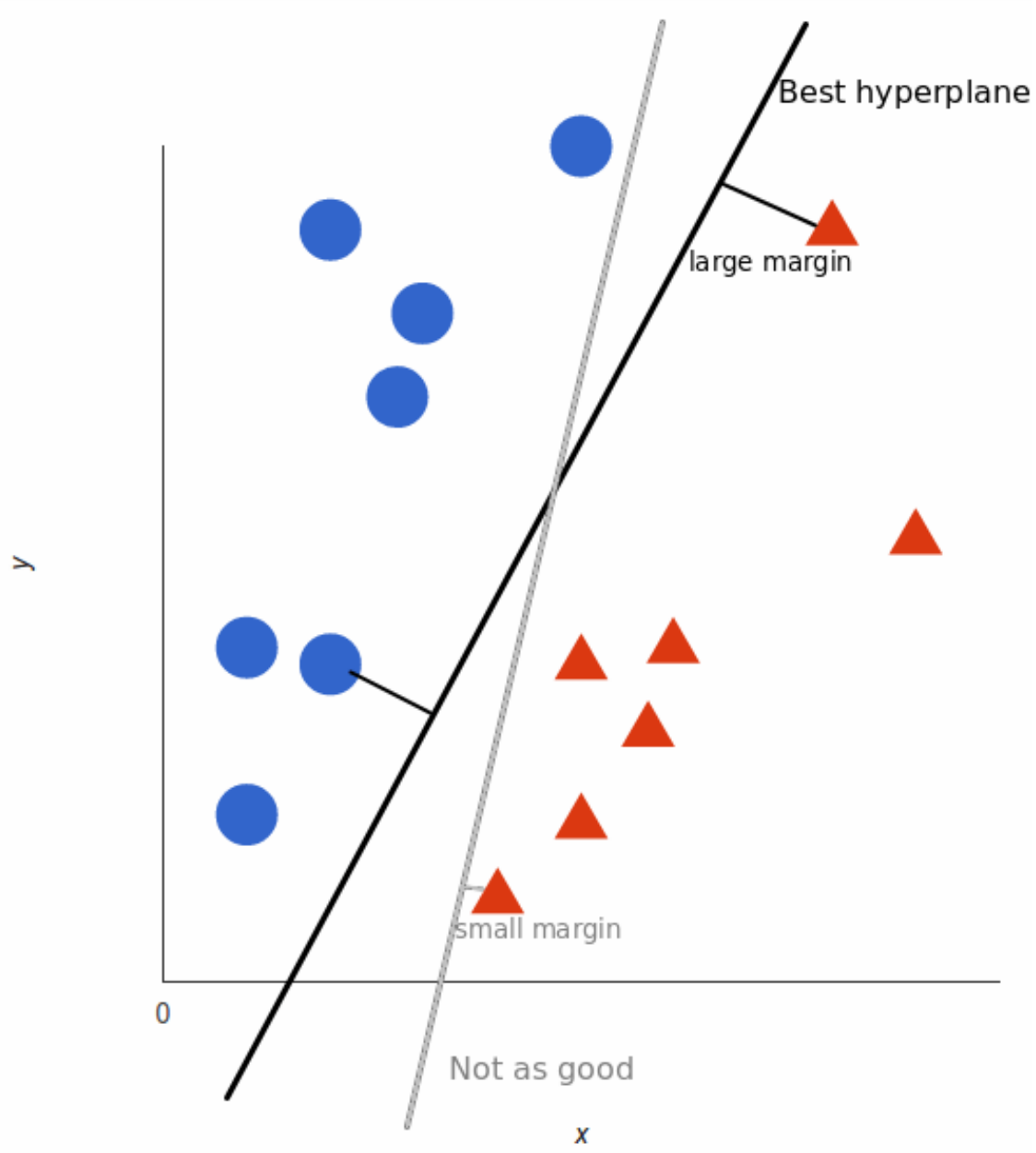
- In the case of classification, an SVM tries to find a boundary that separates the different classes of data with the largest margin, which is the distance between the boundary and the closest data points from each class.
  - This is known as the "maximum margin classifier" and it is used to make predictions about new data points.
- SVMs are particularly useful when the data is not linearly separable, which means that a straight line or a plane cannot be used to separate the different classes of data.
  - The SVM uses a technique called the "kernel trick" to project the data into a higher-dimensional space where it becomes linearly separable.
- SVMs are known to be effective when the number of features is much greater than the number of samples, and when the data is sparse.
  - They are also commonly used in applications such as image and text classification, bioinformatics, and natural language processing.

How does it work?

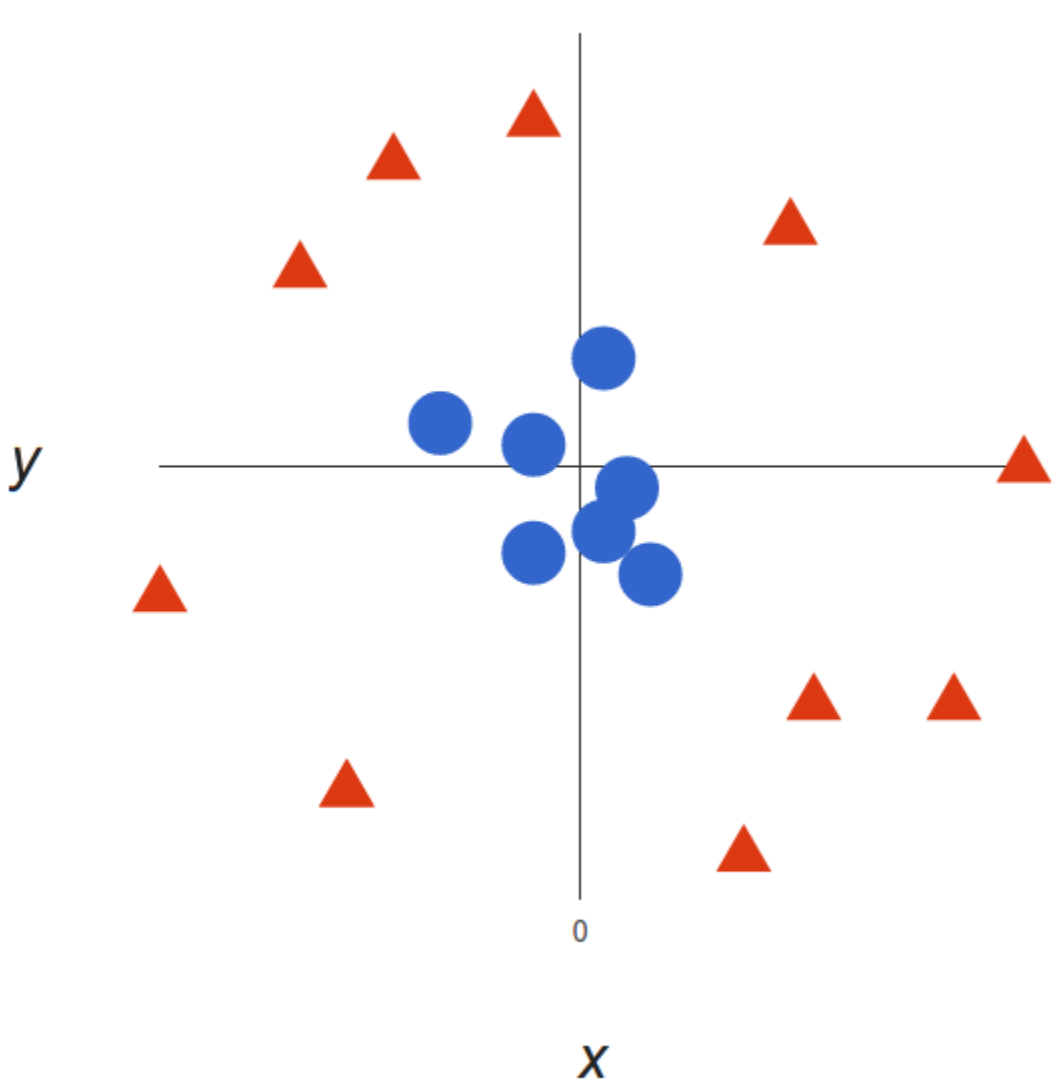


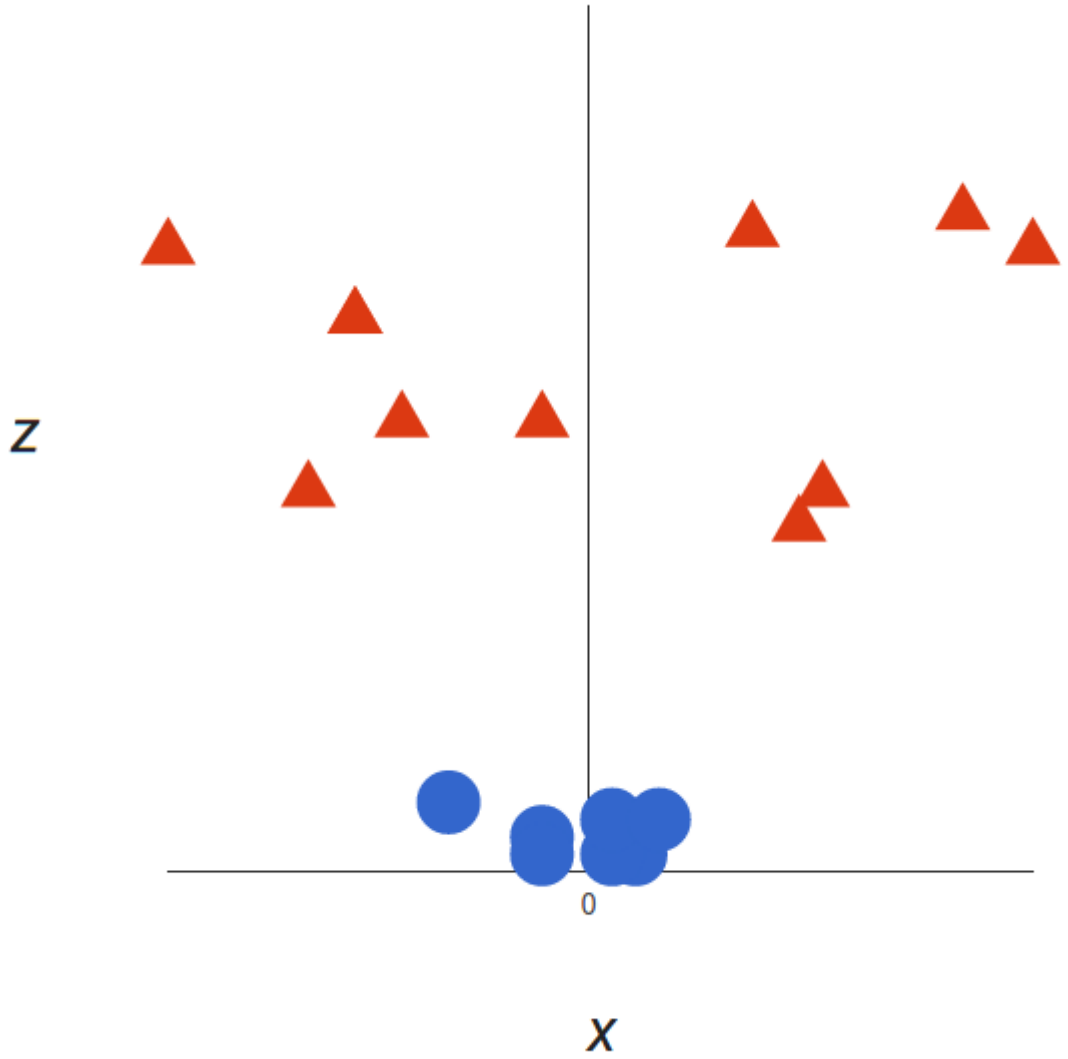


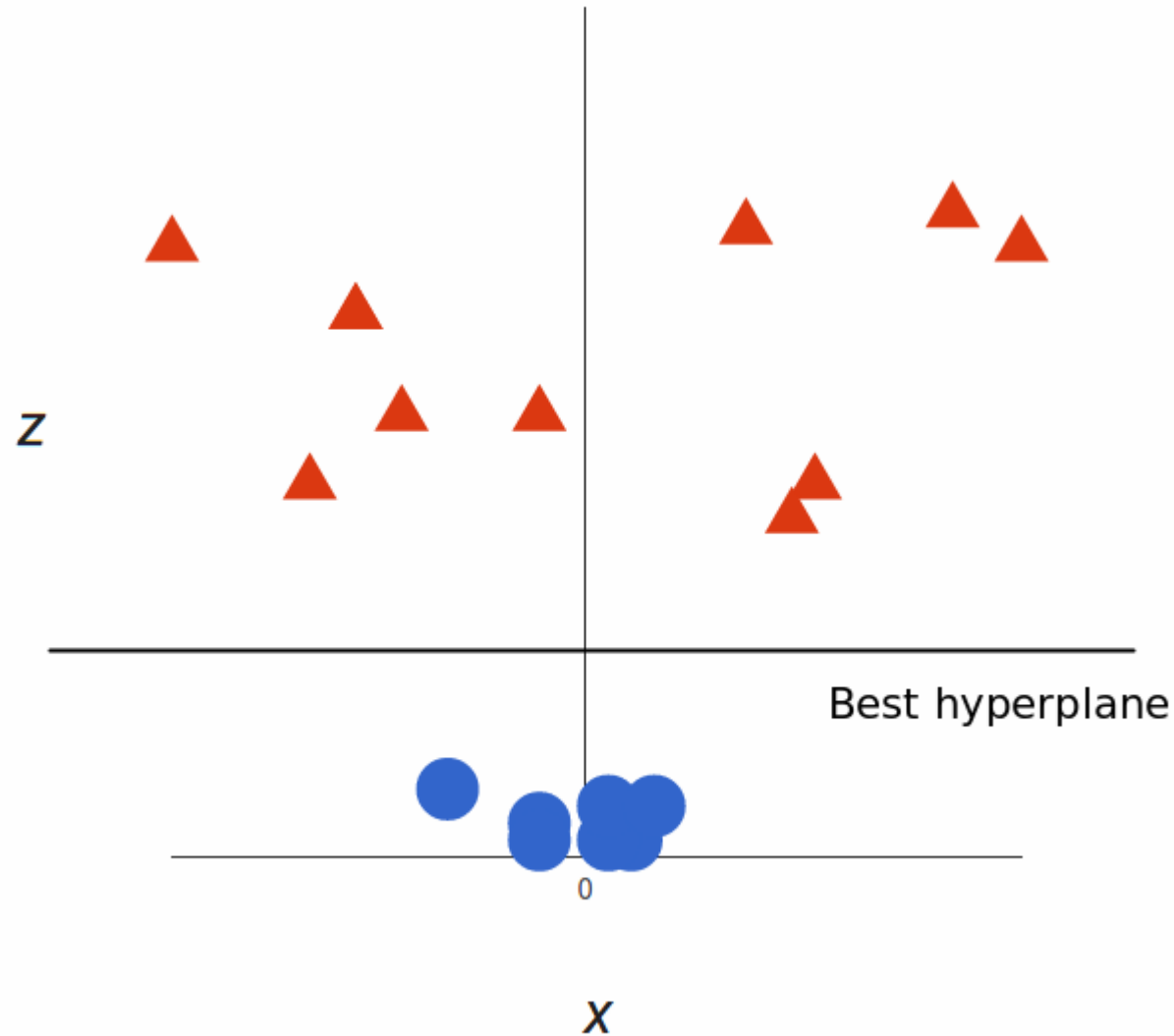
Best hyperplane?



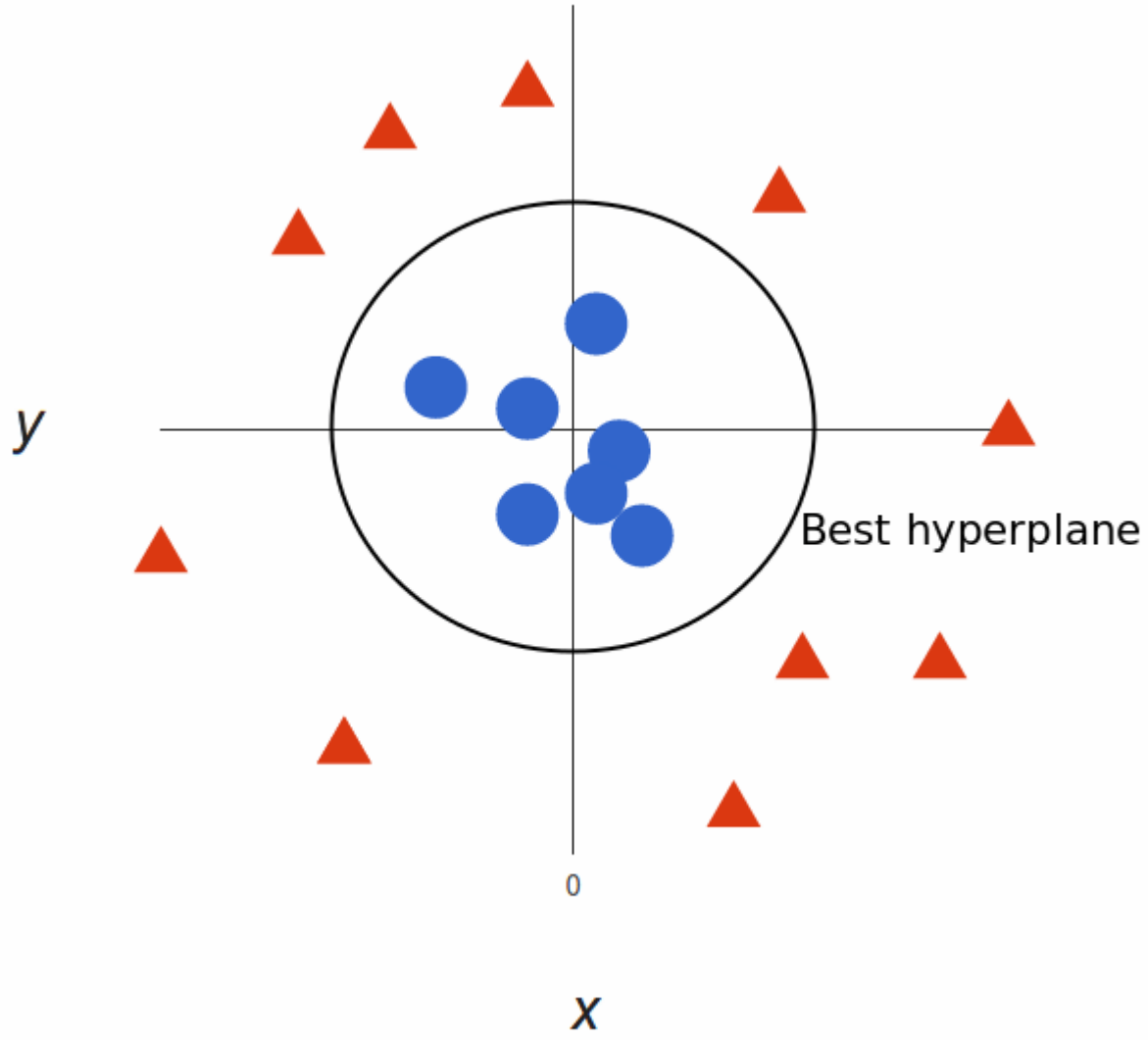
Non-Linear Data






z
Best hyperplane
0
x

Back to original view



Using a kernel

- In our example we found a way to classify nonlinear data by cleverly mapping our space to a higher dimension.
- However, it turns out that calculating this transformation can get pretty computationally expensive: there can be a lot of new dimensions, each one of them possibly involving a complicated calculation.
- Doing this for every vector in the dataset can be a lot of work, so it’d be great if we could find a cheaper solution.

What is a kernel?

- A kernel is a mathematical function that is used in the Support Vector Machine (SVM) algorithm to map the input data points into a high-dimensional feature space.
- The purpose of the kernel is to enable the SVM to find a boundary that separates the different classes of data even when the data is not linearly separable in the original feature space.
- The kernel function takes as input two data points from the original feature space and computes a dot product between them in the high-dimensional feature space.
- This dot product is then used to compute the similarity between the two data points.

Types of kernel functions

- Linear kernel: This kernel is used when the data is linearly separable in the original feature space. It maps the data points to the same dimension as the original feature space.
- Polynomial kernel: This kernel is used when the data is not linearly separable in the original feature space. It maps the data points to a higher-dimensional space where the data becomes linearly separable.
- RBF (Radial Basis Function) kernel: This kernel is also used when the data is not linearly separable in the original feature space. It maps the data points to an infinite-dimensional space where the data becomes linearly separable.
- Sigmoid kernel: This kernel is used when the data is not linearly separable in the original feature space. It maps the data points to a higher-dimensional space where the data becomes linearly separable.

Basic algorithm

- First, the SVM algorithm maps the input data points into a high-dimensional feature space, using a mathematical function called a kernel. The kernel function allows the SVM to find a boundary that separates the different classes of data even when the data is not linearly separable.
- Next, the SVM algorithm finds the hyperplane that maximizes the margin, which is the distance between the boundary and the closest data points from each class. This is known as the "maximum margin classifier" and it is used to make predictions about new data points.
- The SVM algorithm then finds the support vectors, which are the data points that are closest to the boundary and have the greatest impact on determining the position of the boundary.
- Once the support vectors are identified, the SVM algorithm uses them to define the boundary. This boundary is known as the "decision boundary" and it separates the different classes of data in the feature space.
- Finally, the SVM algorithm uses the decision boundary to classify new data points. A new data point is classified based on which side of the boundary it falls on.