Machine Learning

| Kernel Trick

Prof. Dr. Mostafa Elhosseini

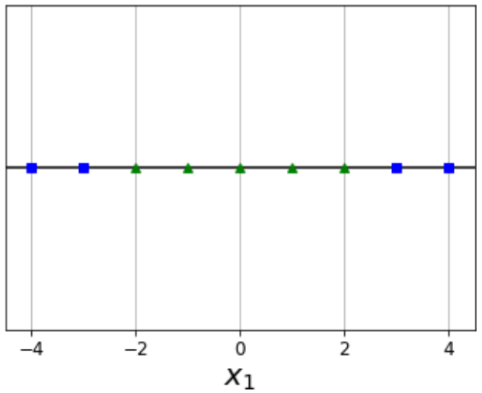
https://youtube.com/ElhosseiniAcademy

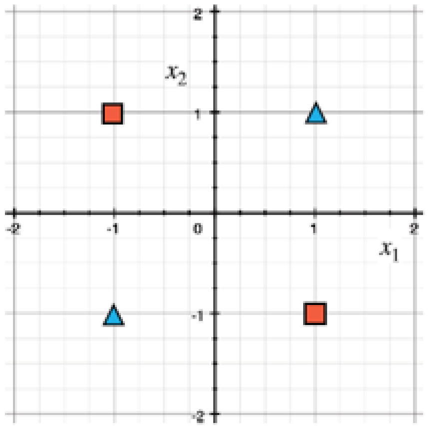
Agenda

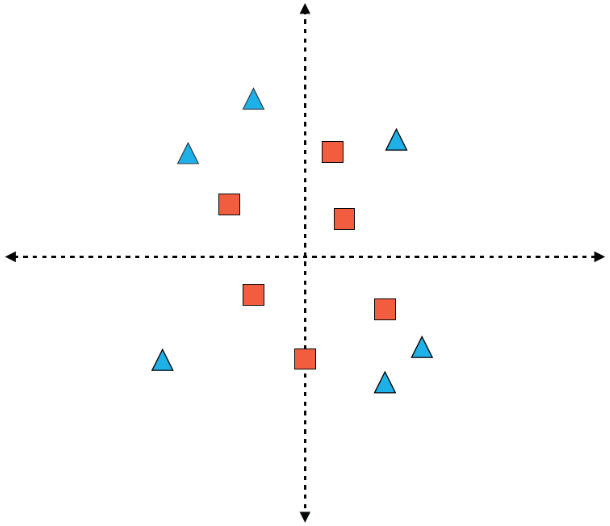
- Nonlinear classification
- Polynomial features
- Kernel Trick

Intro

- Support vector Classifier is pretty cool, because
  - Handle outliers
  - Allow overlapping classification — because it allows misclassification
- But, what if we had **tons of overlap**?

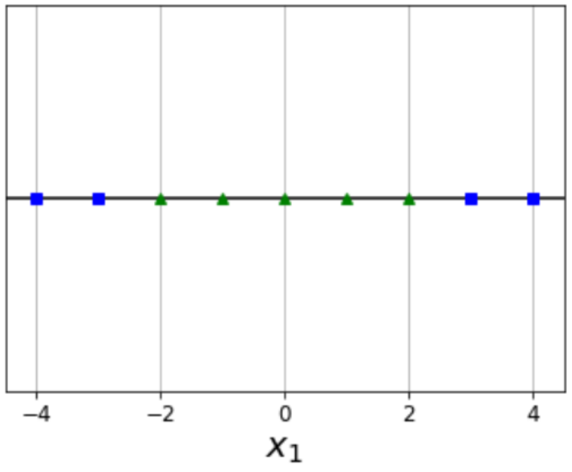


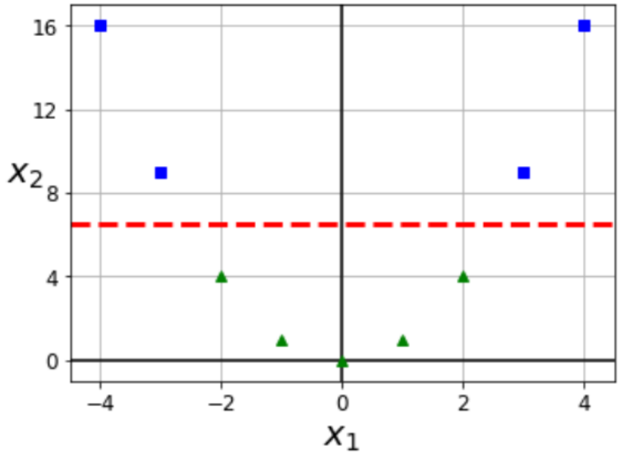




Polynomial features

- If we have a dataset and find that we can't separate it with a linear classifier, what can we do?
- One idea is to **add more columns** to this dataset and hope that the richer dataset is linearly separable.





Polynomial features

- Adding polynomial features is simple to implement and can work great with all sorts of Machine Learning algorithms (not just SVMs), but …
  - At a low polynomial degree, it cannot deal with very complex datasets
  - With a high polynomial degree it creates a huge number of features, making the model too slow

Kernel Trick

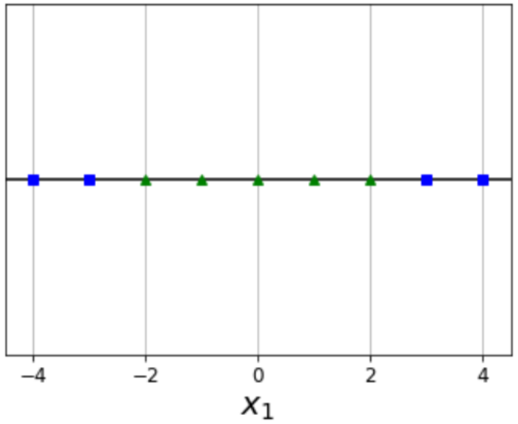
- Fortunately, when using SVMs you can apply an almost miraculous mathematical technique called the **kernel trick**.
- SVM uses something called Kernel function to systematically find support vector classifier in **higher dimensions**
- It makes it possible to get the **same result** as if you added many **polynomial features**, even with very high-degree polynomials, **without** actually having to **add** them
- So there is **no combinatorial explosion** of the number of features since you don't actually add any features

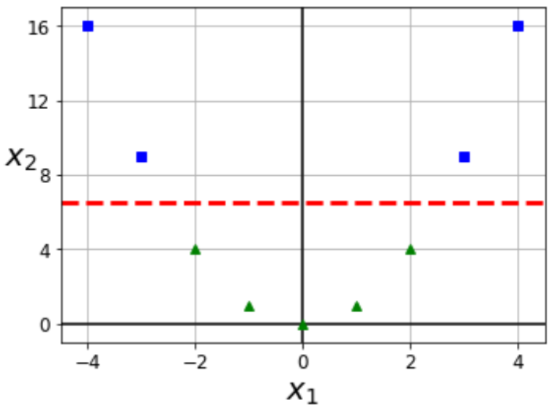
Kernel Trick

- The kernel method consists of **adding more columns** in a **clever way**, building a linear classifier on this new dataset and later removing the columns we added while keeping track of the (now nonlinear) classifier.
- The kernel trick allows the SVM to operate in a **higher-dimensional space** without explicitly computing the coordinates in that space. This makes it computationally **efficient** while still benefiting from the complex transformation.

Kernel Trick

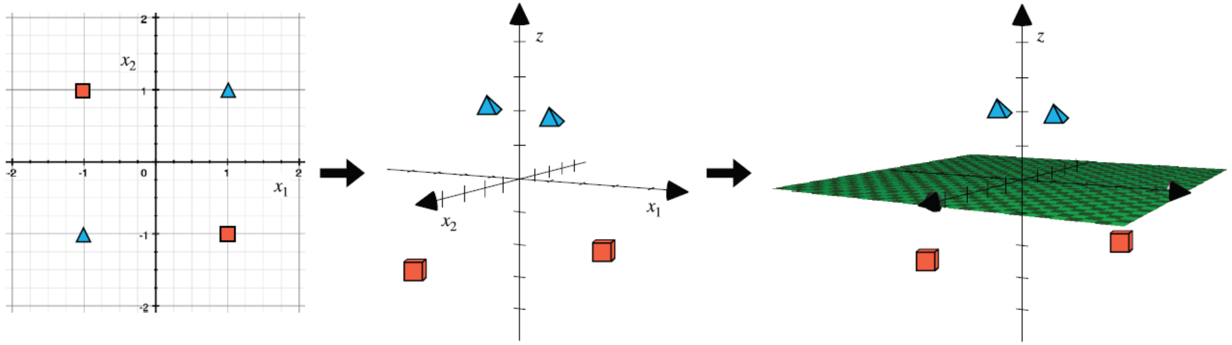
- The kernel method consists of adding more columns in a clever way, building a linear classifier on this new dataset and later removing the columns we added while keeping track of the (now nonlinear) classifier.





Kernel Trick

- Like if the points on your paper all of a sudden start flying into space at different heights. Maybe if we raise the points at different heights in a clever way, we can separate them with a plane. This is the kernel method.



Not separable by a line — Bring triangles up. Bring squares down. — Now separable by a plane

Kernel Trick

- A kernel is a similarity function, which, in short, is a function that tells us if two points are similar or different (e.g., close or far).
- A kernel can give rise to a feature map, which is a map between the space where our dataset lives and a (usually) higher-dimensional space.
- We look at the kernel method as a way of adding columns to our dataset to make the points separable

Kernel Trick

- The two kernels and their corresponding features maps we see in this chapter are:
  - Polynomial kernel and
  - Radial basis function (RBF) kernel.
- Both of them consist of adding columns to our dataset in different, yet very effective, ways.
- Next Lecture