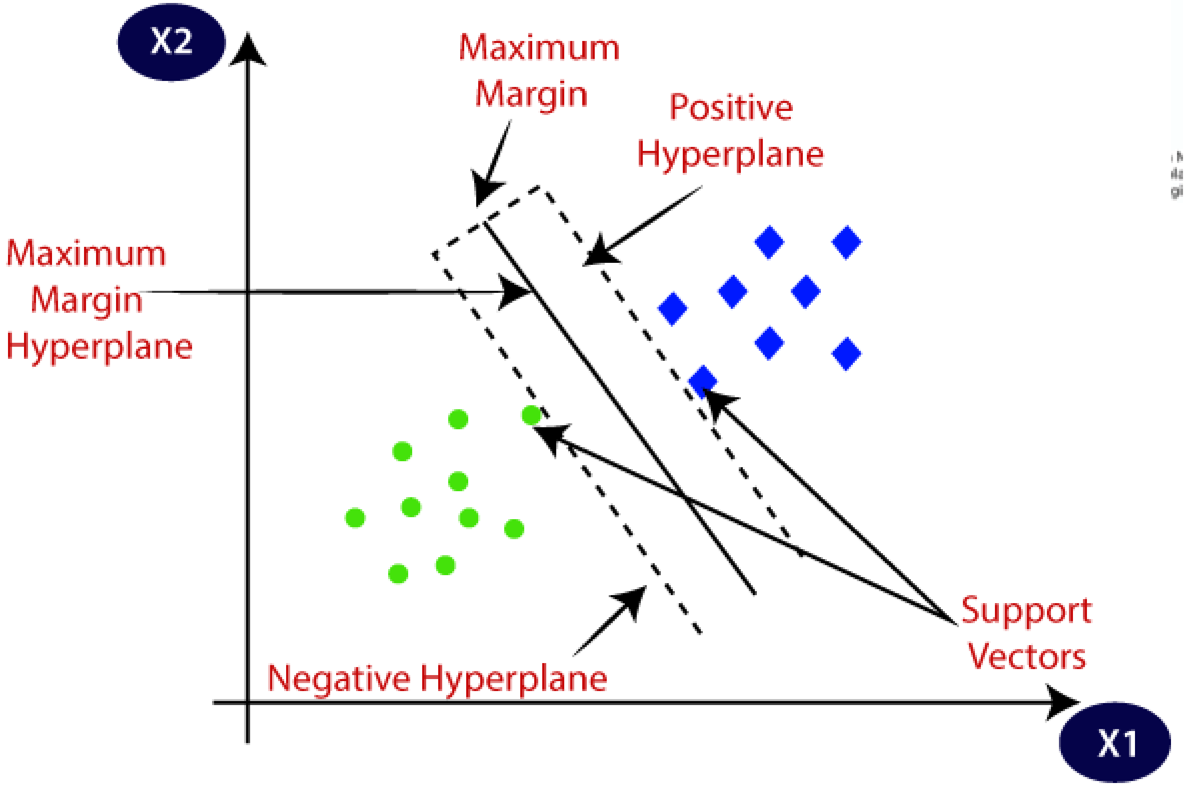
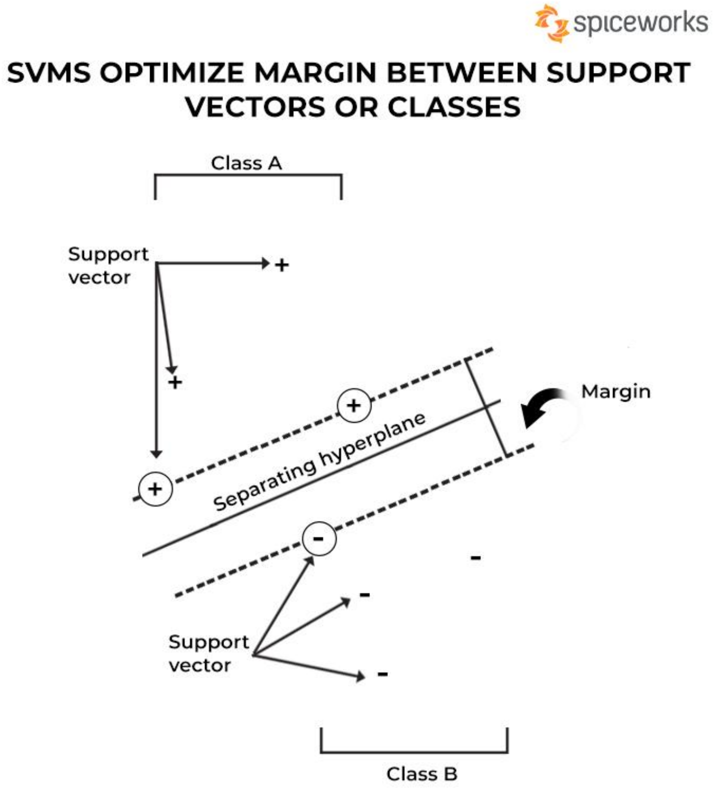
AI Algorithms – 2: SVF, RBF

- **Sayed Ahmed**
- PhD Studies in Electrical and Computer Eng. (McMaster University) (Partially Complete)
- Master of Engineering in Electrical and Computer Engineering (McMaster University)
- MSc in Data Science and Analytics (Toronto Metropolitan University/Ryerson)
- MSc in Computer Science (U of Manitoba)
- BSc. Engineering in Computer Science and Engineering (BUET)
- Extensive experience in Software Development and Engineering (primarily in Canada)
- Significant experience in Teaching
- Taught in Universities, Colleges, and Training Institutes

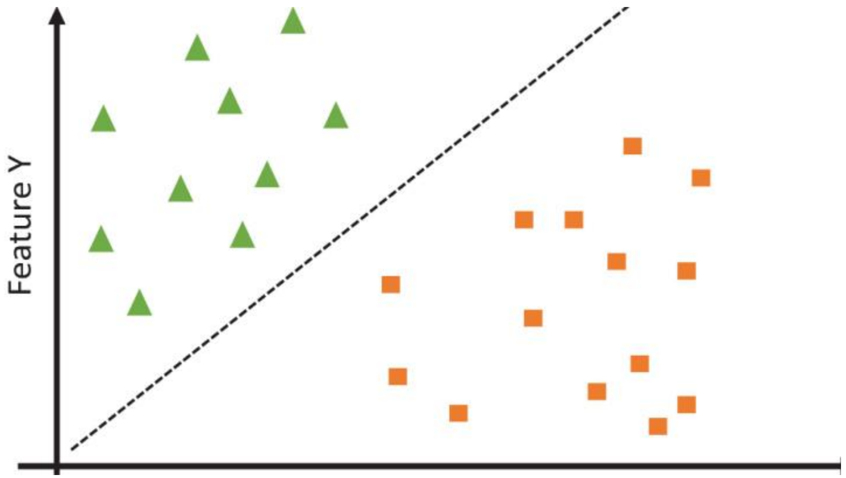
SVM

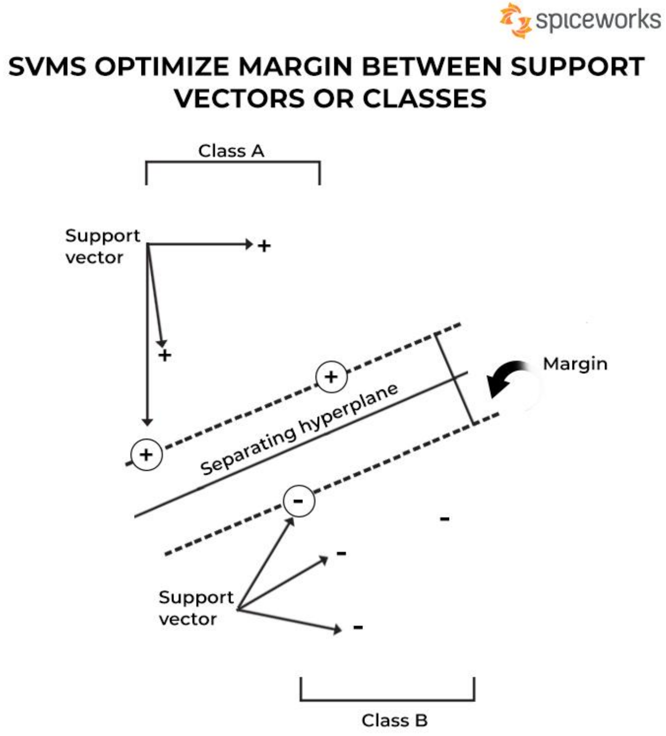


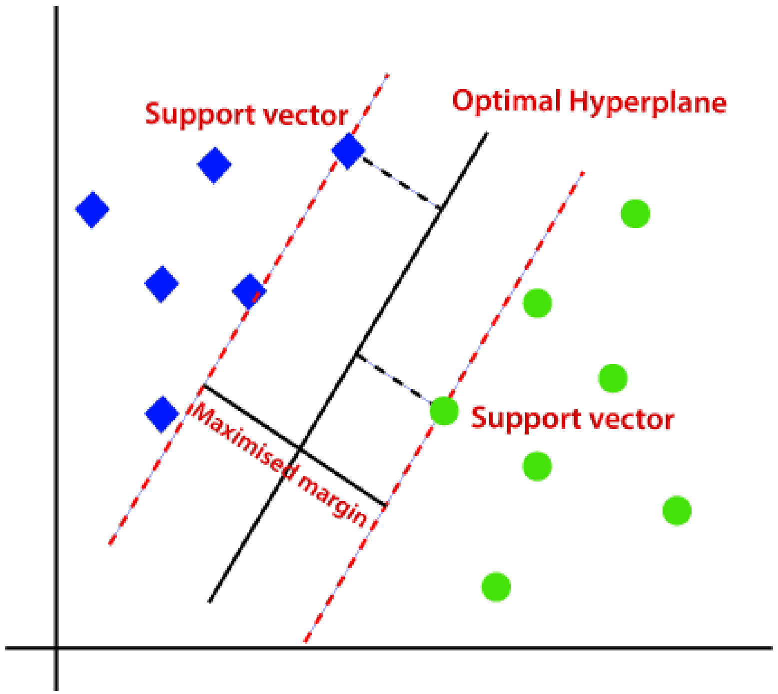


- I know it did not help without explanation. Ref: Google Images

SVM







- https://www.spiceworks.com/tech/big-data/articles/what-is-support-vector-machine/ , https://www.spiceworks.com/tech/big-data/articles/what-is-support-vector-machine/

https://editor.analyticsvidhya.com/uploads/729834.png

SVM?

- **SVM** or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and **work** well for many practical problems. The idea of **SVM** is simple: The algorithm creates a line or a hyperplane which separates the data into classes.

- https://www.slideshare.net/rajshreemuthiah/support-vector-machine-and-associative-classification

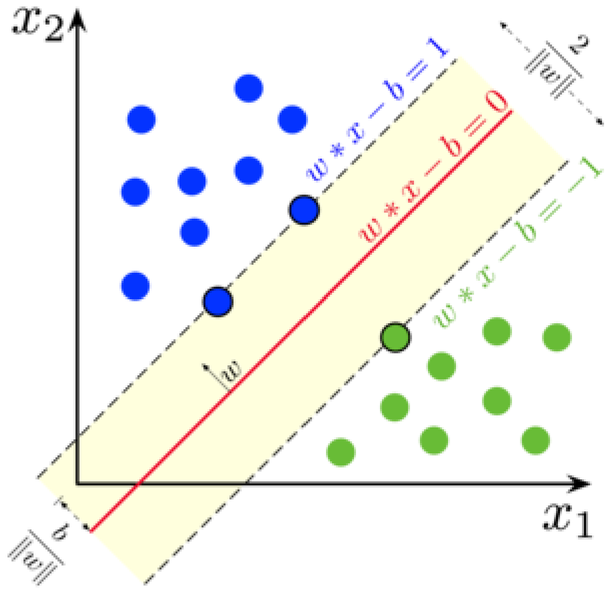
SVM: Linear or Non-Linear or both?

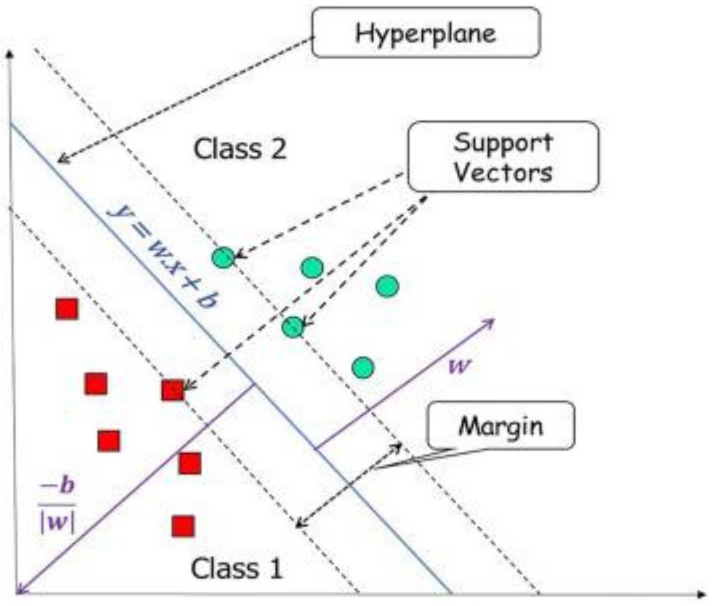
SVM—Support Vector Machines

- A new classification method for both linear and nonlinear data
- It uses a nonlinear mapping to transform the original training data into a higher dimension
- With the new dimension, it searches for the linear optimal separating hyperplane (i.e., “decision boundary”)
- With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane
- SVM finds this hyperplane using support vectors (“essential” training tuples) and margins (defined by the support vectors)

- It’s a bit advanced. Classify with a line or a non-linear line
- Find higher dimensional relationship in the data using a function called Kerber tricks such as RBF

SVM





- https://ars.els-cdn.com/content/image/3-s2.0-B978032385214200001X-f06-02-9780323852142.jpg https://upload.wikimedia.org/wikipedia/commons/thumb/7/72/SVM_margin.png/300px-SVM_margin.png

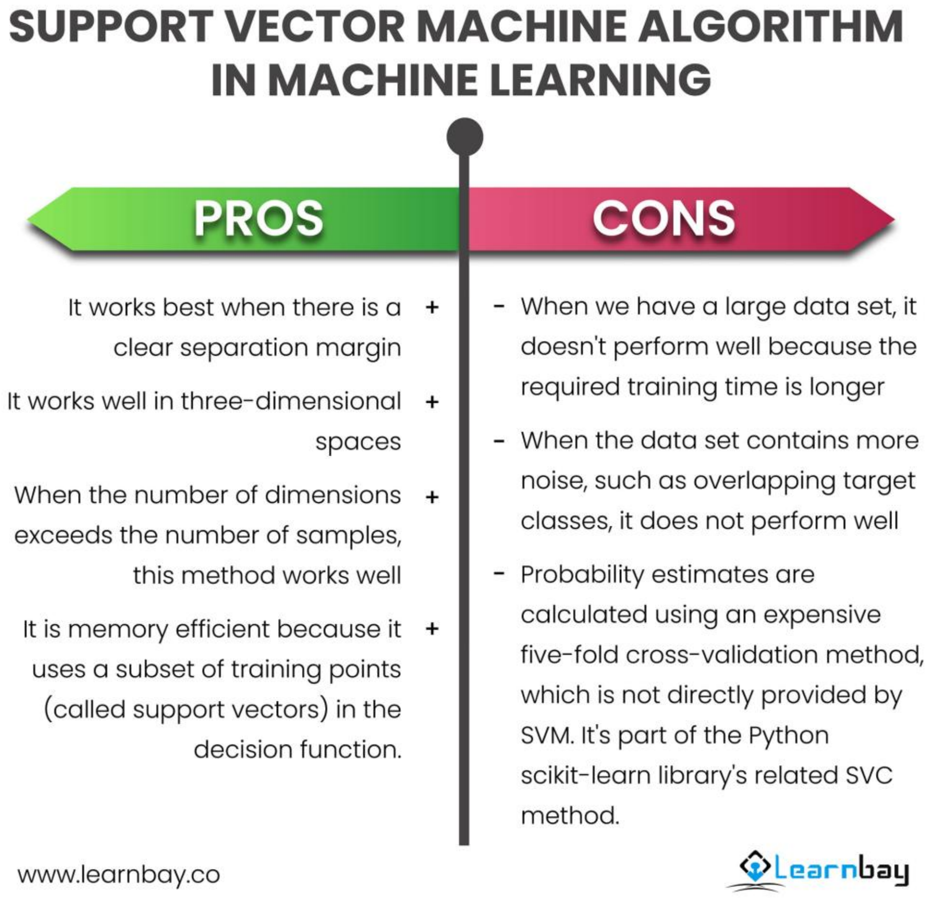
SVM Characteristics

- **SVM**
  - **Classification**
    - Usually only 2 classes
  - **Real valued features** (no categorical ones)
  - **Tens/hundreds of thousands of features**
  - Very **sparse features**
  - **Simple decision boundary**
    - No issues with overfitting
- Example applications
  - Text classification
  - Spam detection
  - Computer vision

- **Decision trees**
  - **Classification & Regression**
    - Multiple (~10) classes
  - **Real valued and categorical features**
  - **Few (hundreds) of features**
  - Usually **dense features**
  - **Complicated decision boundaries**
    - Overfitting! Early stopping
- Example applications
  - User profile classification
  - Landing page bounce prediction

- https://www.researchgate.net/profile/Ankit-Desai-9/post/Why-Support-Vector-MachineSVM-Best-Classifier/attachment/59d6254979197b80779839e2/AS%3A317089980583945%401452611550570/download/_20160112_203156.JPG

SVM Characteristics



Support Vector Machines (SVM) [15] [16] [17]

Advantages	Disadvantages
• Gives good results even if there is not enough information about the data. Also works well with unstructured data. • Solves complex problems with a convenient kernel solution function. • Relatively good scaling of high-dimensional data.	• It is difficult to choose the appropriate kernel solution function. • Training time is long when using large data sets. • It may be difficult to interpret and understand because of problems caused by personal factors and the weights of variables. • The weights of the variables are not constant, thus the contribution of each variable to the output is variant.

- https://www.facebook.com/photo/?fbid=3082863228696212&set=a.1871025246546689, https://www.researchgate.net/publication/338950098/figure/tbl2/AS:10066704

SVM utilizes kernel functions to map the input data points into a higher-dimensional space where the separation between the two classes becomes easier. This allows SVM to solve complex non-linear problems as well. Apr 29, 2023

Medium
https://medium.com › what-is-kernel-trick-in-svm-inter...

What is Kernel Trick in SVM ? Interview questions related to ...

https://medium.com/@Suraj_Yadav/what-is-kernel-trick-in-svm-interview-questions-related-to-kernel-trick-97674401c48d

If AI enables computers to think, computer vision enables them to see, observe and understand. Computer vision works much the same as human vision, except humans have a head start.

IBM
https://www.ibm.com › topics › computer-vision

What is Computer Vision? - IBM

About featured snippets • Feedback

https://www.ibm.com/topics/computer-vision

SVM: Pros and Cons

SVMs: Pros and cons

- Pros
  - Kernel-based framework is very powerful, flexible
  - Often a sparse set of support vectors – compact at test time
  - Work very well in practice, even with very small training sample sizes
  - Solution can be formulated as a quadratic program (next time)
  - Many publicly available SVM packages: e.g. LIBSVM, LIBLINEAR, SVMLight
- Cons
  - Can be tricky to select best kernel function for a problem
  - Computation, memory
    - At training time, must compute kernel values for all example pairs
    - Learning can take a very long time for large-scale problems

Adapted from Lana Lazebnik

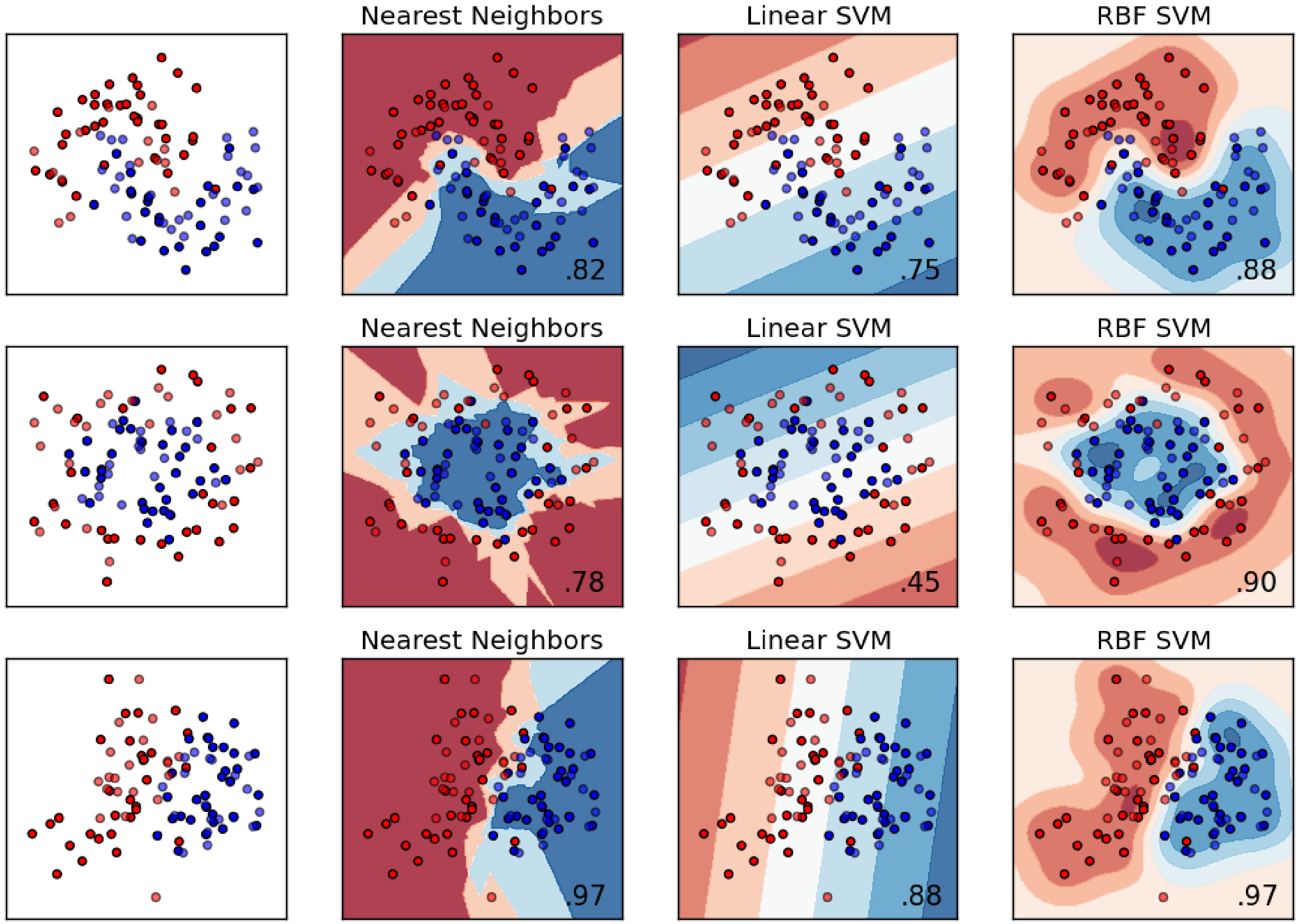
- https://slideplayer.com/slide/14832548/90/images/39/SVMs%3A+Pros+and+cons+Pros+Cons.jpg

SVM, Decision Trees, Naïve Bayes

	Algorithms		
	SVM	**Decision Tree**	**Naïve Bayes**
Advantages	Effective in high dimensional spaces	Computational complexity is not high	Still valid when dealing with small sample size
	Suitable when the sample size is smaller than the number of dimensions	Output is easy to understand and to interpret (i.e., output tree can be visualized).	Easy to extend to multi-class classification problems
	It offers various Kernel functions for non-linear decision boundaries	Can handle numerical and categorical data	Fast, efficient and easy to implement
Disadvantages	When the number of features larger than the number of samples, it is crucial to choose suitable Kernel function and regularization	Propensity to overfit the data	Sensitive to preprocessing of data input
	Complex calculation when there are many class labels	Can be unstable because small variations in data might lead to different results	Only for categorized data
	Not usually employed for continuous numerical variables, mostly for categorical variables	May generate biased tree if some classes are dominant	

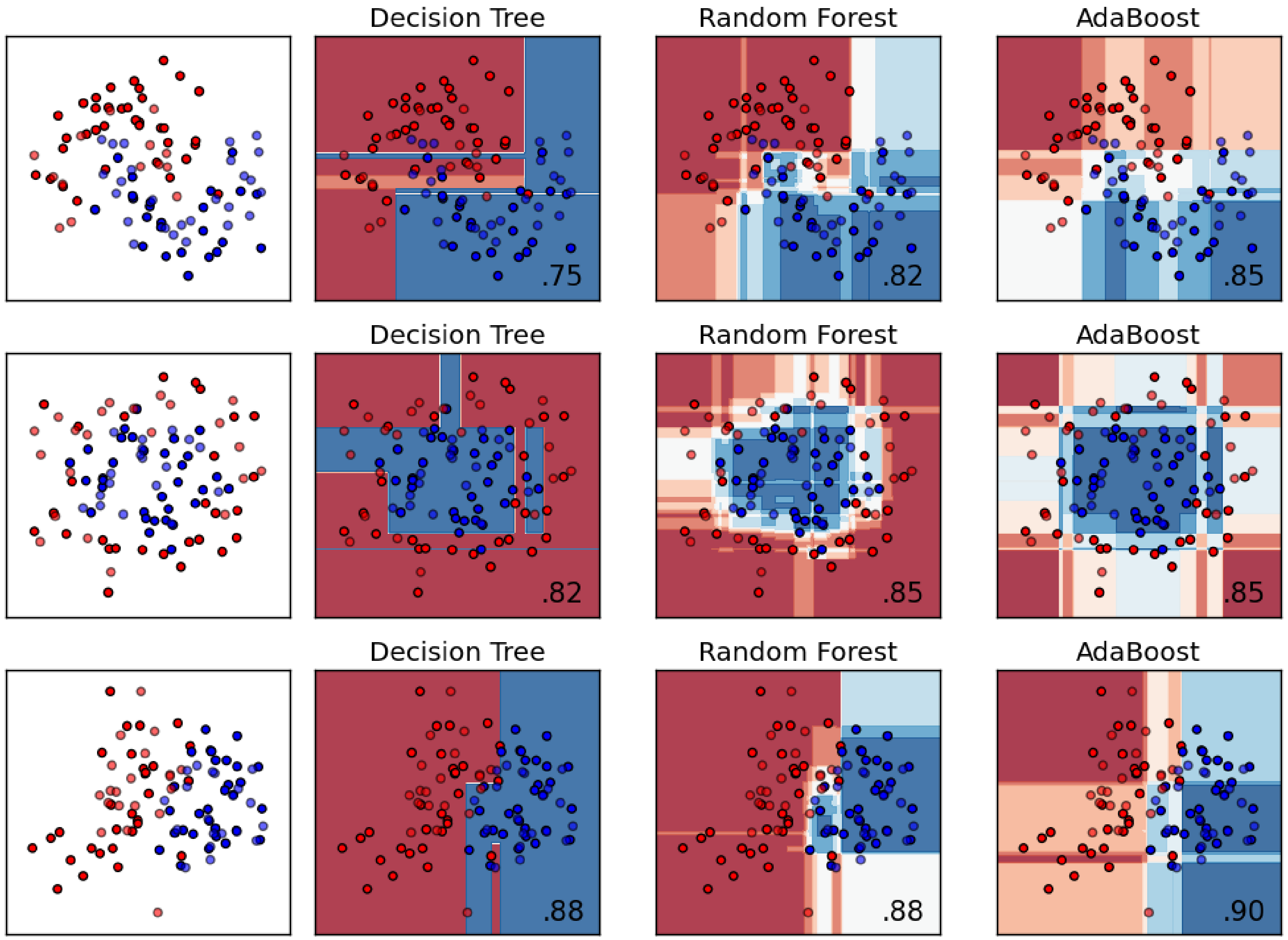
- https://www.researchgate.net/publication/345935933/figure/tbl2/AS:958440910372866@1605521528952/Comparison-of-the-three-Machine-Learning-Algorithms-employed-Support-Vector-Machine.png

Compare: Linear vs RBF SVM



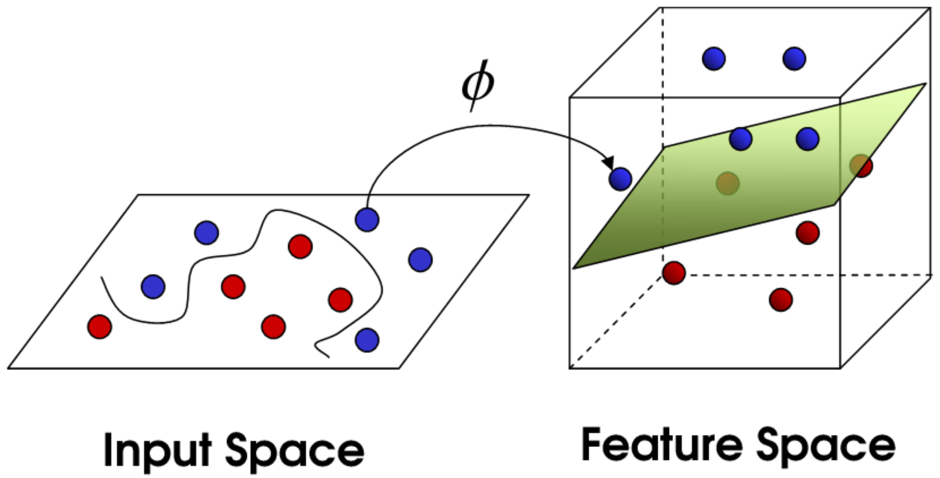
- https://martin-thoma.com/images/2016/01/ml-classifiers-1.png

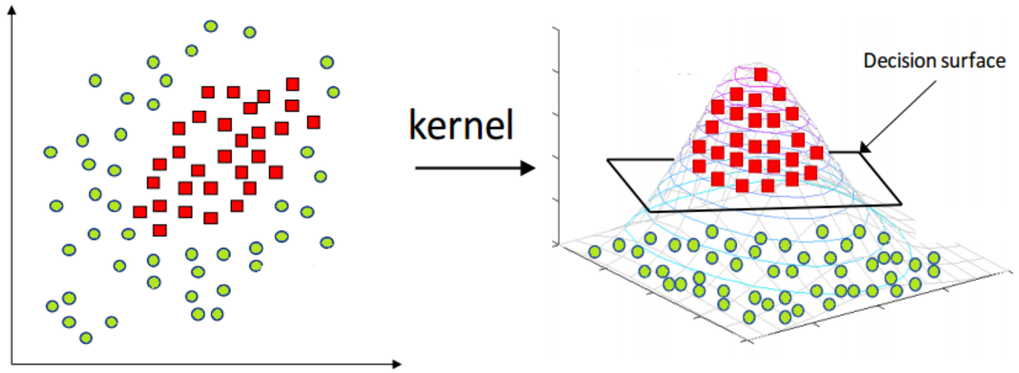
Classification Comparisons

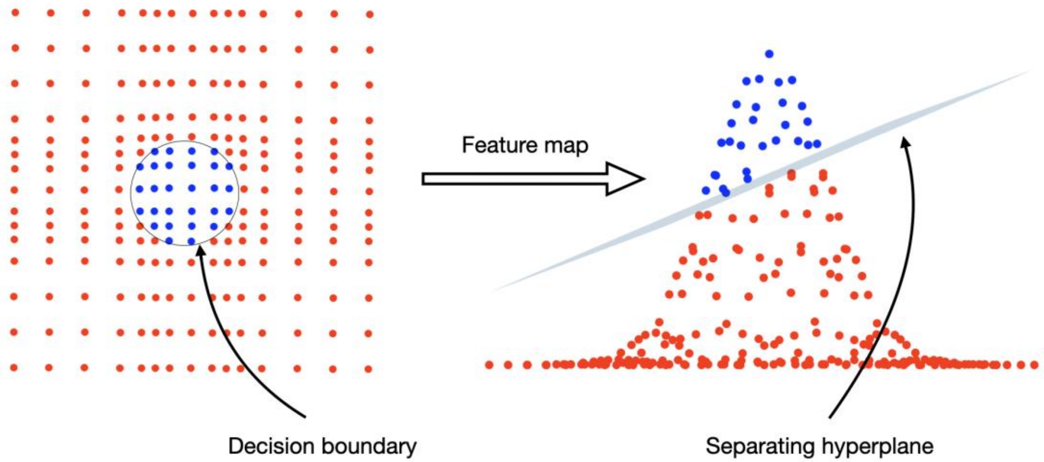


- https://martin-thoma.com/images/2016/01/ml-classifiers-2.png

SVM and Kernel Tricks

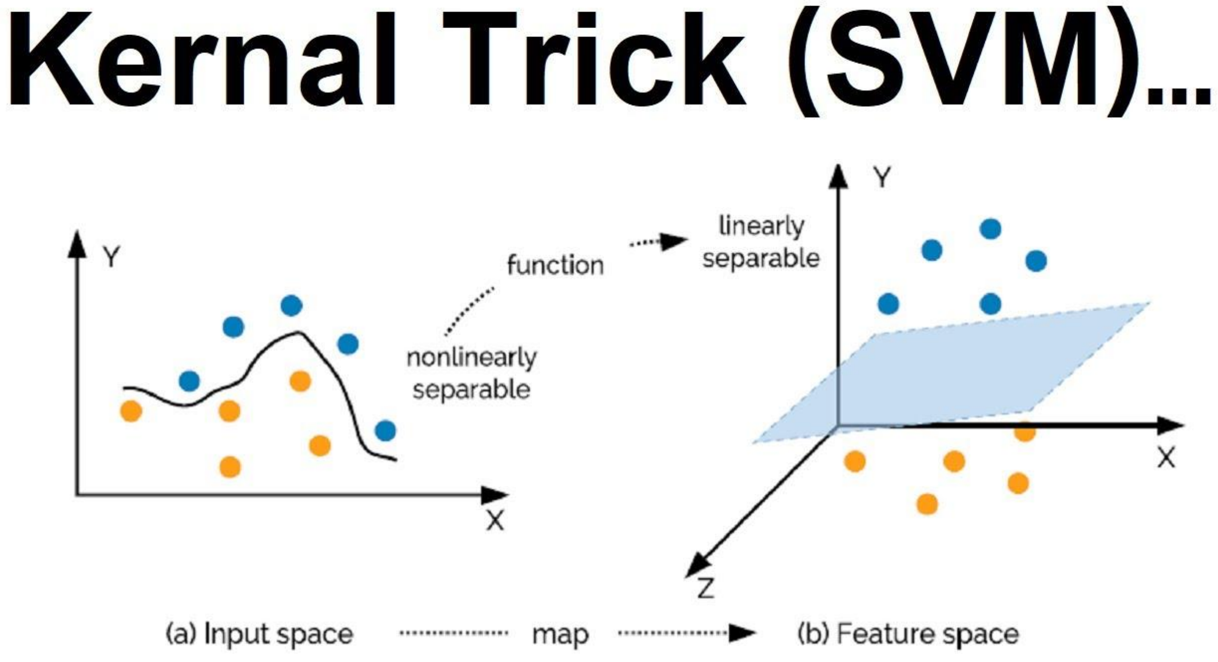






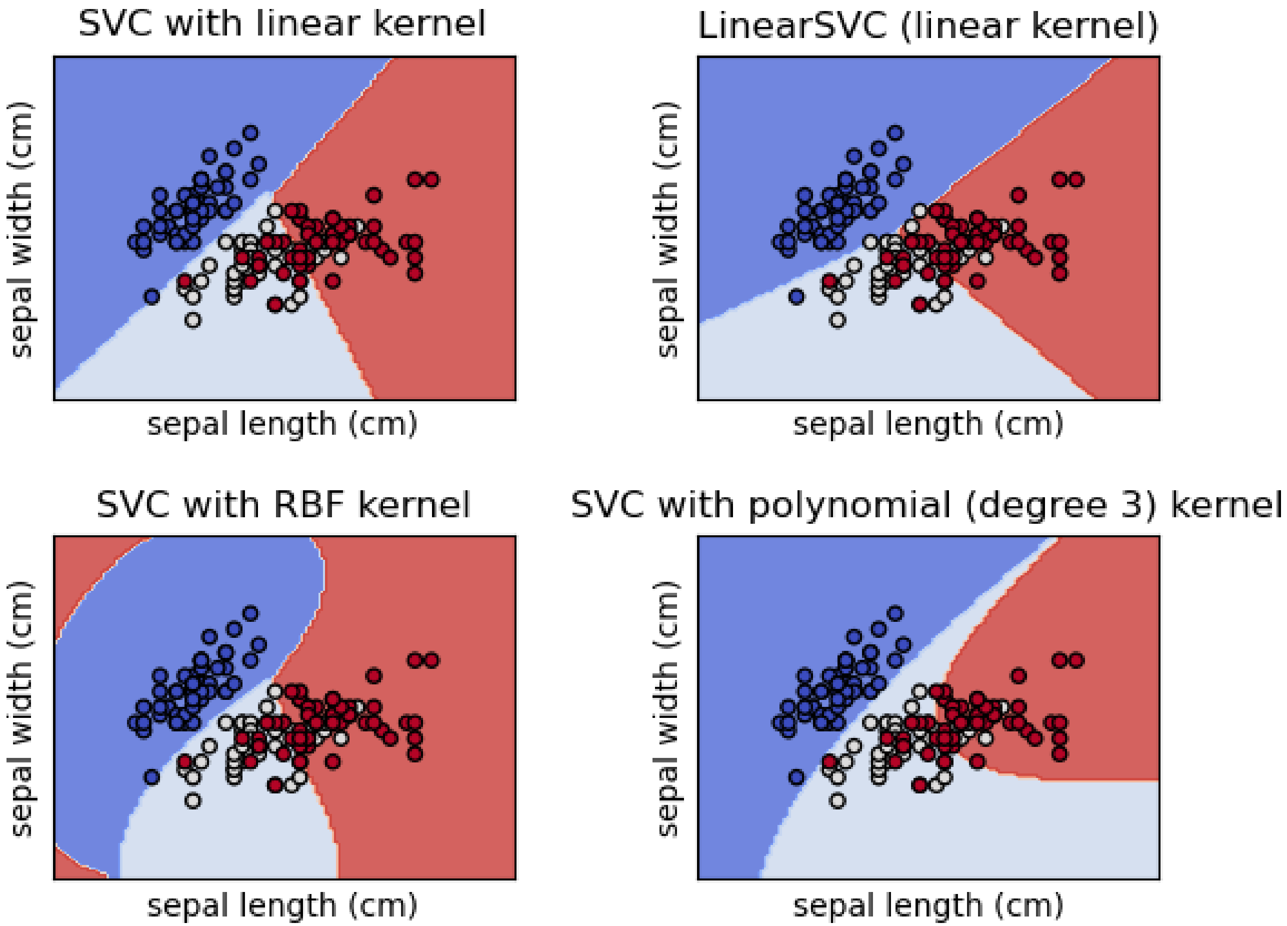
- https://miro.medium.com/v2/resize:fit:872/1*zWzeMGyCc7KvGD9X8lwlnQ.png

Kernel Tricks in SVM



- https://i.ytimg.com/vi/wqSTBCguVyU/maxresdefault.jpg

SVM with RBF Kernels



- https://scikit-learn.org/stable/_images/sphx_glr_plot_iris_svc_001.png

SVM with Gaussian Kernels

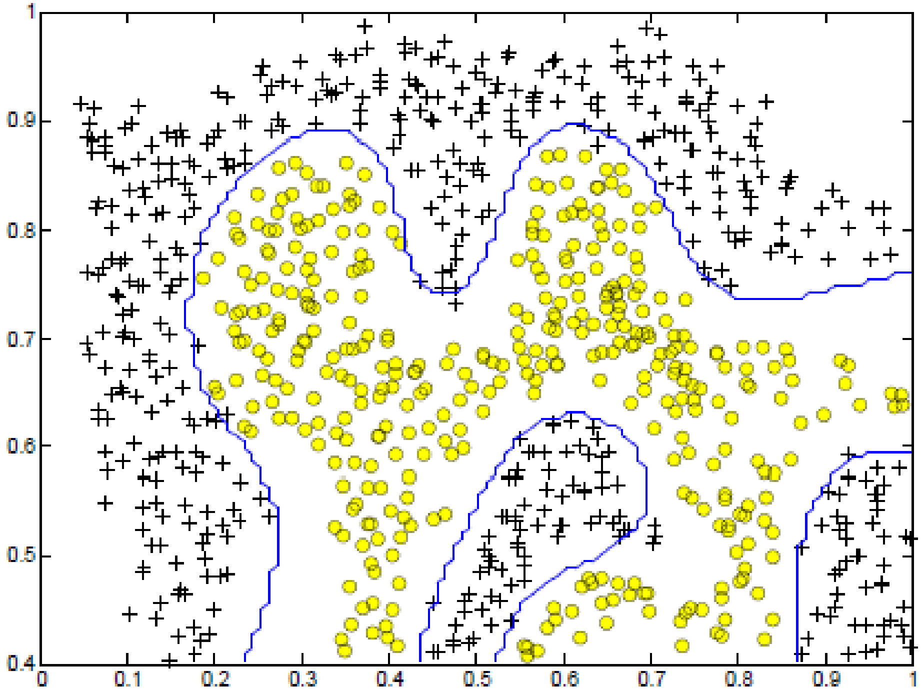


Figure 5: SVM (Gaussian Kernel) Decision Boundary (Example Dataset 2)

- https://i.stack.imgur.com/58LGs.png

Math: Kernel Functions

Kernel	Kernel function, $k(x, y)$	Partial derivative, $\frac{\partial k(x,y)}{\partial x^j}$
Linear	$\boldsymbol{x}^\top \boldsymbol{y}$	y^j
Poly	$(\gamma \boldsymbol{x}^\top \boldsymbol{y} + c_0)^p$	$\gamma p y^j (\gamma \boldsymbol{x}^\top \boldsymbol{y} + c_0)^{p-1}$
RBF	$\exp(-\gamma \lVert \boldsymbol{x} - \boldsymbol{y} \rVert^2)$	$-2\gamma(x^j - y^j)k(\boldsymbol{x},\boldsymbol{y})$
Tanh	$\tanh(\gamma \boldsymbol{x}^\top \boldsymbol{y} + c_0)$	$\gamma y^j \operatorname{sech}^2(\gamma \boldsymbol{x}^\top \boldsymbol{y} + c_0)$
ARD	$v^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{x^d - y^d}{\lambda_d}\right)^2\right)$	$\left(\frac{x^j - y^j}{\lambda_j^2}\right) k(\boldsymbol{x}, \boldsymbol{y})$

https://doi.org/10.1371/journal.pone.0235885.t001

- https://journals.plos.org/plosone/article/figure/image?size=medium&id=10.1371/journal.pone.0235885.t001

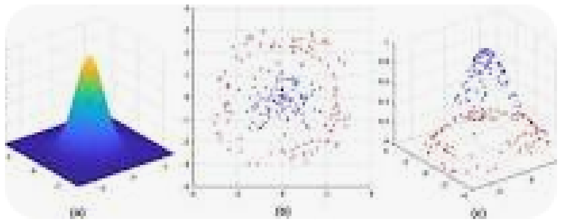
Why Gaussian Kernel in SVM

Why use Gaussian kernel in SVM?

SVMs with kernel functions are created for nonlinearly separable data. These kernel functions are basically polynomial, Gaussian and sigmoid. The Gaussian kernel function allows the separation of nonlinearly separable data by mapping the input vector to Hilbert space.

Mar 5, 2022

springer.com
https://link.springer.com › article



Is the RBF kernel a Gaussian kernel?

Thus RBF kernel is also known as Gaussian Radial Basis Kernel. RBF kernel is most popularly used with K-Nearest Neighbors and Support Vector Machines.

- https://link.springer.com/article/10.1007/s13369-022-06654-3 , https://www.pycodemates.com/2022/10/the-rbf-kernel-in-svm-complete-guide.html

Is Hilbert space a vector space?

In direct analogy with n-dimensional Euclidean space, Hilbert space is a vector space that has a natural inner product, or dot product, providing a distance function. Under this distance function it becomes a complete metric space and, thus, is an example of what mathematicians call a complete inner product space. Jul 5, 2024

Britannica
https://www.britannica.com › science › Hilbert-space

Hilbert space | Linear operators, Banach spaces, Inner product

Which Kernel to use in SVM

How do I know which SVM kernel to use?

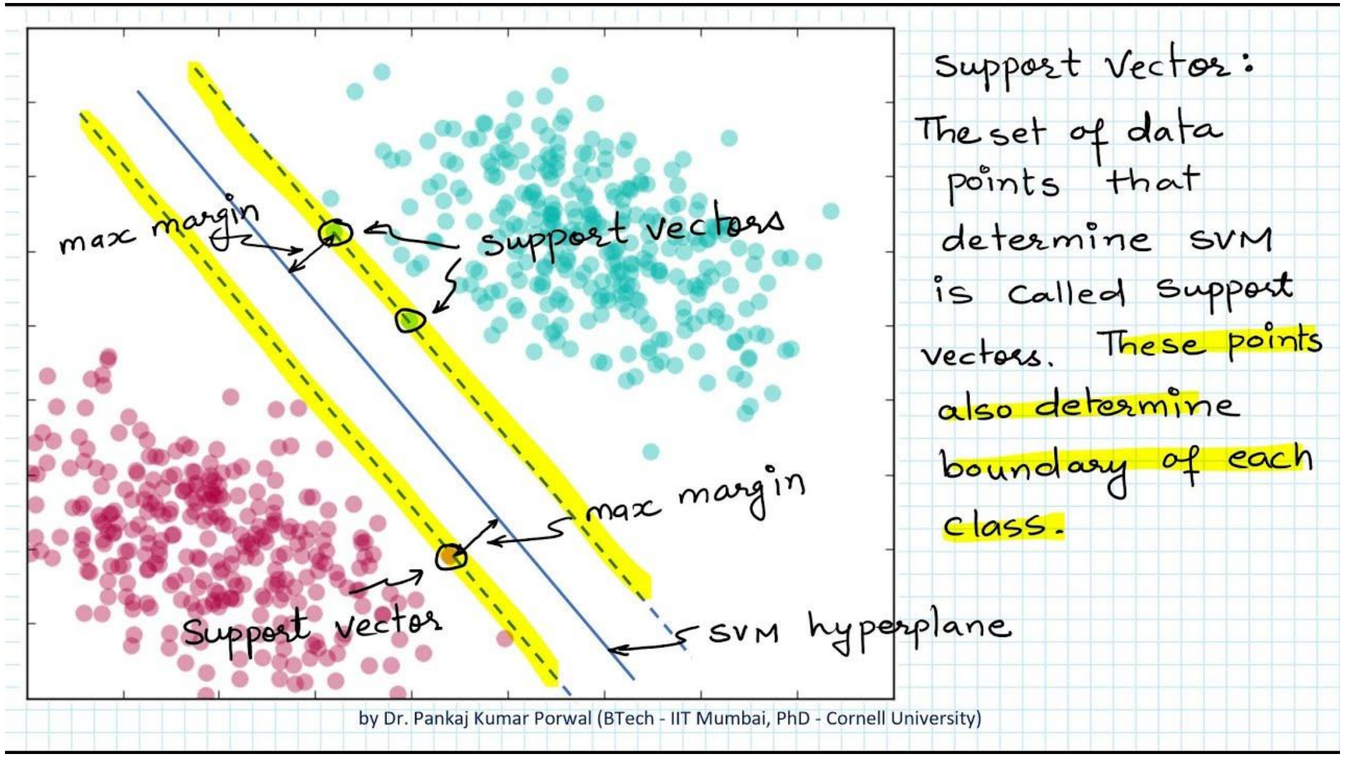
Generally, a linear kernel should be used if the data is linearly separable or has many features, a polynomial kernel if it has nonlinear patterns or interactions between features, an RBF kernel if it has complex and nonlinear patterns or clusters, and a sigmoid kernel if it is binary or looks like a logistic function.

linkedin.com
https://www.linkedin.com › advice › how-do-you-choos...

How do you choose the best kernel function for SVM in industrial ...

- https://www.linkedin.com/advice/0/how-do-you-choose-best-kernel-function-svm

Why use Maximum Margin?



- https://i.ytimg.com/vi/pq88UFYJ2PA/maxresdefault.jpg

Why use Maximum Margin?

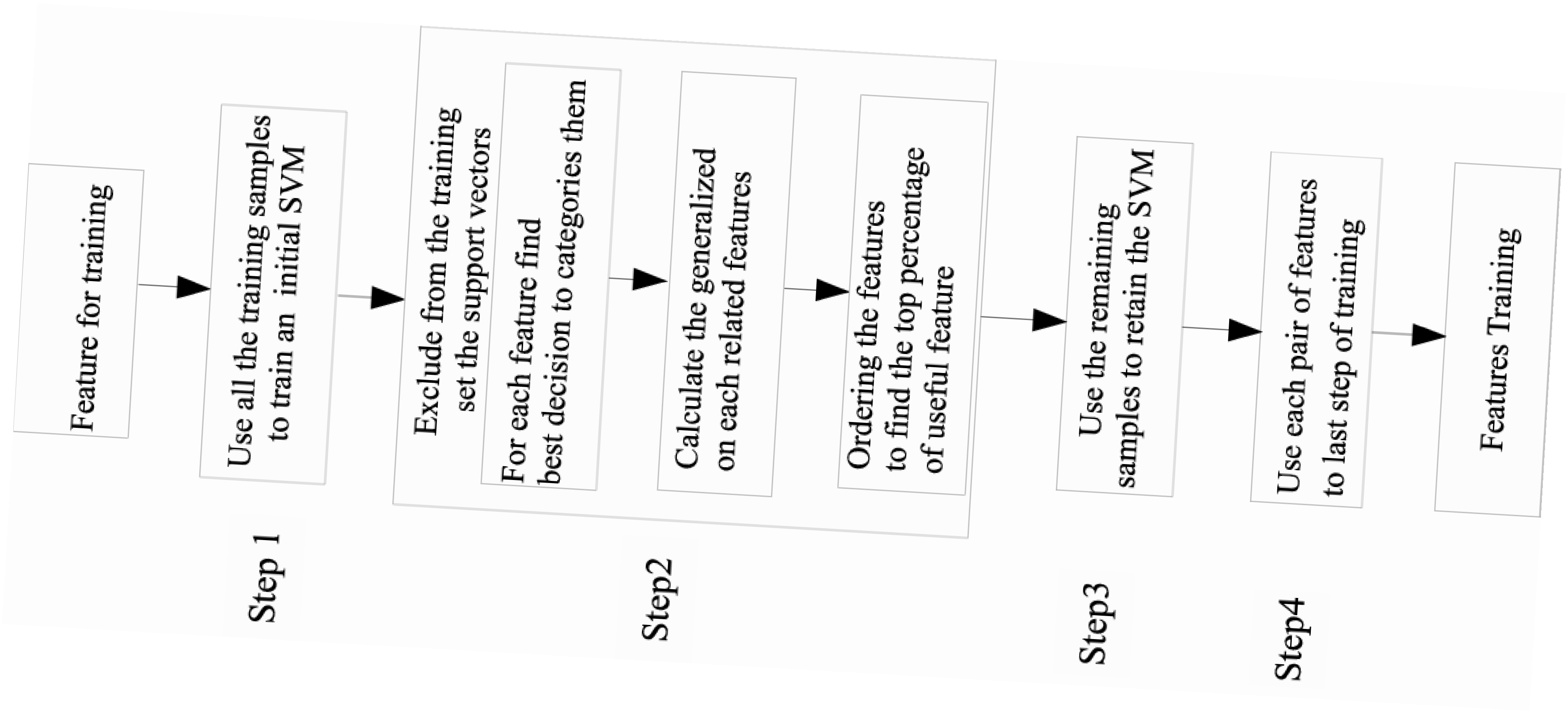
We introduced two reasons why SVM needs to find the maximum margin. First, a large margin can avoid the effect of random noise and reduce overfitting. Second, a larger margin will lead to a smaller VC dimension, reduce the number of potential classifiers, and, therefore, reduce the possibility of generalization error. Jan 6, 2022

Medium
https://medium.com › datascienceray › svm-why-maximi...

SVM: Why Maximize Margin. Machine Learning Interview Note

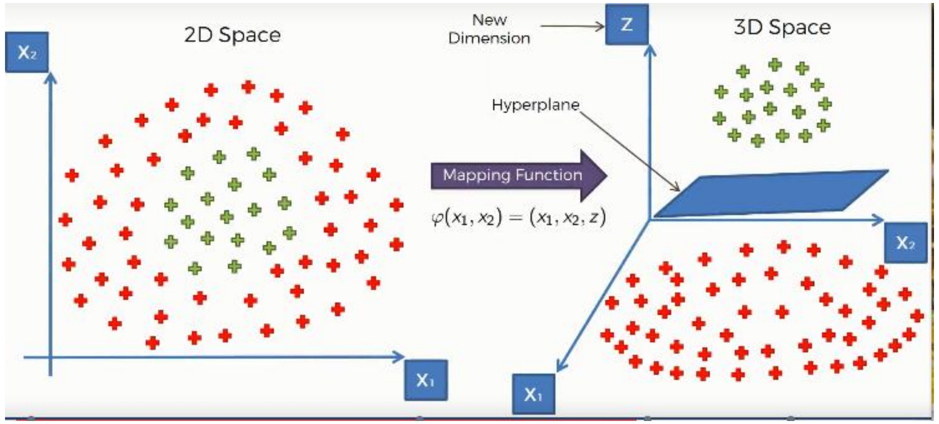
- https://medium.com/datascienceray/svm-why-maximize-margin-b3d69528daf1

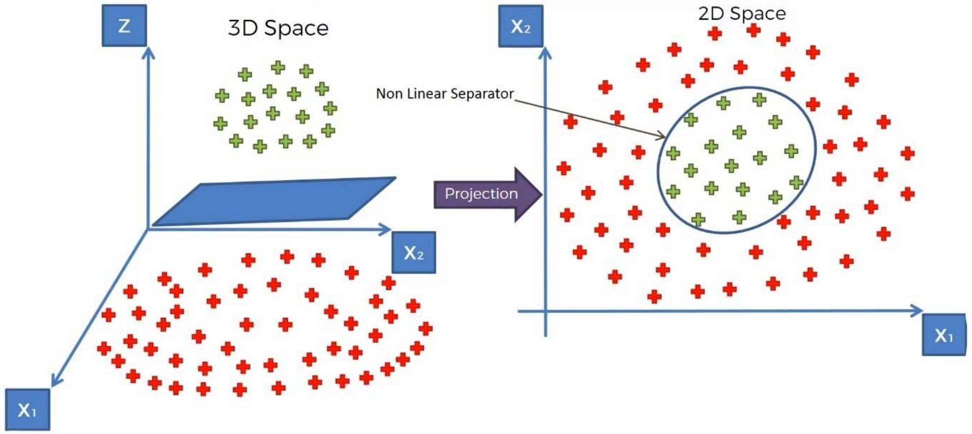
SVM Functionality and Implementation

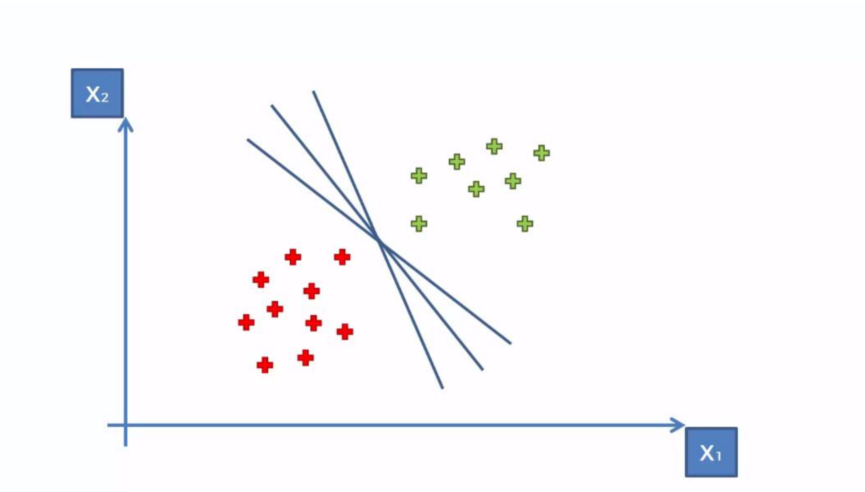


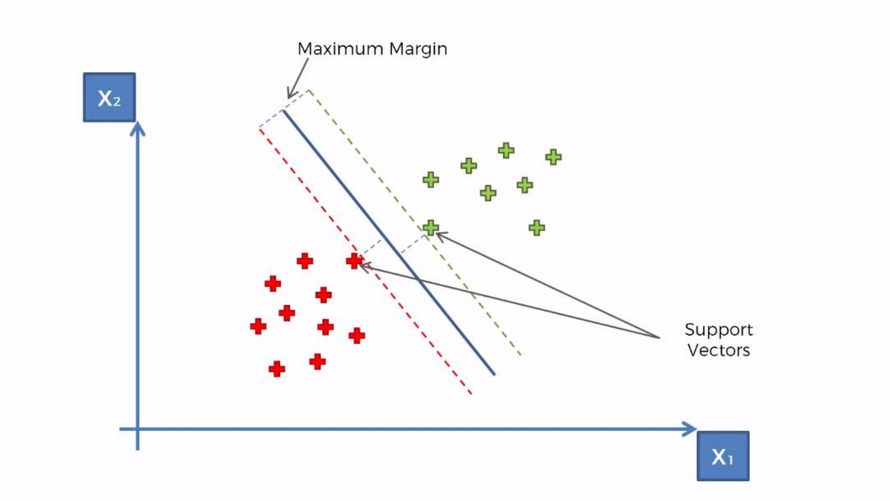
- https://d3i71xaburhd42.cloudfront.net/4293333436f22fab9ad8eba038b7f3422ba03bfd/2-Figure1-1.png

SVM Step by Step









- https://www.aionlinecourse.com/tutorial/machine-learning/support-vector-machine