



KNN & SVM



RESTORY...

Agenda

- K-nearest neighbours
- Support Vector machine
- SVM - användning
- SVM - Kernels

Kod

- SVM
- Kernel - SVM



KNN



RESTORY...

KNN

Klassificering.

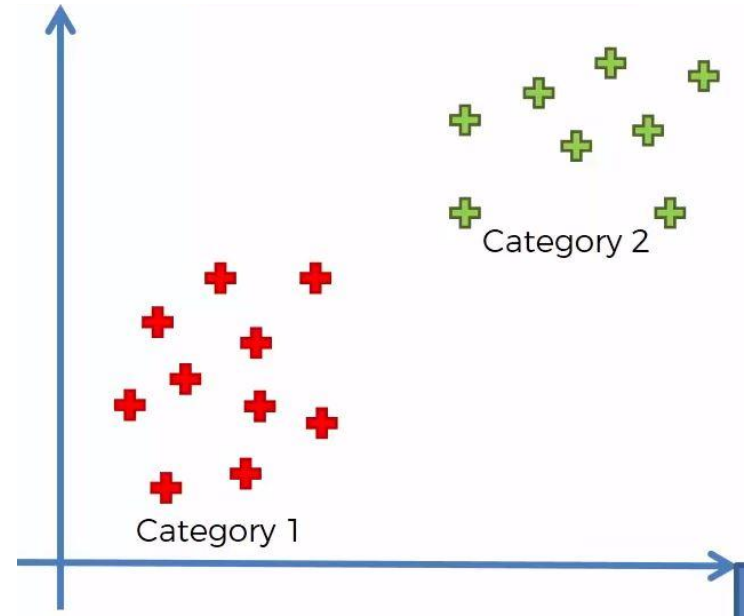
Närmaste grannar.

K betecknar antalet "grannar".

Vilken grupp tillhör en ny punkt?

Mäta avstånd till dess grannar!

Euklidiskt avstånd - mellan två punkter.

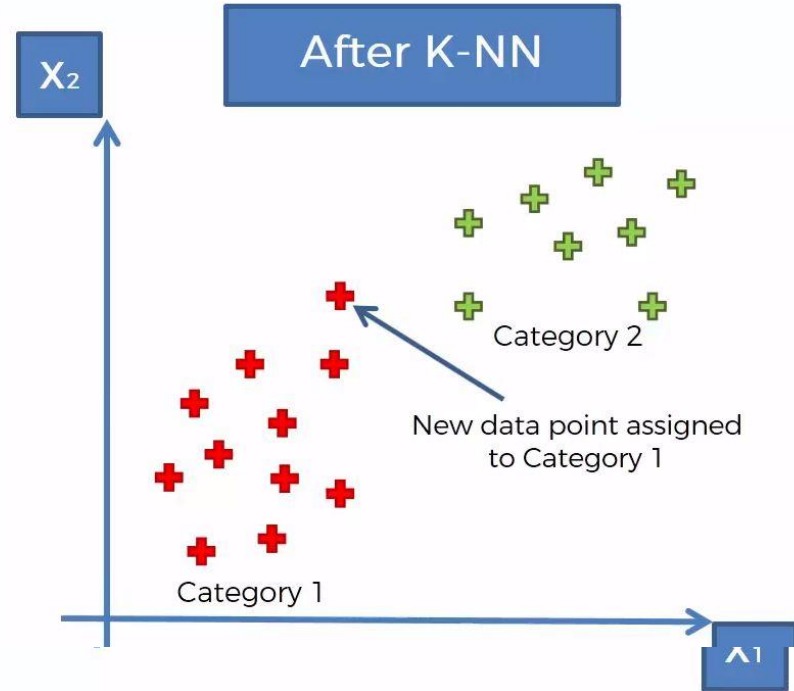


KNN

Lägg till en ny punkt –

Hur ska man klassificera den?

KNN algoritmen undersöker vilken kategori majoriteten av de K stycken närmaste grannarna tillhör.

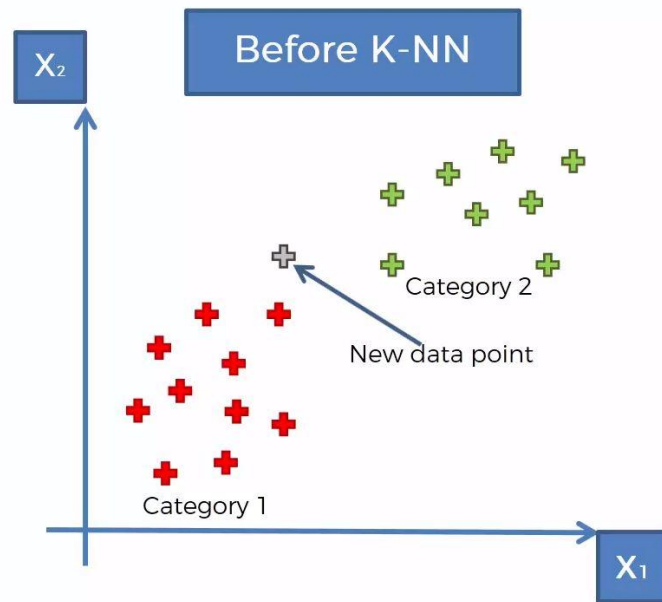


RESTRY...

KNN - Algoritmen

1. Välj antalet närmaste grannar (K) som ska användas.
2. Välj ut de K stycken **närmaste** grannarna till den nya datapunkten.
3. Bland dessa grannar, räkna på hur många tillhör varje kategori.
4. Tilldela den nya datapunkten den kategori som flest grannar tillhör.

KLART!



RESTORY...

KNN - Euklidiskt avstånd

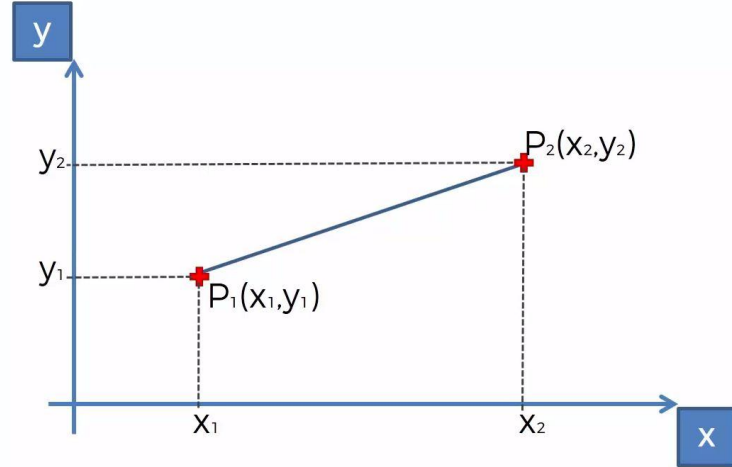
Skillnaden i avståndet mellan x punkterna.

Skillnaden i avståndet mellan y punkterna.

Upphöjt till två - ta bort negativa tecken.

Summera avstånden mellan dimensionerna.

Ta roten ur summan.



$$\text{Euclidean Distance between } P_1 \text{ and } P_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

KNN - Exempel



1. Välj ett antal närmaste grannar (nearest neighbours)
K. `n_neighbors=5`
2. Räkna ut avståndet till punkter och välj ut de 5 närmaste punkterna.
 - a. Välj avståndsformeln.
 - b. Euklidisk norm - mellan en och flera punkter.
 - c. Minkowski - $p=2$ samma som Euklidiskt avstånd.
 - d. [Det finns andra.](#)
3. Räkna hur många av de 5 tillhör varje kategori.
4. Tilldela den nya datapunkten den kategori som flest av de K närmaste grannarna tillhör.



Support Vector Machines

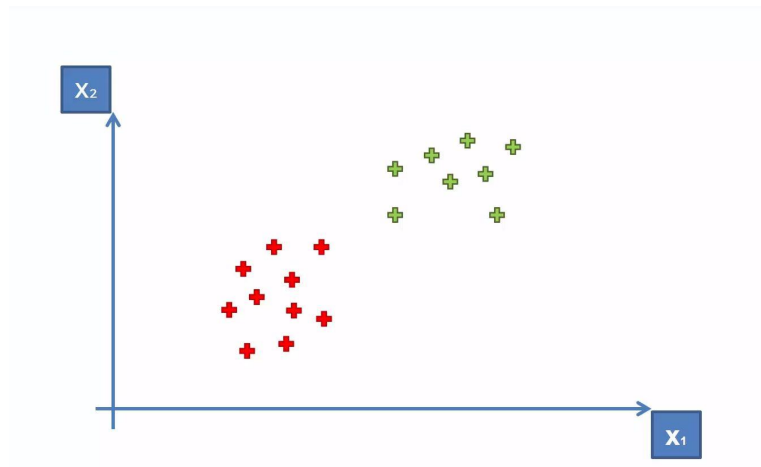


SVM

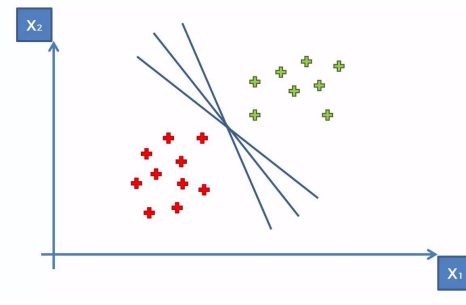
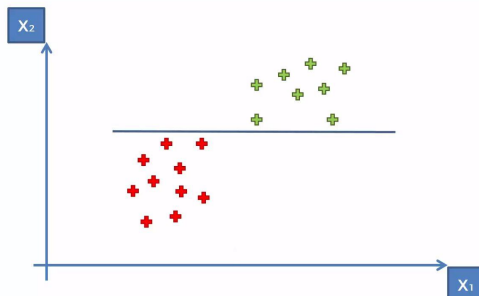
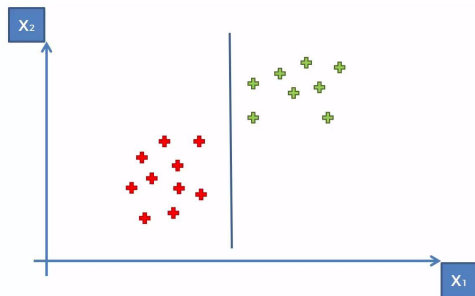
Support Vector Machine

Klassificering

Ny punkt - vilken kategori?



Kan dela upp klasserna på flera sätt.



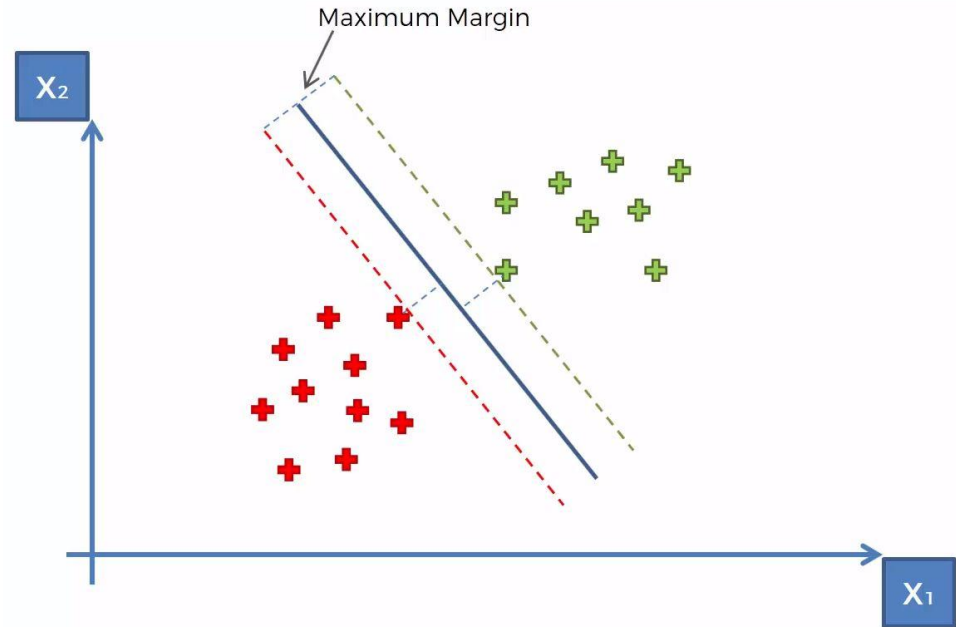
RESTORRY...

SVM - Maximum Margin

Marginal från strecket - tub

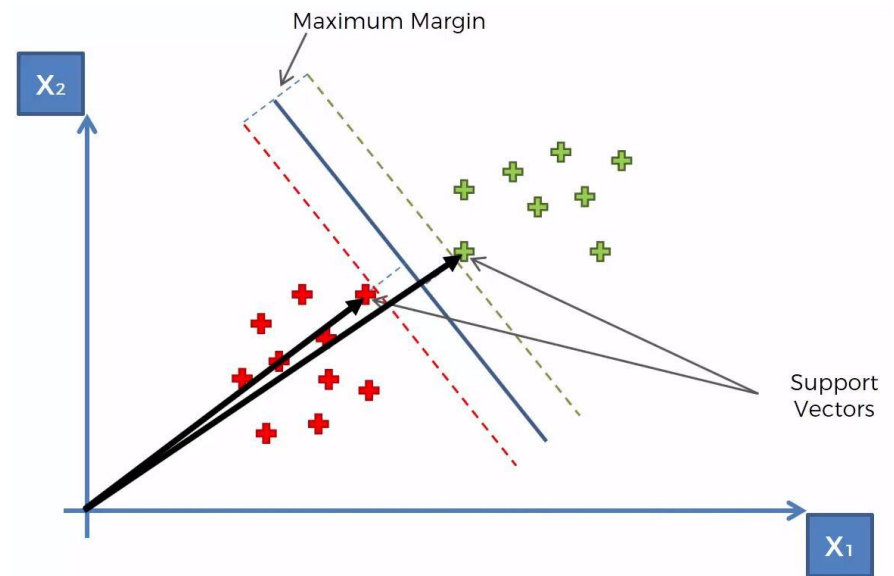
Lika avstånd på varsin sida.

Vill ha så mycket avstånd till de närmaste punkterna som möjligt - max marginal.



SVM - Vektor

- x - värden
 - flera dimensioner
 - en rad
 - en 1-D array
 - vektor
- Supporting the decision boundary



SVM - irl

- Känslor från ansikten
- Speech recognition
 - Separera ord från en ström av ord.
- Bildanalys
 - Har bilden blivit ändrad?
 - *"This could be used in security-based organizations to uncover secret messages. Yes, we can encrypt messages in high-resolution images."*
 - (handskriven)Text till (digital)text
 - Cancer detection
- m.m.



Happy



Sad



Surprised



Angry

SVM - på gränsen

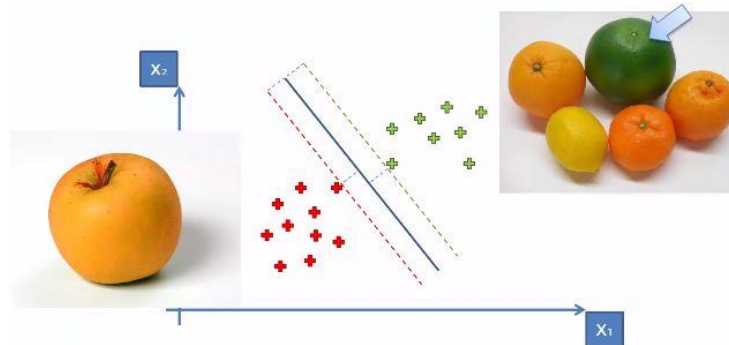
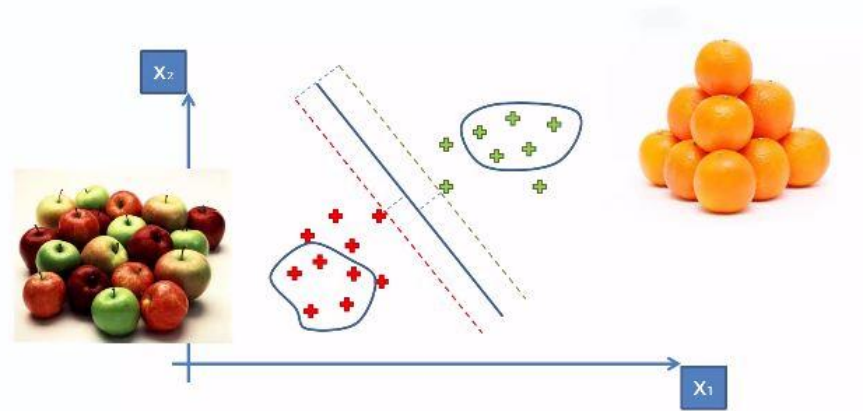
De närmaste punkterna är viktigast!

Många klassificeringsmetoder bygger på typiska värden.

SVM baseras på **gränsfallen**.

Äpplen som är mest lika apelsiner.

Apelsiner som är mest lika äpplen.



SVM - Classifier & Hyperplane

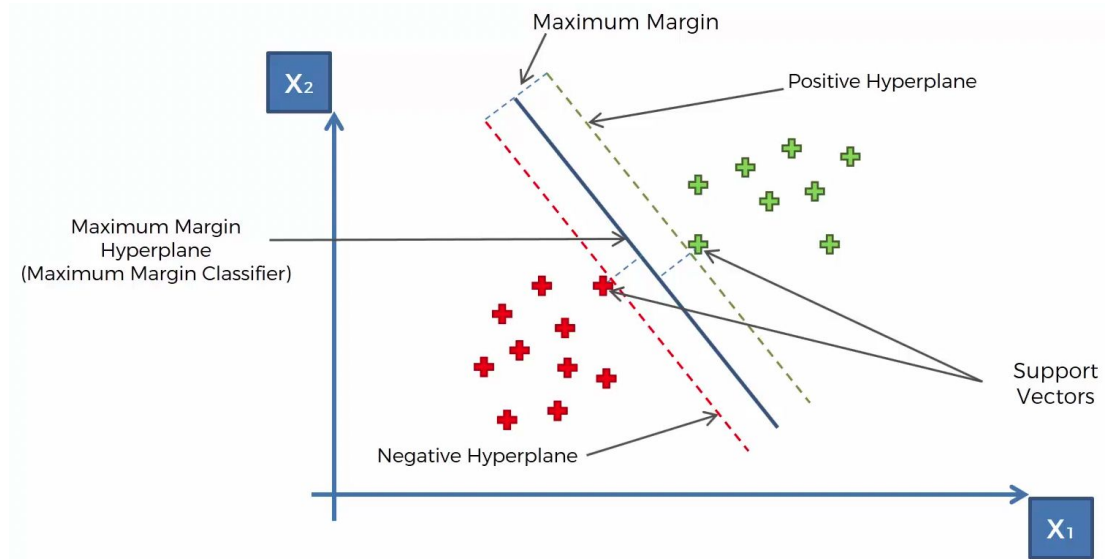
Benämningar på gränslinjen.

Två dimensioner - Classifier

Fler än två dimensioner - Hyperplane

Positivt hyperplane - ovan

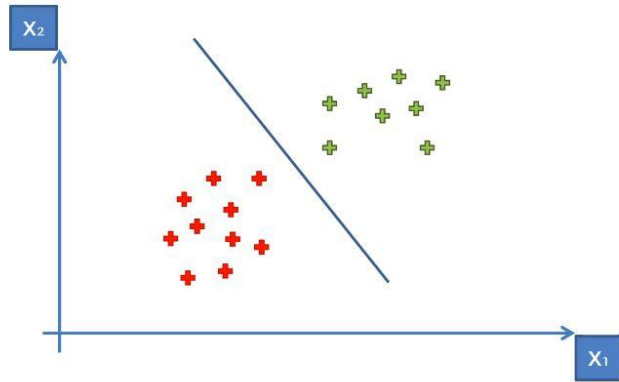
Negativt hyperplane - nedan



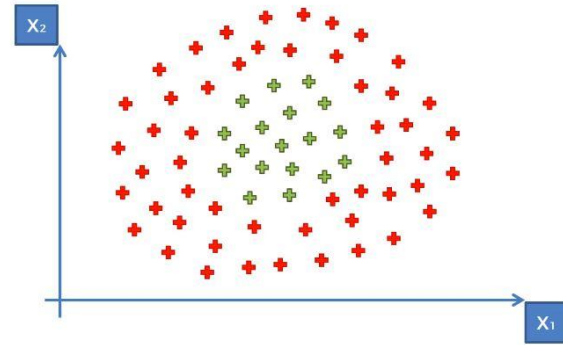
SVM Kernels

SVM - Separationen

Linjärt separerbara punkter



Icke linjärt separerbara punkter

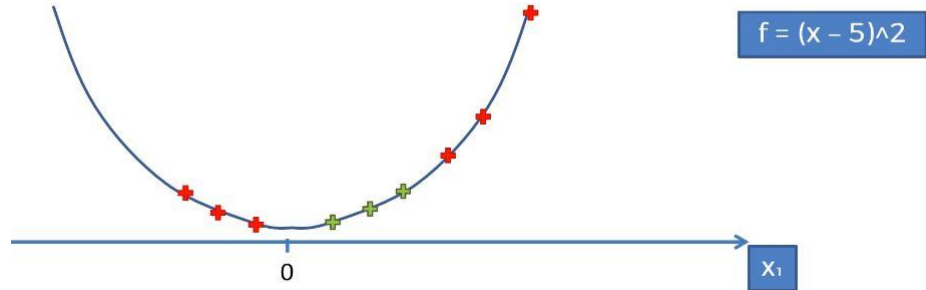


SVM - högre dimensioner

Ikke separerbara punkter i en dimension.

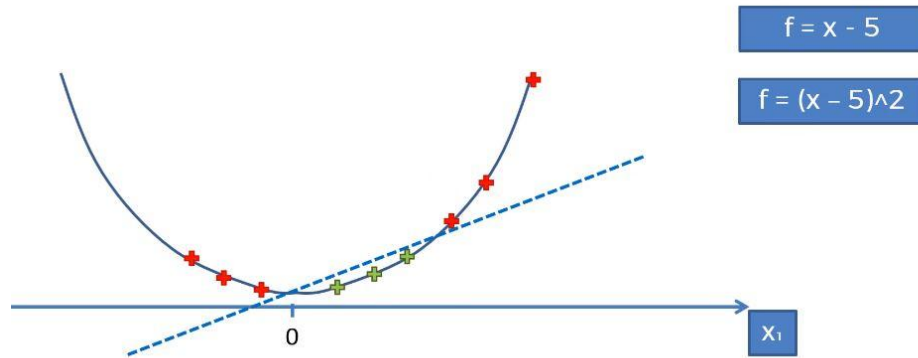


Kan separeras genom att lägga på en dimension.



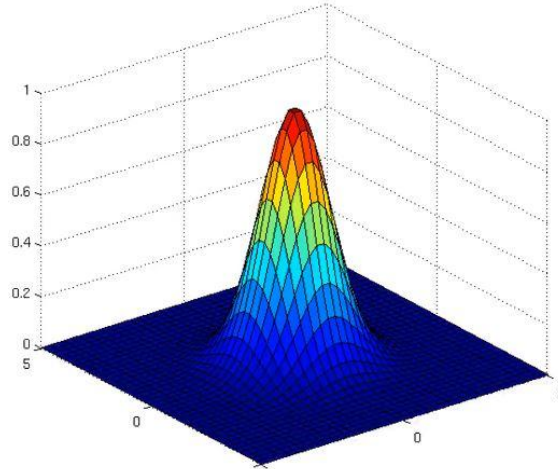
SVM - högre dimensioner

Nu kan en linje separera de två klasserna!



SVM - Kernel trick

Grafisk representation av **rbf** kärnan.



$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x} - \vec{l}^i\|^2}{2\sigma^2}}$$

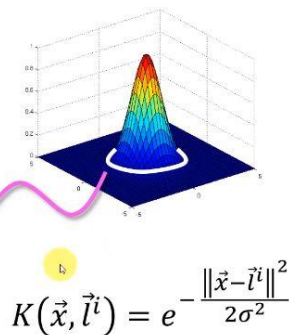
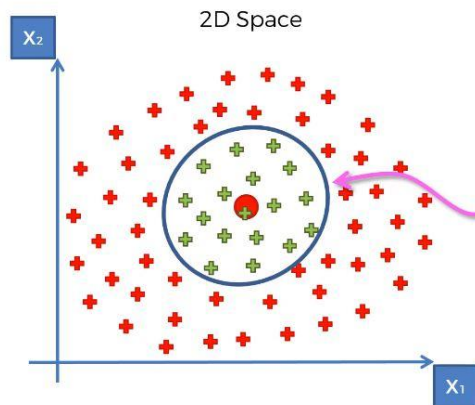
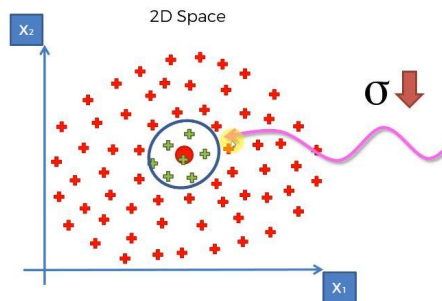
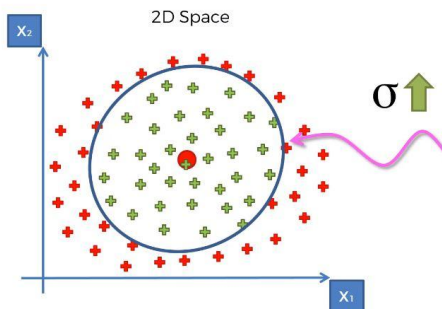
- Radial Basis Function
- Gaussian Function
- Normal distribution / Normalfördelning

SVM - RBF kernel

Projicera RBF kärnan på riktmärket.

Då avgörs vilka punkter som grupperas.

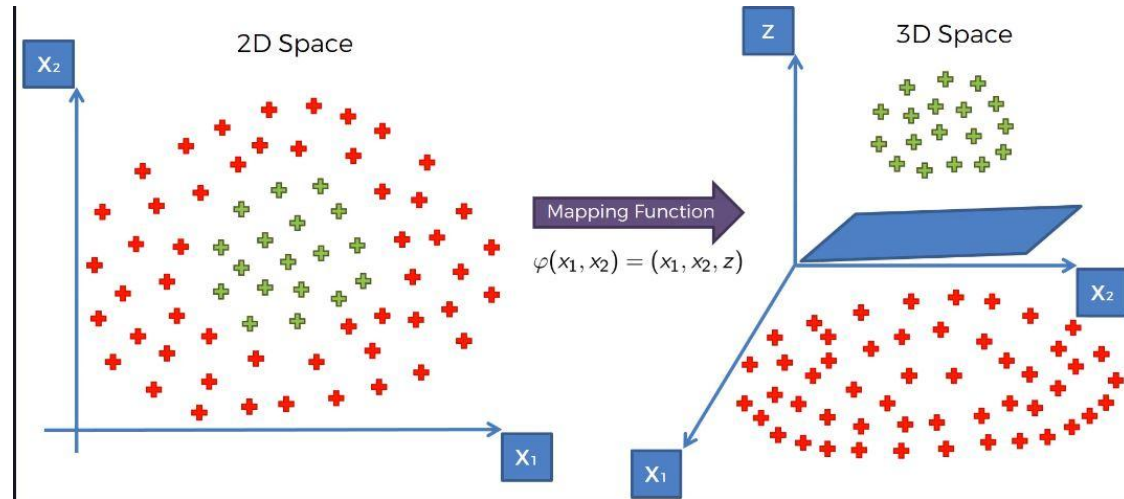
Genom att ändra talet **sigma** kan "trattens" bas **minskas** eller **ökas**.



RESTRÖY...

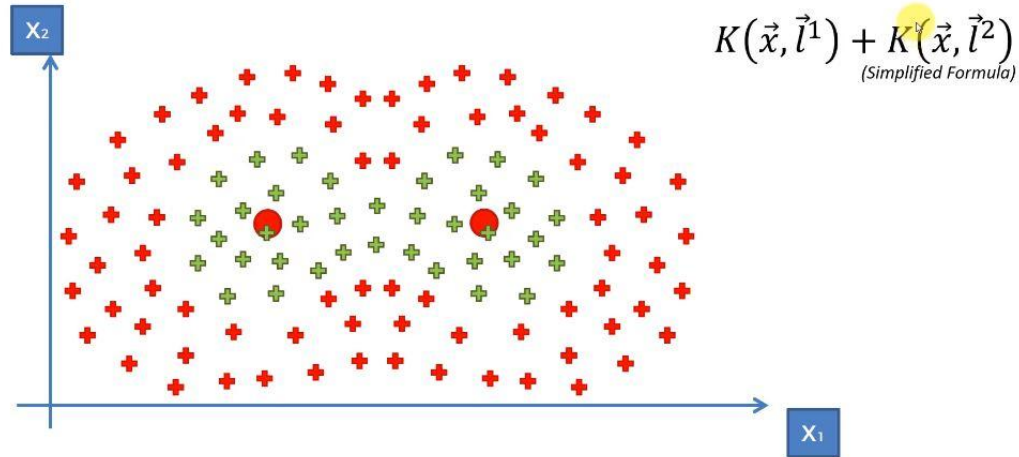
SVM - Kernel

Mapping Function - kärnan är en kartfunktion upp i en högre dimension.



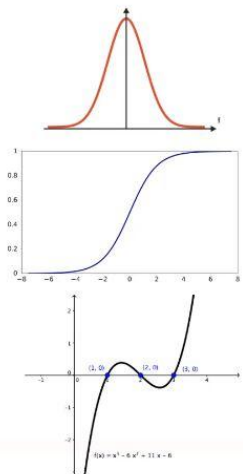
SVM - Multiple kernels

Genom att slå ihop fler kärnor kan mer
komplicerade områden grupperas



SVM - Kernels

Vanligt förekommande kärnor / kernels



Gaussian RBF Kernel

$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x} - \vec{l}^i\|^2}{2\sigma^2}}$$

Sigmoid Kernel

$$K(X, Y) = \tanh(\gamma \cdot X^T Y + r)$$

Polynomial Kernel

$$K(X, Y) = (\gamma \cdot X^T Y + r)^d, \gamma > 0$$



Ickelinjär SVR



RESTRÖY...

Hyperplan

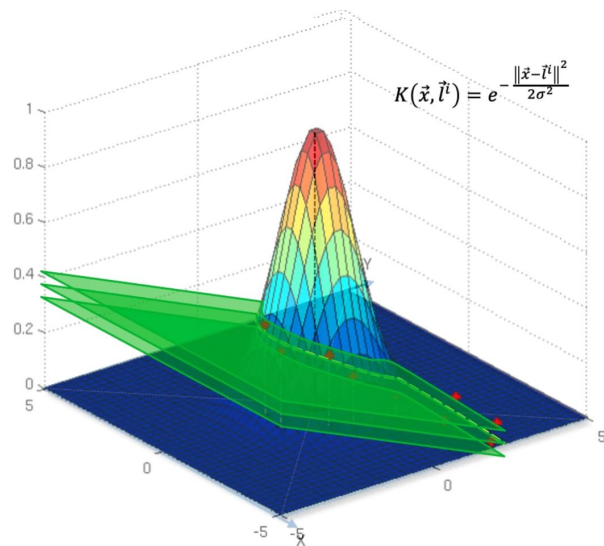
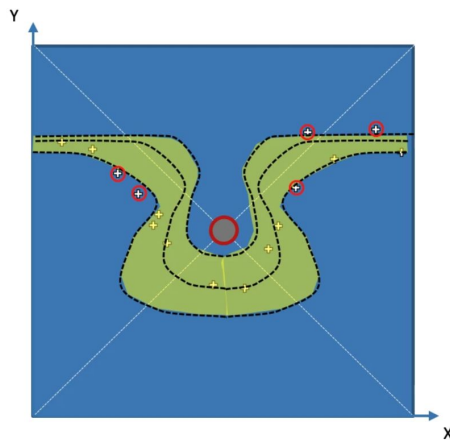
Punkter i två dimensioner.

Får en till dimension.

Kan nu separera de två klasserna med ett hyperplan.

Hyperplanet har marginaler i SVM.

I origindimensionen är inte linjen rak.



Länkar

- [Euklidiskt avstånd](#)
- [sklearn distance metric](#)
- [KNN - vid](#)
- [SVM kernel - main ideas - vid 1](#)
- [SVM polynomial kernel - vid 2](#)
- [SVM rbf kernel - vid 3](#)
- [Kernel example](#)
- [Applications of SVM in real life](#)
- [Weighted average](#)