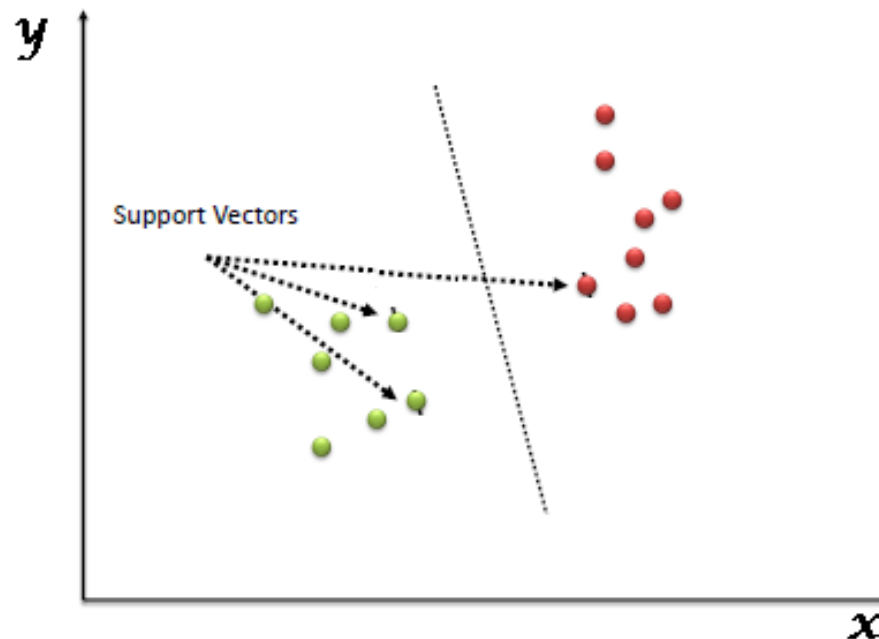


SVM

Support Vector Machines

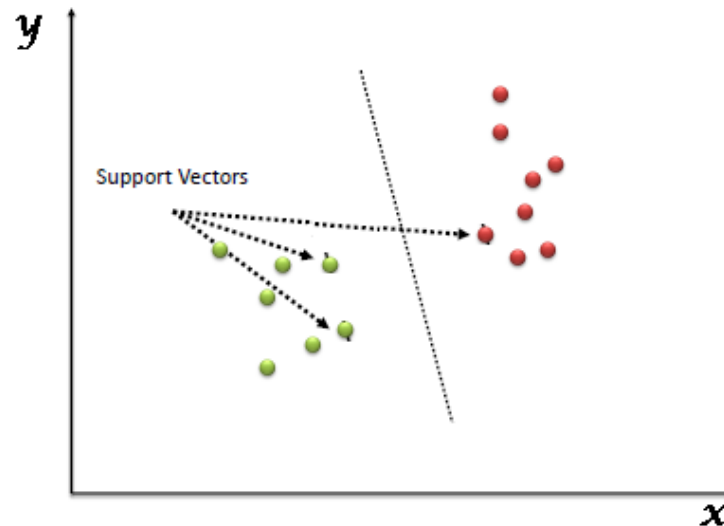
SVM

- Support Vector Machine (SVM)
- Populär supervised Machine Learning algoritm för klassificering (och regression)



SVM

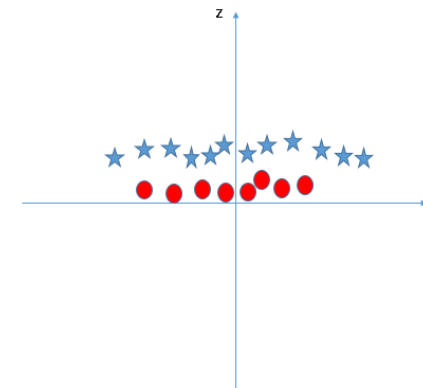
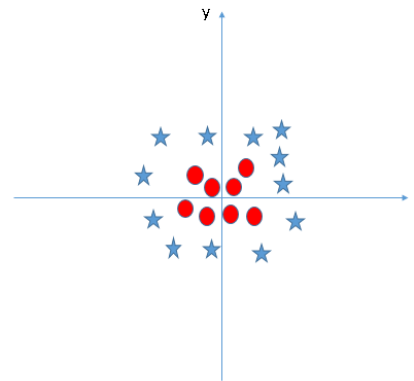
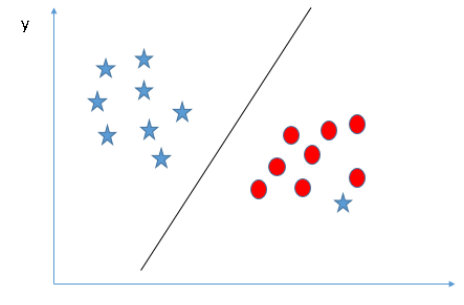
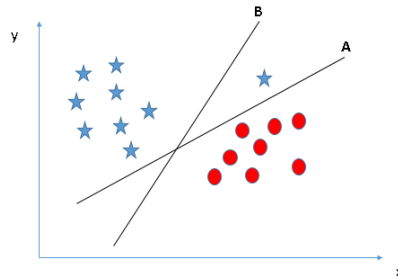
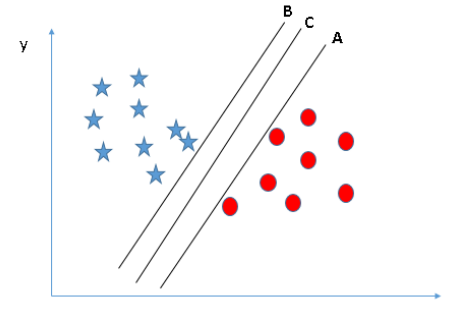
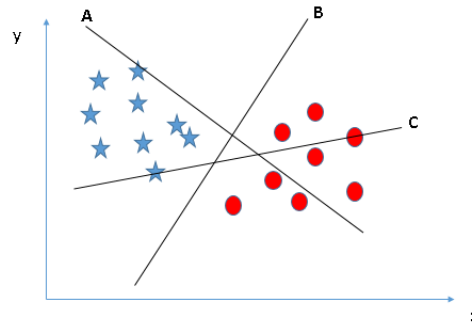
- I SVM plottas varje datapunkt i n -dimensionella planet där n är antal features
- Värdet på varje feature är värdet på koordinaten
- Klassificering sker genom att hitta hyperplanet som skiljer de två klasserna (i binära fallet)
- Hyperplan är ett delrum som har en mindre dimension än omgivningen. I 2-dimensionella utrymmen är hyperplanet 1-dimensionell
- Support Vectors är koordinaterna till varje observation



SVM

Hur fungerar det?

1. Hitta hyperplanet som separerar de två klasserna bäst
2. Maximera avståndet mellan närmaste datapunkt. Distansen kallas margin
3. Ignorera outliers
4. Skapa nya features för att transformera om datan



SVM

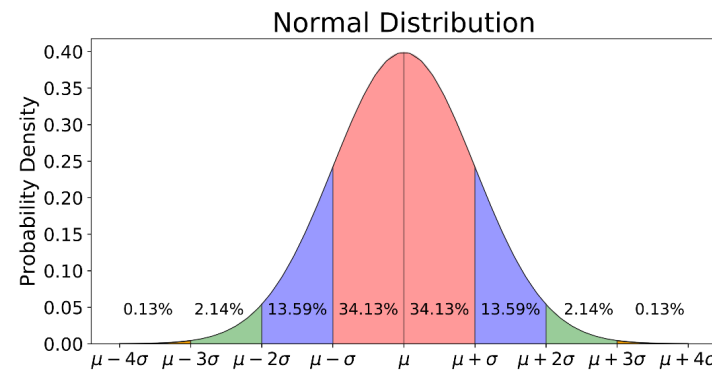
- <https://www.datacamp.com/community/tutorials/support-vector-machines-r>

GMM

Gaussian Mixture Models

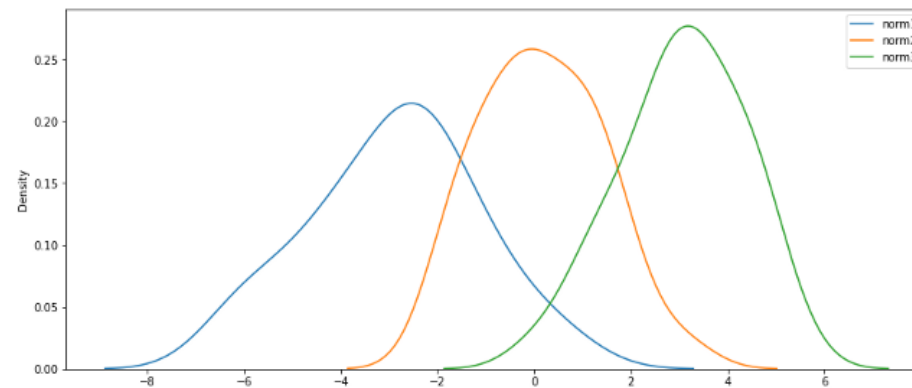
GMM

- Gaussian Mixture Models (GMM)
- Klustringsalgoritm
- Gaussian distribution = Normal distribution
- Idé: trots hela datasetet i sig inte är normalfördelad kan det finnas flera normalfördelningar där varje datapunkt finns



GMM

- GMM skapar kluster genom att anta datan i ett kluster är normalfördelad i klustret
- Det vill säga klustret är koncentrerad kring ett mean som är mittpunkten i klustret och har standard deviation som är hur klustret sprider sig "åt sidorna"
- Varje datapunkt analyseras för att se vilken normalfördelning (kluster) den är närmast
- Det gör modellen genom att kolla på sannolikheten att datapunkten tillhör alla de olika distributionerna
- Högst sannolikhet blir datapunktens kluster



GMM

- På ett mycket djupare plan
<https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95>