

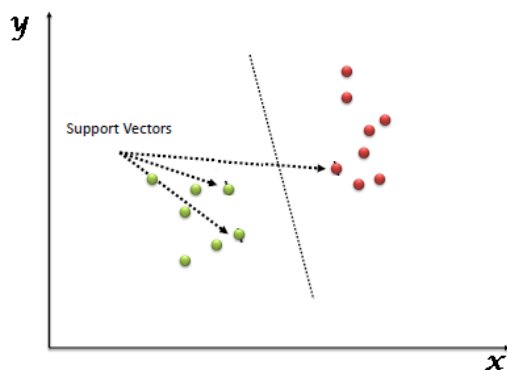
Nama : Ferza Reyaldi
NIM : 09021281924060
Mata Kuliah : Data Mining

Tugas Pertemuan 5 (Pertemuan 15)

Pelajari & tuliskan penjelasan tentang metode atau algoritma Support Vector Machine!

Jawab:

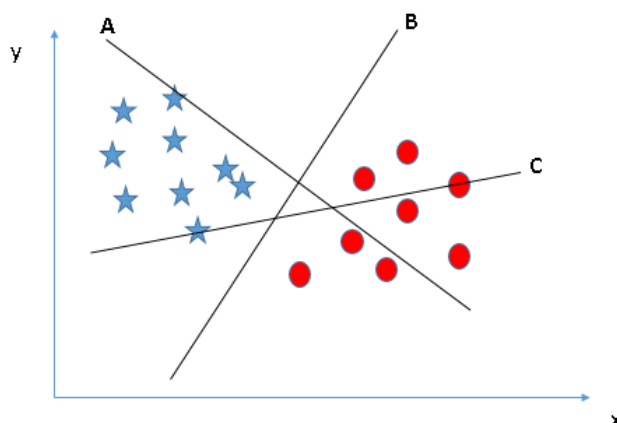
Algoritma Support Vector Machine adalah model *supervised learning* yang menganalisis data yang digunakan untuk analisis klasifikasi dan regresi. Dalam algoritma SVM, kita memplot setiap item data sebagai titik dalam ruang n-dimensi (n adalah sejumlah fitur yang dimiliki) dengan nilai setiap fitur menjadi nilai koordinat tertentu. Kemudian, kita melakukan klasifikasi dengan menemukan hyper-plane yang membedakan kedua kelas dengan sangat baik.



Support Vectors hanyalah koordinat pengamatan individu. Pengklasifikasi SVM adalah perbatasan yang paling baik memisahkan dua kelas (hyper-plane/line).

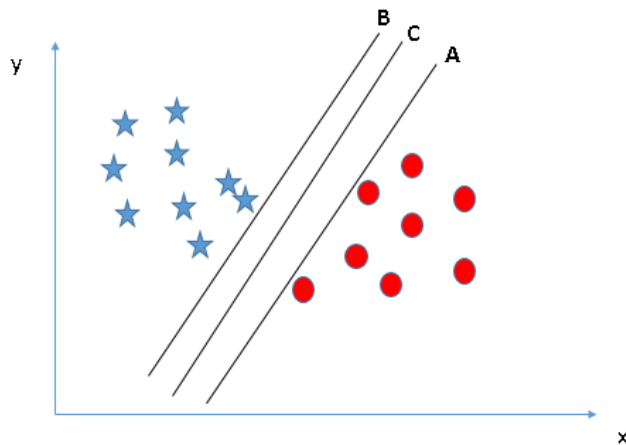
Bagaimana cara mengidentifikasi hyper-plane yang tepat?

- Skenario 1

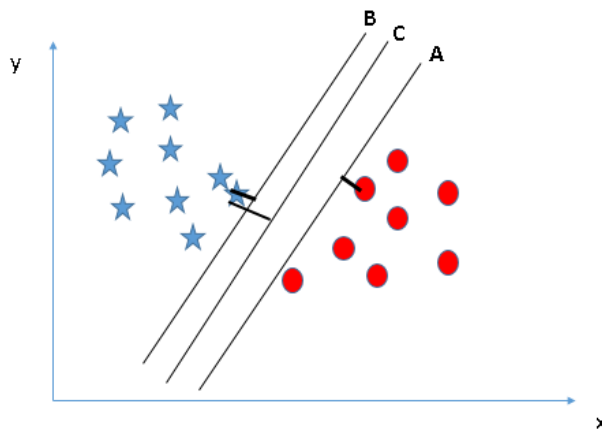


Terdapat 3 hyper-plane (A, B, dan C). Ada aturan praktis untuk mengidentifikasi hyper-plane yang tepat: "Pilih hyper-plane yang memisahkan dua kelas dengan lebih baik". Dalam skenario ini, hyper-plane B telah melakukan pekerjaan ini dengan sangat baik.

- Skenario 2

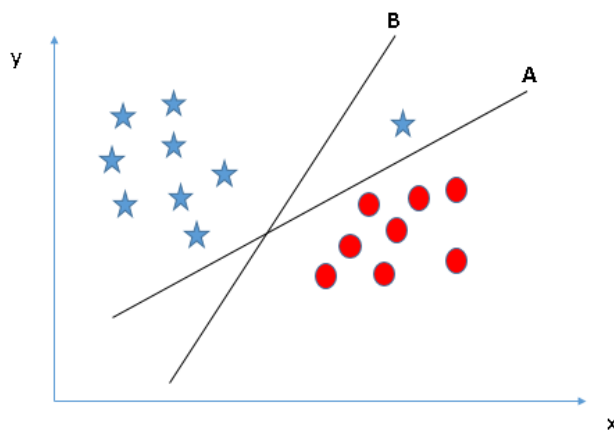


Terdapat 3 hyper-plane (A, B, dan C) dan semuanya memisahkan kelas dengan baik. Cara mencari hype-plane yang tepat adalah memaksimalkan jarak antara titik data terdekat (salah satu kelas) yang disebut sebagai Margin.



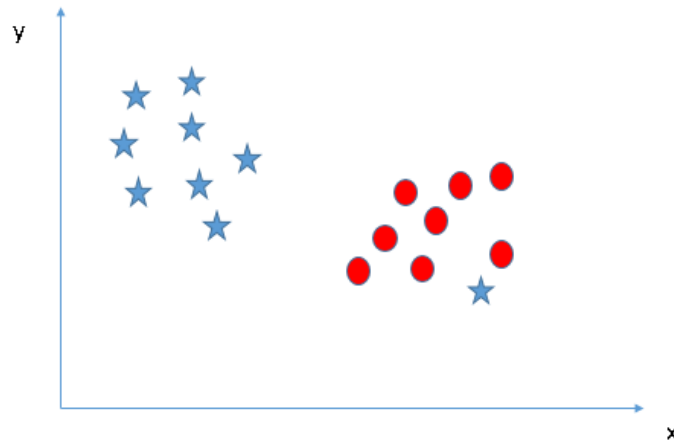
Di atas, Anda dapat melihat bahwa margin untuk hyper-plane C tinggi dibandingkan dengan A dan B. Oleh karena itu, kami menamai hyper-plane yang tepat sebagai C.

- Skenario 3

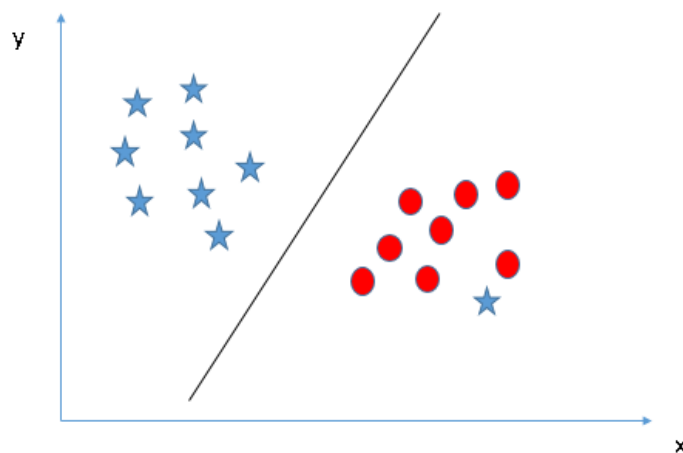


SVM memilih hyper-plane yang mengklasifikasikan kelas secara akurat sebelum memaksimalkan margin. Di sini, hyper-plane B memiliki kesalahan klasifikasi dan A telah mengklasifikasikan semuanya dengan benar. Oleh karena itu, hyper-plane yang tepat adalah A.

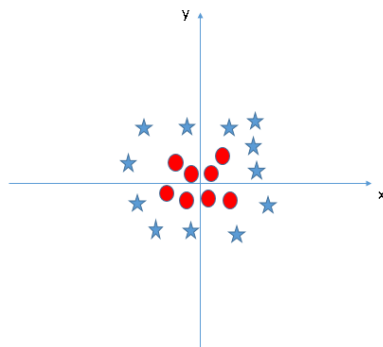
- Skenario 4 (Bisakah kita mengklasifikasikan dua kelas?)



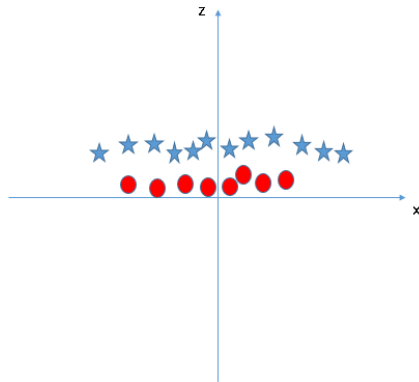
Algoritma SVM memiliki fitur untuk mengabaikan outlier dan menemukan hyperplane yang memiliki margin maksimum. Sehingga berikut hyper-plane yang tepat.



- Skenario 5
Dalam skenario di bawah ini, kita tidak memiliki hyper-plane linier antara dua kelas, jadi bagaimana SVM mengklasifikasikan kedua kelas ini?

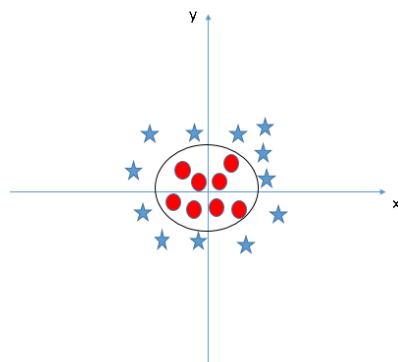


Ini memecahkan masalah ini dengan memperkenalkan fitur tambahan. Di sini, kita akan menambahkan fitur baru $z=x^2+y^2$. Sekarang, mari kita plot titik data pada sumbu x dan z:



algoritma SVM memiliki teknik yang disebut trik kernel. Kernel SVM adalah fungsi yang mengambil ruang input berdimensi rendah dan mengubahnya ke ruang dimensi yang lebih tinggi yaitu mengubah masalah yang tidak dapat dipisahkan menjadi masalah yang dapat dipisahkan. Hal ini sebagian besar berguna dalam masalah pemisahan non-linier. Sederhananya, ia melakukan beberapa transformasi data yang sangat kompleks, kemudian menemukan proses untuk memisahkan data berdasarkan label atau output yang telah Anda tetapkan.

Ketika kita melihat hyper-plane di ruang input asli, itu terlihat seperti lingkaran:



Pros:

- SVM bekerja sangat baik dengan margin pemisahan yang jelas.
- SVM efektif dalam ruang dimensi tinggi.
- SVM efektif dalam kasus di mana jumlah dimensi lebih besar daripada jumlah sampel.
- SVM menggunakan subset poin pelatihan dalam fungsi keputusan (disebut Support Vectors), sehingga juga hemat memori.

Cons:

- SVM tidak berfungsi dengan baik ketika memiliki kumpulan data yang besar karena waktu pelatihan yang dibutuhkan lebih tinggi.

- SVM juga tidak berkinerja baik, ketika kumpulan data memiliki lebih banyak noise, contohnya kelas target tumpang tindih
- SVM tidak secara langsung memberikan perkiraan probabilitas, ini dihitung menggunakan 5-fold cross-validation yang mahal.