Summary Machine Learning

Klasifikasi III (SVM)

Muh.Ikhsan (H071191049)

A. Support Vector Machine

1. Definisi

Support Vector Machine yang disebut sebagai SVM adalah algoritma pembelajaran terawasi yang dapat digunakan untuk masalah klasifikasi dan regresi sebagai Support Vector Classifier (SVC) dan Support Vector Regression (SVR). SVM ini digunakan untuk kumpulan data yang kecil karena terlalu lama untuk diproses. Memiliki prinsip dasar untuk melakukan klasifikasi dengan menggunakan batas pemisah.

SVM pertama kali diperkenalkan pada tahun 1992 oleh Vapnik dengan Partner Boser dan Guyon. Prinsip dasar SVM adalah classifier linear, yang kemudian dikembangkan untuk memecahkan masalah non-linear dengan mengintegrasikan konsep trik kernel ke dalam area kerja dimensi tinggi. SVM dapat mengklasifikasikan data linier dan non linier. Variabel-variabel prediktor merupakan data input sedangkan variabel target yang saling bergantung merupakan output. SVM bertujuan untuk menemukan fungsi klasifikasi terbaik dan untuk membedakan antara anggota dari dua kelas dalam data training. Matrik untuk konsep fungsi klasifikasi "terbaik" dapat diwujudkan secara geometris. Untuk dataset terpisah secara linear, fungsi klasifikasi linier berhubungan dengan hyperplane pemisah f (x) yang melewati tengah dua kelas, memisahkan keduanya.

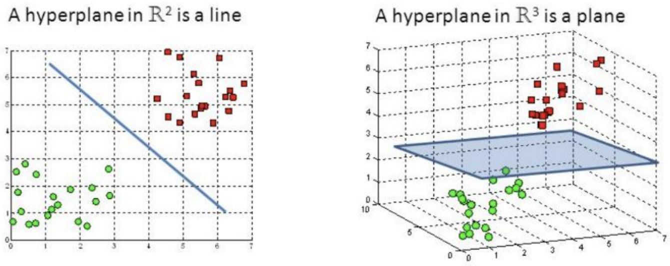
Model algoritma SVM merupakan salah satu algoritma dari metode klasifikasi, yang bekerja dengan cara mencari suatu garis (hyperplane) untuk memisahkan dua kelompok data. SVM menggunakan prinsip mencari margin maksimum pada batas (hyperplane) untuk menemukan batas pemisah terbaik antara tiap kelas. Bagaimana bentuk batas pemisah ini bergantung pada kernel yang akan kita gunakan. Untuk sistem dengan data tak terpisahkan, SVM akan menggunakan kernel untuk mentransformasi data masukan kedalam dimensional yang lebih tinggi (Jika awalnya hanya menggunakan 2 dimensi maka akan dinaikkan ke 3 dimensi (kedalaman)).

Terdapat beberapa kernel yang dapat digunakan:

a. Linear kernel
b. Polinomial Kernel
c. Radial Basis Function Kernel

2. Dasar dibalik SVM

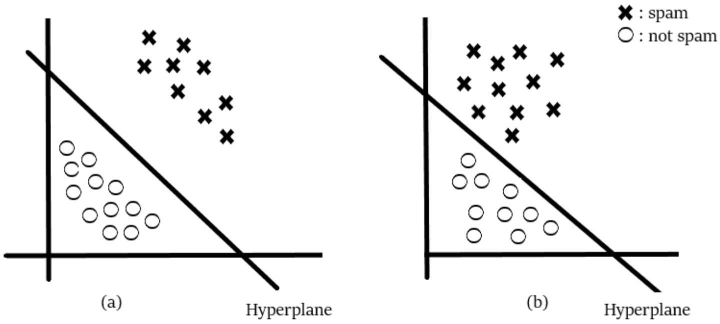
   SVM didasarkan pada gagasan untuk menemukan hyperplane yang paling baik memisahkan fitur ke dalam domain yang berbeda.



3. Contoh Kasus

   Pertimbangkan situasi berikut :

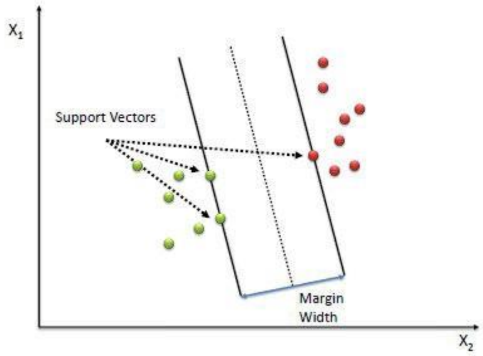
   Ada stalker yang mengirimi Anda email dan sekarang Anda ingin merancang sebuah fungsi (hyperplane) yang akan dengan jelas membedakan kedua kasus tersebut, sehingga setiap kali Anda menerima email dari penguntit itu akan diklasifikasikan sebagai spam. Berikut ini adalah gambar dua kasus di mana hyperplane digambar, mana yang akan Anda pilih dan mengapa? luangkan waktu sejenak untuk menganalisis situasi.



   Email dalam gambar(a) diklasifikasikan dengan jelas dan lebih meyakinkan dibandingkan dengan gambar(b). Pada dasarnya, SVM terdiri dari gagasan untuk menghasilkan hyperplane Optimal yang akan dengan jelas mengklasifikasikan kelas yang berbeda (dalam hal ini mereka adalah kelas biner).

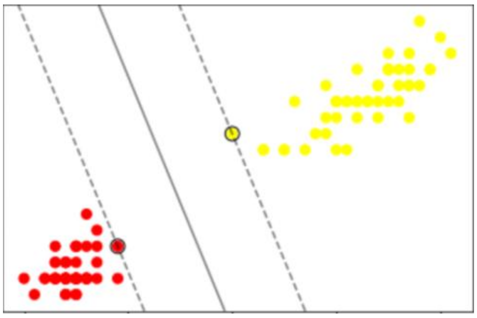
4. Terminologi yang digunakan dalam SVM

   Titik-titik yang paling dekat dengan *hyperplane* disebut sebagai titik *support vector* dan jarak vektor dari *hyperplane* disebut *margin*.



Intuisi dasar untuk dikembangkan di sini adalah bahwa semakin jauh titik SV, dari hyperplane, semakin besar kemungkinan mengklasifikasikan titik dengan benar di wilayah atau kelasnya masing-masing. Titik SV sangat penting dalam menentukan hyperplane karena jika posisi vektor berubah posisi hyperplane berubah. Secara teknis hyperplane ini juga bisa disebut sebagai margin maximizing hyperplane. Margin maksimum adalah pilihan yang aman karena jika terjadi sedikit kesalahan pada data maka akan memberikan kemungkinan terkecil terjadi kesalahan klasifikasi.

5. Hyperplane



Hyperplane adalah garis pemisah terbaik antara dua kelas. Untuk mencari hyperplane dapat dilakukan dengan mencari margin hyperplane. dan mencari titik maksimum. Margin adalah jarak antara data terdekat di antara dua kelas yang berbeda, yang disebut dengan support vector. Garis solid merah pada gambar menunjukkan hyperplane yang terbaik, karena terletak tepat diantara kedua class, sedangkan support vector dilambangkan dengan titik merah dan kuning yang berada di dalam lingkaran hitam.

Hyperplane klasifikasi linear SVM dinotasikan

$$f(x) = w, x + b = 0$$

Dari persamaan di atas di dapatkan pertidaksamaan kelas +1 (negatif)

$$w \cdot x + b \leq +1$$

Pertidaksamaan kelas -1:

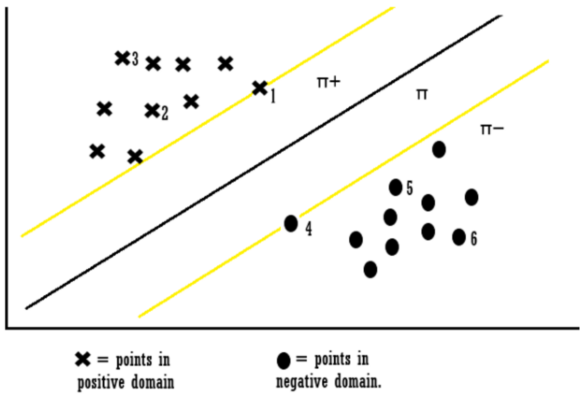
$$w \cdot x + b \geq -1$$

pusat. Dengan mengoptimalkan nilai jarak antara hyperplane dan titik berikutnya, margin terbesar dapat ditemukan, yaitu $\frac{1}{|W|}$. Ini dapat dirumuskan sebagai masalah pemrograman kuadratik (QP)

$$\min \frac{1}{2}\left||w|\right|^2 = \min \frac{1}{2}(w_1^2 + w_2^2)$$
$$y_i(w_{xi} + b) \geq 1, i = 1,2,3, \dots, N$$

6. Hard margin SVM

Asumsikan 3 hyperplane yaitu (π, +,) sedemikian rupa sehingga 'π+' sejajar dengan 'π' yang melalui vektor pendukung di sisi positif dan 'π–' sejajar dengan 'π' yang melalui vektor pendukung di sisi negatif.



Persamaan setiao hyperplane dapat dianggap sebagai :

$$\pi = b + w^T X = 0$$
$$\pi^+ = b + w^T X = 1$$
$$\pi^- = b + w^T X = -1$$

untuk titik $X1$:

$$y_1 = 1$$
$$y_1(w^T x_1 + b) > 1$$

ketika titik X1 kita dapat mengatakan bahwa titik itu terletak pada hyperplane dan persamaan menentukan bahwa produk dari output aktual kita dan persamaan hyperplane adalah 1 yang berarti titik tersebut diklasifikasikan dengan benar dalam domain positif.

untuk titik $X3$:

$$y_1 = -1$$
$$y_1(w^T x_1 + b) = 1$$

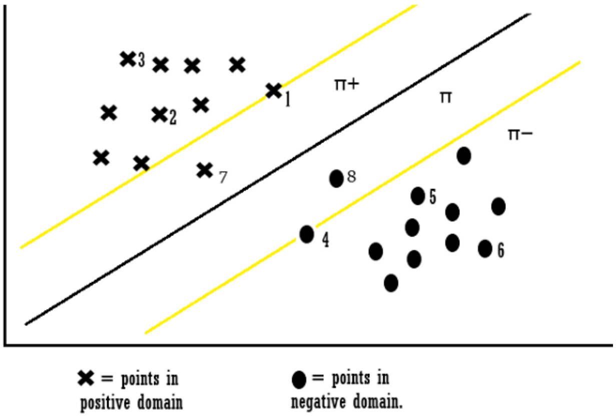
ketika titik X3 kita dapat mengatakan bahwa titik itu menjauh dari hyperplane dan persamaan menentukan bahwa produk dari output *aktual kita dan persamaan hyperplane lebih besar 1 yang berarti titik tersebut diklasifikasikan dengan benar dalam domain positif.*

untuk titik *X6* :

$$y_1 = -1$$
$$y_1(w^T x_1 + b) > 1$$

ketika titik X6 kita dapat mengatakan bahwa titik itu terletak jauh dari hyperplane di wilayah negatif dan persamaan menentukan bahwa produk dari output aktual kita dan persamaan hyperplane lebih besar 1 yang berarti titik tersebut diklasifikasikan dengan benar dalam domain negatif.

Berikut Batasan yang tidak diklasifikasikan



untuk titik X7:

Ketika $Xi = 7$ *titik diklasifikasikan salah karena untuk titik* 7 $wT + b$ *akan lebih kecil dari satu dan ini melanggar batasan. Jadi kami menemukan kesalahan klasifikasi karena pelanggaran kendala. Demikian pula, kita juga dapat mengatakan untuk poin* $Xi = 8$.

Jadi berdasarkan data diatas dapat disimpulkan bahwa :

If $Yi(WT^* X_i + b) \geq 1$:
    Then X_i *memiliki klasifikasi yang benar*
Else :
    X_i *tidak diklasifikasi dengan benar*

Jadi kita dapat melihat bahwa jika titik-titik tersebut dapat dipisahkan secara linier maka hanya hyperplane kita yang dapat membedakannya dan jika ada outlier pada data maka model tidak dapat memisahkannya.

Jadi jenis SVM ini disebut sebagai SVM margin keras (karena kami memiliki batasan yang sangat ketat untuk mengklasifikasikan setiap titik data dengan benar).

7. Soft margin SVM
   Ide ini didasarkan pada premis sederhana: memungkinkan SVM untuk membuat sejumlah outlier dan menjaga margin selebar mungkin sehingga poin lain masih dapat diklasifikasikan dengan benar . Ini dapat dilakukan hanya dengan memodifikasi tujuan SVM. Untuk alasan ini, diperkenalkan variabel Slack (ε) baru yang disebut Xi.
   Jika kita memasukkan ε ke dalam persamaan kita sebelumnya, kita dapat menulis ulang sebagai
   $$y_i(w^T x_i + b) \geq 1 - \varepsilon_i$$

   Di sini, sisi kiri pertidaksamaan dapat dianggap seperti kepercayaan klasifikasi. Skor kepercayaan 1 menunjukkan bahwa pengklasifikasi telah mengklasifikasikan titik dengan benar. Namun, jika skor kepercayaan kurang dari 1, itu berarti pengklasifikasi tidak mengklasifikasikan titik dengan benar.

8. Kelebihan dan Kekurangan SVM
   a. Kelebihan
      1) Generalisasi dapat diartikan sebagai kemampuan suatu metode SVM untuk mengklasifikasikan pola yang tidak berisi data yang digunakan dalam fase pembelajaran metode ini.
      2) Curse Of Dimensionality adalah masalah yang biasanya dihadapi ketika proses pengenalan pola ketika memperkirakan parameter. Karena jumlah sampel data relatif kecil dibandingkan dengan ruang data vektor,sehingga semakin tinggi ruang vektor yang diproses, ini mengarah pada konsekuensi yang memerlukan jumlah data dengan tiga dimensi.
      3) Feasibility, SVM dapat diimplementasikan dengan sangat mudah karena proses penentuan support vektor dalam masalah QP dapat dirumuskan. Jadi jika kita memiliki perpustakaan untuk menyelesaikan masalah QP, SVM itu sendiri dapat digunakan dengan sangat mudah.

   b. Kekurangan
      1) Sulit digunakan untuk pengolahan data yang mempunyai jumlah data yang besar.
      2) Metode SVM secara teoritis dikembangkan untuk masalah klasifikasi dengan dua atau lebih kelas. Namun masing-masing strategi ini memiliki kelemahan, sehingga untuk pengembangan

SVM pada masalah yang lebih dari dua kelas masih menjadi topik penelitian terbuka

B. Support Vector Regression (SVR)

Support Vector Regression (SVR) merupakan suatu metode SVM yang diterapkan pada kasus regresi. Menurut (Scholkopt dan Smola, 2012), SVR bertujuan untuk menemukan sebuah fungsi f(x) sebagai suatu hyperplane (garis pemisah) berupa fungsi regresi yang mana sesuai dengan semua input data dengan membuat error (ε) sekecil mungkin. Menurut Santoso (2007), misalkan dipunyai l set data training, $(Xi, Yi)\ , i\ =\ 1,2, \ldots, l$ dimana Xi merupakan vektor input

$$x = \{x_1, x_2, \ldots, x_n\} \subseteq \Re^n$$

dan output skalar 2 dan l adalah banyaknya data training. Dengan metode SVR diperoleh fungsi regresi sebagai berikut

$$f(x) = w^T \phi(x) + b$$

Dimana

w : vector pembobot berdimensi *l*

$\phi(x)$: fungsi yang memetakan *x* pada ruang dengan dimensi *l*

b : bias

Menurut Rezzy et al. (2017), Agar mendapatkan generalisasi yang baik untuk fungsi regresi f(x), dapat dilakukan dengan cara meminimalkan norm dari w. Oleh karena itu perlu adanya penyelesaian problem optimasi:

$$\min \frac{1}{2} \left||w|\right|^2$$

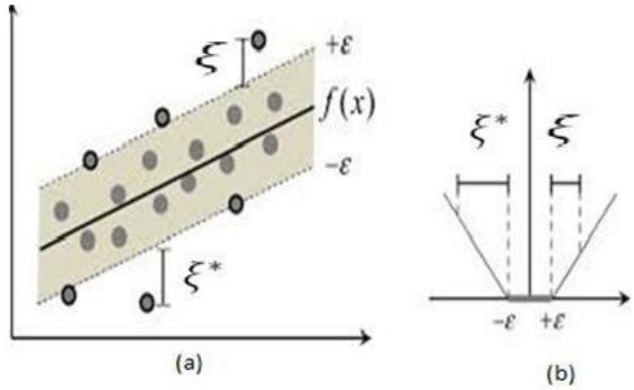
Dengan syarat

$$y_i - w^T \phi(x_i) - b < \mathcal{E}$$

$$w^T \phi(x_i) - y_i + b \leq \mathcal{E}, dimana\ i = 1,2, \ldots l$$

Pada persamaan diatas diasumsikan bahwa semua titik berada dalam rentang

$f(x) \pm \mathcal{E}$, dalam hal ketidaklayakan (*infesiable)* dimana ada beberapa titik yang mungkin keluar dari rentang $f(x) \pm \mathcal{E}$ sehingga dapat ditambahkan variable ξ dan ξ^* untuk mengatasi masalah pembatas yang tidak layak dalam masalah optimasi.



Gambar diatas menjelaskan bahwa semua titik di luar margin akan dikenai pinalti sebesar C. Selanjutnya, masalah optimasi di atas dapat diformulasikan sebagai berikut:

$$\min\{\frac{1}{2}||w||^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)\}$$

dengan syarat

$y_i - w^T\phi(x_i) - b - \xi_i \leq \mathcal{E}$

$w^T\phi(x_i) - y_i + b - \xi_i^* \leq \mathcal{E}$

$\xi_i, \xi_i^* \geq 0, dimana\ i = 1,2,..,l$

Loss function merupakan fungsi yang menunjukkan hubungan antara error dengan bagaimana error ini dikenai pinalti. Perbedaan loss function akan menghasilkan formulasi SVR yang berbeda (Santoso, 2007). Menurut Amanda (2014), loss function yang paling sederhana ε-insensitive loss function. Formulasi ε-insensitive loss function sebagai berikut:

$$L_{\mathcal{E}}(y) = \begin{cases} 0, untuk\ |f(x) - y| < \mathcal{E} \\ |f(x) - y| - \mathcal{E}, untuk\ yang\ lain \end{cases}$$

Langkah-langkah peramalan menggunakan SVR

1. Membagi data runtun waktu (time series) menjadi dua, yaitu data training dan testing.
2. Menentukan jenis fungsi kernel dan loss function yang digunakan untuk peramalan.
3. Menentukan nilai dari C dan parameter kernel.
4. Mencari nilai beta dan bias.
5. Melakukan peramalan terhadap data testing.
6. Menghitung nilai MAPE (Mean Absolute Percentage Error).
7. Menentukan fungsi kernel, loss function, dan parameter terbaik, dimana fungsi kernel, loss function, dan parameter terbaik tersebut menghasilkan MAPE terkecil.
8. Melakukan peramalan ke depan menggunakan fungsi kernel, loss function, dan parameter terbaik