Metode Support Vector Machines Pada Klasterisasi K-Means Data Nonlinear Separable

Cori Pitoy
Jurusan Matematika FMIPA UNIMA
cory_pitoy@unima.ac.id

ABSTRAK

Data dalam dunia nyata jarang memiliki struktur linear sehingga dapat dengan sulit dipisahkan secara linear (*linearly separable*). Data yang tak terpisahkan secara linear (*non-linearly separable*) pada klasterisasi K-Means dapat dianalisis menggunakan pendekatan *Support Vector Machines* (SVM), dimana ruang input (*input space*) untuk data awal yang kompleks dalam ruang berdimensi rendah dipetakan ke ruang fitur (*feature space*) yang berdimensi tinggi melalui sebuah transformasi Kernel, dalam 2 tahap. Tahap pertama, data $x_1, \dots, x_n$ pada ruang data ^{d} diekstrak melalui pemetaan menggunakan fungsi kernel $\Phi(x_i)$ ke ruang fitur yang berdimensi lebih tinggi, $\Phi: {}^d \rightarrow$, sehingga *linearly separable.* Selanjutnya dicari *optimal hyperplane* secara linear oleh SVM.

Kata kunci : SVM, Klasterisasi K-Means, Data Nonlinear Separable, Kernel

ABSTRACT

Data in the real world rarely has a linear structure so that it can be difficult to linearly separate (linearly separable). Observations that are linear inseparable (non-linearly separable) in the K-Means klastering can be analized using the Support Vector Machines (SVM) approach, where the data space for complex initial observations in low dimensional space is mapped to feature space which has a high dimension through a Kernel transformation, which was modified into two stages. The first step, observations $x_1, \dots, x_n$ in the ^{d} data space is extracted through mapping using the kernel function $\Phi(x_i)$ to the higher dimension feature space, $\Phi: {}^d \rightarrow$, so that it is linearly separable. Furthermore, it is sought a linear optimal hyperplane by SVM.

Keywords: SVM, K-Means Klastering, Nonlinear Separable Data, Kernel

PENDAHULUAN

Klasterisasi K-Means (MacQueen, 1967) merupakan salah satu metode klasterisasi data non hirarki yang mempartisi data ke dalam k klaster sehingga data dengan karakteristik yang sama dikelompokkan ke dalam klaster yang sama sedangkan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam klaster lainnya. Adapun tujuannya adalah untuk meminimalisasikan variasi di dalam suatu klaster dan memaksimalisasikan variasi antar klaster.

Tahapan algoritma K-Means (Johnson and Witchern (1988):

1. Tentukan besarnya k (banyaknya klaster), dan sentroid di tiap klaster.
2. Hitung jarak antara setiap data dengan setiap sentroid. Masukkan tiap-tiap data ke suatu klaster berdasarkan jarak terdekat dengan sentroid klaster yang berpadanan. Jarak biasanya adalah jarak Euclid. Hitung kembali sentroid untuk tiap klaster yang baru terbentuk.
3. Ulangi langkah 2 sampai tidak ada lagi pemindahan data antar klaster.

Algoritma K-Means membutuhkan 3 parameter yang spesifik: banyaknya klaster k, inisialisasi klaster dan ukuran jarak yang digunakan (Jain, 2010).

Keunggulan metode K-Means adalah mudah diterapkan, sederhana, efisien dan keberhasilannya telah terbukti secara empirik (Jain, 2010),

serta waktu yang dibutuhkan untuk menjalankannya relatif lebih cepat (Russell and Norvig, 2010). Untuk jumlah variabel yang banyak dengan ukuran k kecil, perhitungan K-Means lebih cepat dibanding *hierarchical klastering* (Santini, 2016).

Terdapat beberapa kelemahan yang membatasi algoritma klasterisasi K-Means, diantaranya tidak dapat mengekstrak semua informasi dari data bila densitas probabilitas data bukan Gaussian (Lee and Choi, 2007), penentuan sentroid awal yang bersifat acak (Sakthi and Thanamani, 2011) dan sensitif terhadap data yang diduga sebagai outlier (Singh and Singh, 2013).

Permasalahan

Metode K-Means akan bekerja dengan baik pada data yang berbentuk *spherical* (*spherical* K-Means). Untuk data berbentuk *spherical* klasterisasi K-Means memberikan hasil yang lebih baik dari Info-K Means yang dikembang-kan oleh Wu (2012). Namun kenyataannya, tidak semua data memiliki struktur linear. Beberapa kelompok data menyebar dengan struktur tertentu dan memerlukan suatu metode pemecahan khusus. Gambar 1. Adalah ilustrasi kelompok data dengan berbagai bentuk khusus (non linear) yang kalau dimodelkan dengan metode K-Means, (*Hard* atau *Fuzzy* K-Means) akan memberikan hasil yang tidak mewakili keadaan kelompok data tersebut.

Gambar 1. Data Menyebar Nonlinearly Separable

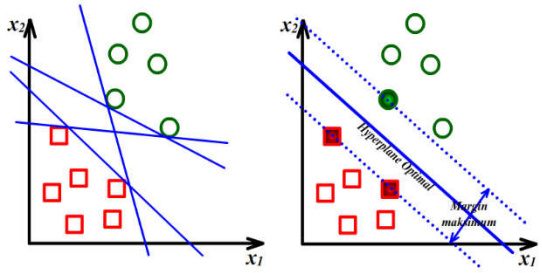
HASIL DAN PEMBAHASAN

Support Vector Machines

SVM dikenal sebagai teknik pembelajaran mesin (*machine learning*) yang paling mutakhir setelah pembelajaran mesin sebelumnya yang dikenal sebagai Neural Network (NN). Prinsip dasar SVM adalah klasifikasi data secara linear. SVM menampilkan klasifikasi melalui konstruksi *hyperplane* berdimensi-N yang secara optimal memisahkan/membagi data ke dalam 2 kategori. *Hyperplane* pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya.

Dengan demikian tujuan dari pemodelan SVM adalah untuk menemukan *hyperplane* yang dapat membagi/memilah klaster-klaster sebagai berikut, kasus dari suatu kategori dari variabel target berada pada satu sisi dari *plane* dan kasus dengan kategori lainnya berada pada sisi lain dari plane tersebut. Vektor-vektor berada pada *hyperplane* dinamakan *Support Vectors* (SV).

Analisis SVM berupaya menemukan suatu garis *hyperplane* yang berorientasi-kan bahwa margin diantara SV maksimum (Gambar 2). Upaya mencari lokasi *hyperplane* optimal ini merupakan inti dari proses pembelajaran pada SVM.



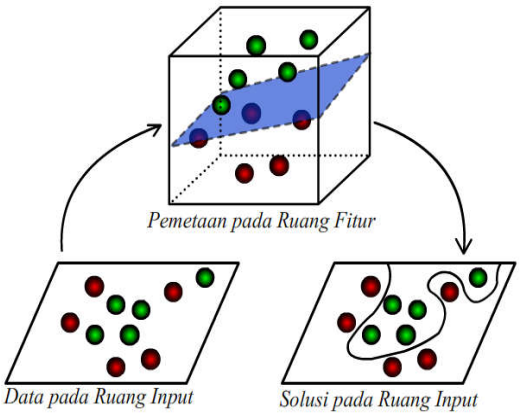
Gambar 2. Support Vector (SV)

Non Linear SVM

Umumnya, masalah dalam dunia nyata (*real world problem*) jarang yang bersifat linearly separable. Kebanyakan masalah-masalah tersebut bersifat non linear. Untuk menyelesaikan kasus non linear, perhitungan SVM dimodifikasi menjadi dua tahap, dimana dalamnya dimanfaatkan konsep yang disebut *Kernel trick.* Proses pemetaan dilakukan:

1. Data $x_1,x_2,\ldots,x_n$ in ^{d}.
   Pemetaan Φ dari input space ke ruang berdimensi lebih tinggi (>d) .
2. Φ : $^d \rightarrow$, : *feature space* , Batas klasifikasi linear dalam merupakan batas klasifikasi non linear dalam input space original.

SVM yang dibentuk dikenal dengan nama SVM non-linear



Gambar 3. Pemetaan data dari Ruang Input ke Ruang Fitur dan Solusinya pada Ruang Input

SVM merupakan salah satu varian dari *linear machine* sehingga hanya dapat digunakan untuk menyelesaikan masalah yang sifatnya *linearly separable*. Prinsip dasar SVM adalah klasifikasi data secara linear, yang dilakukan melalui konstruksi *hyperplane* berdimensi-*d*, yang secara optimal memisahkan data ke dalam 2 kategori.

Hyperplane $\boldsymbol{H}$ pada ^{d} adalah himpunan titik $\boldsymbol{x}=(x_1,\ldots,x_n)^T$, yang memenuhi $w_1x_1 + \quad + w_nx_n = b$ atau $\boldsymbol{w} \bullet \boldsymbol{x} = b$, $\boldsymbol{w}$ adalah *weight vector*, b adalah bias, yaitu konstan yang mewakili posisi bidang relatif terhadap pusat koordinat, dan $\boldsymbol{w} \bullet \boldsymbol{x}$ adalah perkalian titik antara $\boldsymbol{w}$ dan $\boldsymbol{x}$. *Hyperplane* $\boldsymbol{H}$ ortogonal terhadap $\boldsymbol{w} = (w_1, \ldots, w_n) \neq \boldsymbol{0}$ dalam ^{d}. Vektor $\boldsymbol{w}$ disebut vektor normal bagi $\boldsymbol{H}$.

Formulasi Primal

Misalkan dilakukan klasifikasi linear terhadap $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}^T$ dengan $y_i \in \{1,-1\}$ adalah label kelas dari x_i.

Fungsi keputusan :

$f(\mathbf{x}) = \text{sign}(\boldsymbol{w} \bullet \boldsymbol{x} + \text{b})$ ……(1)

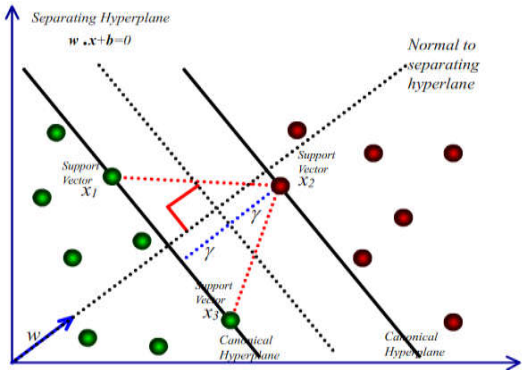
dan *canonical hyperplane* :

$\boldsymbol{w} \bullet x_i + \text{b} \geq +1$ for $y_i = +1$ ....(2)

$\boldsymbol{w} \bullet x_i + \text{b} \leq \quad 1$ for $y_i = 1$....(3)

Bila proyeksi ortogonal $\boldsymbol{u}$ ke $\boldsymbol{v}$ adalah $\boldsymbol{w}_1 = \frac{\boldsymbol{u} \bullet \boldsymbol{v}}{\|\boldsymbol{v}\|_2^2} \boldsymbol{v}$, dan $\|\boldsymbol{w}_1\| = \frac{\boldsymbol{u} \bullet \boldsymbol{v}}{\|\boldsymbol{v}\|_2}$, maka panjang proyeksi ortogonal $(x_1 \quad x_2)$ ke normal *separating hyperplane* $\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}$ adalah

$2\gamma = (((x_1 \quad x_2) \bullet \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}) / \left\| \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2} \right\|_2) = ((x_1 \quad x_2) \bullet$



$\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}$)(Campbell&Ying,2011)

Gambar 4. Proyeksi (x_1-x_2) ke Normal ke Separating Hyperplane $2\gamma = \boldsymbol{w}/\|\boldsymbol{w}\|_2$

Dengan demikian, jarak antar kelas adalah jarak antara kedua *canonical hyperplane* (*margin band*) adalah :

$\boldsymbol{w} \bullet (x_1\text{-}x_2)) = 2 \Rightarrow (\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2} \bullet (x_1\text{-}x_2)) = \frac{2}{\|\boldsymbol{w}\|_2}$ .....(4)

dan *margin width* adalah :

$\gamma = \frac{1}{\|\boldsymbol{w}\|_2}$.....(5)

Hyperplane terbaik diperoleh dengan memaksimalkan $2\gamma = \frac{2}{\|\boldsymbol{w}\|_2}$, yang setara dengan meminimalkan $\frac{1}{2}\|\boldsymbol{w}\|_2^2$. Dengan menggunakan konstrain (2&3) pada $y_i\ (x_i \bullet \boldsymbol{w} + \mathrm{b}) \quad 1 \geq 0,$ pencarian bidang pemisah terbaik dengan γ terbesar :

$min \frac{1}{2}\|\boldsymbol{w}\|_2^2$ .....(6)

s.t. $y_i\ (x_i \bullet \boldsymbol{w} + \mathrm{b}) \quad 1 \geq 0,\ \forall_i$ .....(7)

yang merupakan masalah optimasi konstrain di mana akan diminimalkan fungsi tujuan (6) yang tergantung pada konstrain (7).

Formulasi optimasi konstrain diatas diminimalkan dengan pengganda Lagrange, yang dimulai dengan pembentukan fungsi Lagrange yang terdiri dari penjumlahan fungsi tujuan dan n konstrain yang dikalikan dengan masing-masing pengganda Lagrange, dan disebut formulasi primal (*primal formulation*) :

$L_p(\boldsymbol{w}, b, \alpha) \equiv \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \quad \sum_{i=1}^{n} \alpha_i(y_i\ (x_i \bullet \boldsymbol{w} + \mathrm{b}) \quad 1)$ .....(8)

Dimana $\alpha_i \geq 0$ adalah pengganda Lagrange.

Meminimumkan $L_p(\boldsymbol{w}, b, \alpha)$dengan mencari turunan $L(\boldsymbol{w}, b, \alpha)$ terhadap b and $\boldsymbol{w}$ dan menjadikan nol :

$\frac{\partial}{\partial b} L_p(\boldsymbol{w}, b, \alpha) = \quad \sum_{i=1}^{n} \alpha_i y_i = 0$ .....(9)

$\frac{\partial}{\partial \boldsymbol{w}} L_p(\boldsymbol{w}, b, \alpha) = \boldsymbol{w} \quad \sum_{i=1}^{n} \alpha_i y_i x_i = 0\boldsymbol{w} = \sum_{i=1}^{n} \alpha_i y_i x_i$.....(10)

Subtitusi $\boldsymbol{w}$ pada (10) ke $L_p(\boldsymbol{w}, b, \alpha)$ pada (8), diperoleh formulasi dual (*dual formulation*) dengan konstrain yang berbeda, yang dikenal sebagai Wolfe dual:

$\max_{\alpha} \ L_D = W(\alpha) = \sum_{i=1}^{n} \alpha_i$

$\frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \bullet x_j)$.....(11)

yang dimaksimalkan sehubungan dengan α_i, dan tergantung pada konstrain:

$\alpha_i \geq 0,$

$\sum_{i=1}^{n} \alpha_i y_i\ 0$.....(12)

Dual objective dalam (11) adalah bentuk kuadrat dalam parameter α_i, dan digambarkan sebagai masalah *quadratic programming* (QP). Selanjutnya konstrain tambahan (12), tersebut merupakan konstrain QP.

Karena $\min_{w,b} L_P = \max_{\alpha} L_D$, pencarian *separating hyperplane* terbaik dengan :

$\min_{w,b} L_P = \max_{\alpha} L_D \equiv \sum_{i=1}^{n} \alpha_i$

$\frac{1}{2}\sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i\, y_j\, x_i \bullet x_j$.....(13)

s.t. $\alpha_i \geq 0\ ,\ \sum_{i=1}^{n} \alpha_i y_i = 0$.....(14)

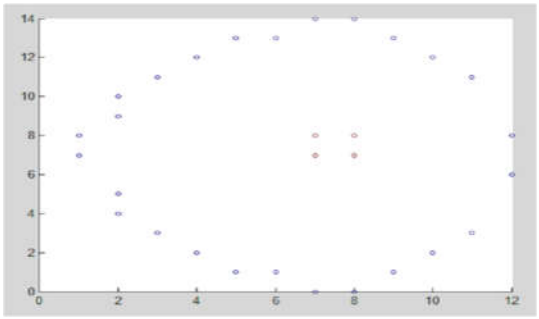
Untuk tiap data pelatihan terdapat nilai α_i dan solusinya kebanyakan α_i =0. Data pelatihan yang bernilai $\alpha_i > 0$ adalah SV (Campbell & Ying 2011). SV merupakan bagian *training set* yang paling informatif, yang mempengaruhi fungsi keputusan.

Data Nonlinear Separable

Untuk kasus *nonlinear separable*, perhitungan SVM dimodifikasi menjadi dua tahap, dengan memanfaatkan konsep *kernel trick* $\Phi(x_k)$. Pertama data $x_1, \ldots, x_n$ pada ^{d} yang bersifat *nonlinear separable* dipetakan terlebih dulu ke ruang berdimensi yang lebih tinggi ($> d$) , $\Phi :$ $^d \rightarrow$, : ruang fitur, sehingga *linearly separable*. Pada ruang yang berdimensi lebih tinggi, diharapkan data dapat lebih terstruktur dan mudah dipisahkan (Souza, 2010).

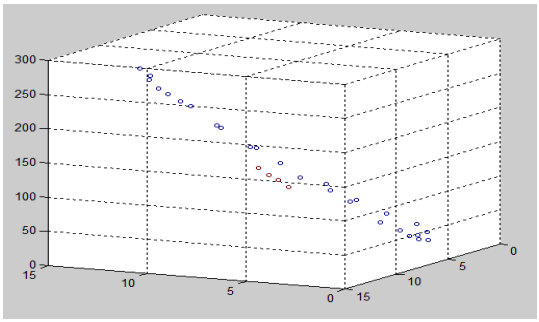
Pemetaan Φ

Pemetaan ke ruang fitur : Φ: $\boldsymbol{x} =$ ^{2} $\rightarrow$ $=$ ^{3}, $(x_1, x_2) \rightarrow (x_1, x_2, x_1^2 + x_2^2)$. Pemetaan yang dilakukan :Φ : $\boldsymbol{x} =$ ^{2} $\rightarrow$ $=$ ^{3}, $(x_1, x_2) \rightarrow (x_1, x_2, x_1^2 + x_2^2)$. Dataset dengan 2 atribut dan terdiri atas 2 kelas (positif dan negatif) berikut ini : {(1,7),(1,8),(2,4),(2,5),(2,9), (2,10),(3,3), (3,11)(4,2), (4,12),(5,1), (5,13), (6,1), (6,13),(7,14),(7,0), (7,8), (7,7), (8,4), (8,0),(8,8),(8,7), (9,1), (9,13), (10,2), (10,12),(11,3),(11,11),(12,6),(12,8)}. Plot data ke dalam ruang 2 dimensi, data tidak dapat dipisahkan secara linear :



Gambar 5. Plot Dua Dimensi Data Nonlinear Separable

Pemetaan yang dilakukan : Φ : $\boldsymbol{x} =$ $^2 \rightarrow$ $=$ ^{3}, $(x_1, x_2) \rightarrow (x_1, x_2, x_1^2 + x_2^2)$ yaitu $\Phi_1(\boldsymbol{x}) = (x_1)$, $\Phi_2(\boldsymbol{x}) = (x_2)$ dan $\Phi_3(\boldsymbol{x}) = (x_1^2 + x_2^2)$. Plot tiga dimensi hasil pemetaan sebagai berikut :



Gambar 6. Plot Tiga Dimensi Hasil Pemetaan Data Terpisahkan Secara Linear

Pada ruang fitur, selanjutnya dicari *optimal hyperplane* secara linear oleh SVM. SVM tetap bekerja sebagai *linear clasifier*, pada ruang dimensi baru yang lebih tinggi. Bentuk fungsional pemetaan $\Phi(x_i)$ tidak perlu diketahui, karena secara implisit telah didefinisikan:

$K(x_i, x_j) = \Phi(x_i) \bullet \Phi(x_j)$ pada $x_i \bullet x_j \rightarrow \Phi(x_i) \bullet \Phi(x_j)$.....(15)

Pada klasifikasi biner dengan kernelnya, pembelajaran memaksimalkan (11) pada ruang fitur, menjadi :$W(\alpha) = \sum_{i=1}^{n} \alpha_i$ $\frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$...(16) konstrain $\alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i y_i = 0$pada (12).

Untuk $y_j = +1$: $\min_{\{i|y_i = +1|\}} [\boldsymbol{w} \bullet \boldsymbol{x}_i + b] =$

$\min_{\{i|y_i = +1|\}} \left[\sum_{i,j=1}^{n} \alpha_i y_j K(x_i \bullet x_j)\right] + b =$ 1.....(17)

menggunakan $\boldsymbol{w} = \sum_{i=1}^{n} \alpha_i y_i x_i$ pada (10), dengan ekspresi yang sama untuk data berlabel $y_i =$ 1. Dari observasi ini diperoleh deduksi untuk *b*:

$b = -\frac{1}{2}\left(\max_{\{i|y_i = 1|\}} \left[\sum_{i,j=1}^{n} \alpha_i y_j K(x_i \bullet x_j)\right] + \min_{\{i|y_i = +1|\}} \left[\sum_{i,j=1}^{n} \alpha_i y_j K(x_i \bullet x_j)\right] \right)$ ....(18)

Untuk konstruksi SVM klasifikasi biner, tempatkan data (x_i, y_i) ke (16) dan memaksimalkan $W(\alpha)$ dengan konstrain (12). Dari nilai-nilai optimal α_i, yang ditunjukkan oleh α_i, akan dihitung bias b menggunakan (18). Untuk input baru vektor **z**, kelas yang diprediksi didasarkan pada tanda dari :

$\Phi(\boldsymbol{z}) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, \boldsymbol{z}) + b$ .. ...(19)

dimana b adalah nilai bias pada optimalitas. Ekspresi ini diperoleh dari substitusi $\boldsymbol{w} = \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i)$ dari (10) ke dalam fungsi keputusan (1), yaitu, $f(\mathbf{z}) = \text{sign}(\boldsymbol{w} \bullet \Phi(\mathbf{z}) + b)$

$= \text{sign} (\sum_{i=1}^{n} \alpha_i y_i K(x_i, \mathbf{z}) + b)$. .....(20)

Selanjutnya mengacu pada solusi $(\boldsymbol{\alpha}, \boldsymbol{b})$, sebagai hipotesis pemodelan data. Dari sudut pandang teori optimasi dan kondisi Karush-Kuhn-Tucker (KKT), $\alpha_i(y_i(\boldsymbol{w} \bullet x_i + b) \quad 1) = 0$ .....(21)

yang disimpulkan,baik $y_i(\boldsymbol{w} \bullet x_i + b)$ >1

(non SV) dan karena $\alpha_i = 0$ atau karena $y_i(\boldsymbol{w} \bullet x_i + b) = 1$ (SV) yang di-mungkinkan untuk $\alpha_i > 0$. Data dengan α_i besar berpengaruh besar pada orientasi hyperplane dan berpengaruh signifikan pada fungsi keputusan.

KESIMPULAN DAN SARAN

Simpulan

Untuk menyelesaikan kasus non linear pada klasterisasi, dapat digunakan pendekatan SVM, dimana *data space* untuk data awal yang kompleks dalam ruang berdimensi rendah dipetakan ke ruang fitur yang berdimensi tinggi melalui transformasi *Gaussian Kernel* (agar data lebih terstruktur dan mudah dipisahkan secara linear), dengan tahapan:

1. Data $x_1, x_2, \ldots, x_n$ pada ^{d}.
   Pemetaan Φ dari input space ke ruang berdimensi lebih tinggi (>d) .
2. Φ : $^{d} \rightarrow$, : *feature space,* batas klasifikasi linear pada merupakan batas klasifikasi non linear pada ^{d}.

Konstruksi SVM klasifikasi biner berusaha menemukan dengan α_i terbesar yang akan digunakan untuk membentuk hyperplane dengan margin optimal yang berpengaruh signifikan fungsi keputusan yang dinyatakan dengan fungsi tanda, yaitu menjadi anggota atau tidak :

$f(\mathbf{z}) = \text{sign}(\boldsymbol{w} \bullet \Phi(\mathbf{z}) + b)$

$= \text{sign} (\sum_{i=1}^{n} \alpha_i y_i K(x_i, \mathbf{z}) + b)$.

Saran

Deteksi pola sebaran data awal sangat disarankan sebelum klasterisasi dilakukan terhadap sejumlah obyek maupun terhadap sejumlah peubah. Bila data awal saling berkorelasi atau berpola non linear, klasterisasi dapat dilakukan dengan menggunakan teknik SVM atau melalui transformasi data ke ruang fitur yang berdimensi lebih tinggi dengan menggunakan transformasi kernel, diantaranya kernel Gaussian.

Bila dimensi data hasil transformasi terlalu tinggi, dapat dilakukan reduksi dimensi dengan menggunakan transformasi kernel lainnya seperti Kernel PCA (Cristianini and Taylor, 2000).

DAFTAR PUSTAKA

Cambbell, C. And Y. Ying. (2011). Learning With Support Vector Machines. Morgan & Claypool. 2011.

Cristianini, N and J.S. Taylor. (2000). An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press. New York.

Jain, A. K. (2010). 'Jain, Lansing - 2009 - Data Clustering 50 Years Beyond

K-Means 1 Anil K . Jain Michigan State University', 19th International Conference in Pattern Recognition (ICPR), pp. 651–666. doi: 10.1016/j.patrec.2009.09.011.

Johnson, R.A. and D.W. Wichern. (1988). Applied Multivariate Statistical Analysis. Prentice Hall. Englewood. New Jersey.

Lee, Y. and Choi, S. (2007). 'Minimum Entropy , k -Means , Spectral Clustering', Biometrics Technology Research, (1).

MacQueen, J. B. (1967). 'Kmeans Some Methods for classification and Analysis of Multivariate Observations', 5th Berkeley Symposium on Mathematical Statistics and Probability 1967, 1(233), pp. 281–297. doi: citeulike-article-id:6083430.

Russell, S. and Norvig, P. (2010). Artificial Inteligence: A Modern Approach. 3 rd. Prentice Hall Series.

Sakthi, M. and Thanamani, A. S. (2011). 'An Effective Determination of Initial Centroids in K-Means Clustering Using Kernel PCA', 2(3), pp. 955–959. Available at: http://www.ijcsit.com/docs/Volume 2/vol2issue3/ijcsit2011020303.pdf.

Santini. (2016). Advantages & Disadvantages of K-Means and Hierarchical clustering (Unsupervised Learning)', pp. 1-5.

Singh, N. and Singh, D. (2013). 'The Improved K-Means with Particle Swarm Optimization', Journal of Information Engineering and Applications, 3(11), pp. 1–8.

Souza, C. (2010). Kernel Functions for Machine Learning Applications. http:// crsouza.blogspot.com/2010/03/kernel-functions-for-machine-learning.html

Taylor, J. S. And N. Cristianini. (2004). Kernel Methods for Pattern Analysis. Cambridge University Press.New York