

Teknik Pengukuran: Metode Klasifikasi *Support Vector Machine* (SVM) pada Data Pengukuran

Eufrat Tsaqib Qasthari ¹

Departemen Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA)
Universitas Indonesia¹, Depok, 16424, Indonesia

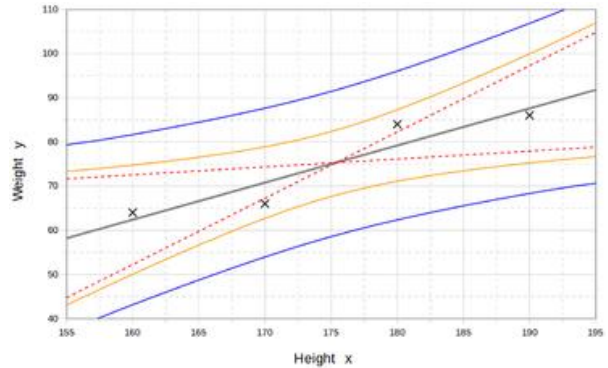
(eufrat.tsaqib@ui.ac.id¹)

1. Klasifikasi Data Pengukuran

Data pengukuran yang berbagai macam kelas atau kategorinya perlu dilakukan sebuah klasifikasi agar pengertian datanya dapat diekstraksi dengan mudah dan dapat dimengerti oleh pengamat. Proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui disebut dengan klasifikasi data. Metode klasifikasi yang beragam antara lain Naive Bayes, Fuzzy dan k-Nearest Neighbor. Pada klasifikasi data dapat dibagi ke 3-fase yaitu fase pelatihan (training phase) dimana model di bangun dari hasil pelatihan dari data, lalu fase pengujian (testing phase) dimana model akan diberikan beberapa tes sample yang diambil secara random pada data yang sudah ada namun bukan data pada *training phase* umumnya fase ini akan dipecah menjadi tahap validasi dan *testing*. Fase terakhir pada klasifikasi data adalah tahap penggunaan (*usage phase*) dimana permodelan akan digunakan untuk memprediksikan dan mengklasifikasi data yang belum diketahui polanya.

2. Regresi Data Pengukuran

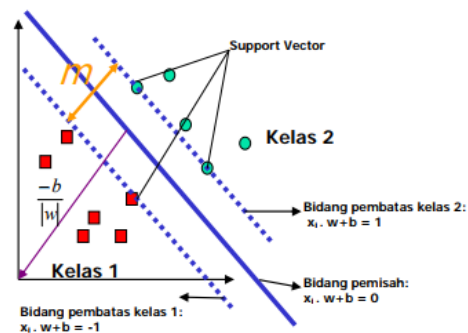
Pada permodelan statistik, analisis regresi bertujuan untuk mengestimasi hubungan antar variabel. Contohnya pada regresi linier yang mengestimasi hubungan antara variabel terikat dan bebas pada suatu data. Umumnya regresi akan membantu pengamat untuk melihat perubahan data pada variabel terikat pada setiap perubahan variabel bebas. Pada regresi data pengukuran akan dikenal jenis prediksi Intrapolasi dan Ekstrapolasi, prediksi Intrapolasi adalah metode prediksi berdasarkan data-data yang diperoleh sedangkan metode Ekstrapolasi adalah metode prediksi berdasarkan data diluar data-data yang diperoleh.



Pada gambar diatas diperlihatkan prediksi Intrapolasi dengan garis hitam lurus yang paling seimbang dari data-data yang ada sebelumnya. Sedangkan, garis merah adalah garis ekstrem. Lalu estimasi data diperlihatkan dari garis lengkung kuning pertama dan prediksi Ekstrapolasi dapat diperlihatkan oleh garis lengkung biru kedua.

3. Linier Support Vector Machine (SVM)

Support Vector Machine atau SVM adalah metode regresi atau pengklasifikasian data berdasarkan data-data sebelumnya dan permodelannya di supervisi terlebih dahulu. SVM termasuk kedalam jenis klasifikator yang biner, linier dan non probabilistik. SVM menggunakan decision boundary (batas keputusan) yang akan menentukan klasifikasi dari data-data pelatihan sehingga dapat dibentuk sebuah model linier atau hyperplane yang paling optimal untuk mengklasifikasikan data data tersebut.



SVM menggunakan model linier sebagai decision boundary dalam bentuk sebagai berikut:

$$y(x) = w^t \phi(x) + b$$

Dimana x adalah vektor input, w adalah parameter bobot, $\phi(x)$ adalah fungsi basis, dan b adalah suatu bias. Sehingga, sebuah *hyperplane* yaitu $y(x)$ dengan memaksimalkan margin didapatkan fungsinya. Margin maksimum adalah pilihan yang paling aman karena terjadi paling sedikit kesalahan pada klasifikasi. Apabila data dapat terklasifikasi secara linier maka data x ke n dapat dicari jaraknya dengan

$$\frac{t_n y(x_n)}{\|w\|} = \frac{t_n (w^t \phi(x_n) + b)}{\|w\|}$$

Margin adalah jarak decision boundary dengan data terdekat oleh karena itu memaksimalkan margin dapat menggunakan

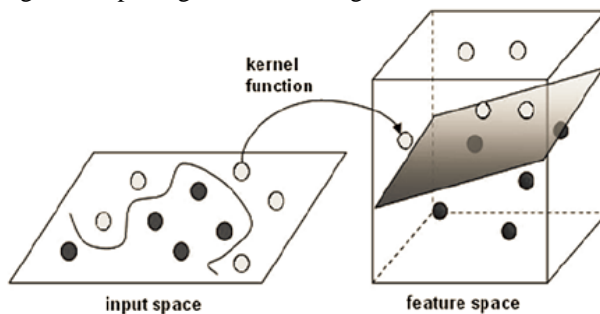
$$\operatorname{argmax} \left\{ \frac{1}{\|w\|} \min [t_n (w^t \phi(x_n) + b)] \right\}$$

4. Non-Linier Support Vector Machine (SVM)

Pada dunia nyata data-data pengukuran umumnya tidak dapat diklasifikasi dengan menggunakan metode SVM linier, namun dapat diklasifikasi dengan menggunakan SVM non linier. Oleh karena itu SVM dimodifikasi dengan memasukkan fungsi kernel. Fungsi kernel adalah fungsi k yang mana untuk semua vektor input x dan z akan memenuhi kondisi

$$k(x, z) = \phi(x)^t \phi(z)$$

Dimana ϕ adalah fungsi pemetaan dari ruang input ke ruang fitur. Penggambaran transformasi data dari ruang input ke ruang fitur dapat digambarkan sebagai berikut:



Pada umumnya kernel-kernel yang dapat dipakai pada SVM adalah sebagai berikut:

Tabel 1- Kernel yang umum dipakai dalam SVM

Jenis Kernel	Definisi
Polynomial	$K(\bar{x}_i, \bar{x}_j) = (\bar{x}_i \cdot \bar{x}_j + 1)^p$
Gaussian	$K(\bar{x}_i, \bar{x}_j) = \exp\left(-\frac{\ \bar{x}_i - \bar{x}_j\ ^2}{2\sigma^2}\right)$
Sigmoid	$K(\bar{x}_i, \bar{x}_j) = \tanh(\alpha \bar{x}_i \cdot \bar{x}_j + \beta)$

Gambar dibawah ini merupakan perbedaan antara linier SVM dan non-linier SVM. Selain kernel yang direkomendasikan untuk diuji pertama kali adalah kernel RBF (*Radial Basis Function*) karena memiliki kinerja yang sama dengan kernel linier dan memiliki perilaku sama seperti fungsi kernel *sigmoid* seperti pada tabel diatas.

5. Multiclass SVM

Pada umumnya *Support Vector Machine* hanya dapat mengklasifikasi dua kelas karena SVM termasuk ke klasifikator biner (dua jenis). Namun penelitian lebih lanjut dapat mengklasifikasi SVM menjadi lebih dari dua kelas. Ada dua cara untuk mengimplementasikan Multiclass SVM yaitu menggabungkan beberapa SVM biner.

Referensi

- <http://sutikno.blog.undip.ac.id/files/2011/11/tutorial-svm-bahasa-indonesia-oleh-krisantus.pdf> diakses 26 Desember 2017, 12.30 WIB
- <http://mti.binus.ac.id/2017/11/24/trik-kernel/> , diakses 26 Desember 2017, 12.30 WIB
- http://ocw.ui.ac.id/pluginfile.php/269/mod_resource/content/0/7.1%20Support%20Vector%20Machine.pdf diakses 26 Desember 2017, 12.36 WIB
- <http://asnugroho.net/papers/ikcsvm.pdf> diakses 25 Desember 2017, 12.40 WIB

