

Muhammad Harits

1103223159

Bab 5: Support Vector Machines (SVM)

Ringkasan Bab

Bab ini membahas algoritma Support Vector Machine (SVM) — salah satu metode yang sangat kuat dan serbaguna dalam Machine Learning, terutama sebelum era Deep Learning. SVM sangat efektif pada dataset dengan ukuran kecil hingga menengah yang memiliki pola kompleks.

Konsep Utama

1. Large Margin Classification

Inti dari SVM adalah mencari garis pemisah (decision boundary) antar kelas yang tidak hanya memisahkan data dengan benar, tetapi juga memiliki margin atau jarak pemisah yang selebar mungkin. Semakin lebar marginnya, semakin baik kemampuan model melakukan generalisasi ke data baru.

2. Support Vectors

Support vectors adalah titik-titik data yang posisinya berada tepat di tepi margin. Hanya titik-titik inilah yang menentukan batas keputusan model; data yang jauh dari margin tidak memengaruhi keputusan model sama sekali.

3. Hard Margin vs. Soft Margin

- **Hard Margin**
Mengharuskan seluruh data berada di luar margin tanpa ada pelanggaran. Pendekatan ini hanya cocok jika data benar-benar dapat dipisahkan secara linear dan tidak mengandung outlier.
- **Soft Margin**
Lebih fleksibel: memperbolehkan beberapa data memasuki margin atau salah klasifikasi jika hal tersebut dapat menghasilkan generalisasi yang lebih baik. Dikendalikan oleh hyperparameter C :
 - C kecil \rightarrow margin lebar (regularisasi tinggi)
 - C besar \rightarrow margin sempit (regularisasi rendah)

4. SVM Non-Linear dan Kernel Trick

Untuk data yang tidak terpisahkan secara linear, SVM menggunakan Kernel Trick — teknik yang secara matematis memproyeksikan data ke ruang berdimensi lebih tinggi sehingga menjadi linear separable, tanpa benar-benar melakukan perhitungan dimensi tinggi tersebut. Ini membuat SVM tetap efisien secara komputasi.

Beberapa kernel yang umum digunakan:

- Polynomial Kernel — cocok untuk data dengan pola hubungan polinomial
- Gaussian RBF Kernel — sangat fleksibel dan menjadi pilihan default untuk banyak kasus

5. SVM untuk Regresi (SVR)

Selain klasifikasi, SVM juga dapat digunakan untuk regresi. Pada regresi, konsep margin digunakan dengan tujuan yang berbeda: model mencoba memasukkan sebanyak mungkin titik data ke dalam margin. Lebar margin dikendalikan oleh hyperparameter ϵ (epsilon).