

Support Vector Machines

Bab 5 memperkenalkan Support Vector Machines (SVM) ¹, sebuah algoritma klasifikasi yang sangat kuat dan serbaguna². Konsep inti dari SVM adalah Large Margin Classification ³, yaitu memilih *hyperplane* yang memisahkan kelas-kelas data dengan margin terbesar untuk memastikan model memiliki generalisasi yang baik⁴.

Konsep Inti dan Klasifikasi Margin

Keputusan posisi *hyperplane* ini hanya dipengaruhi oleh beberapa titik data yang disebut *support vectors*⁵⁵. Titik-titik ini adalah yang berada paling dekat dengan batas keputusan (*decision boundary*)⁶⁶⁶. Karena hanya vektor-vektor ini yang penting, SVM tetap efektif bahkan pada *dataset* besar dengan dimensi fitur yang tinggi⁷.

Untuk data yang tidak dapat dipisahkan secara sempurna, SVM menggunakan Soft Margin Classification⁸. Metode ini memperbolehkan beberapa pelanggaran margin untuk meningkatkan fleksibilitas model⁹. Tingkat hukuman terhadap pelanggaran ini dikontrol oleh parameter C ¹⁰. Nilai C yang besar menghasilkan margin yang ketat, sedangkan nilai C yang kecil menghasilkan margin yang lebih lebar dengan toleransi *misclassification* yang lebih tinggi¹¹.

Menangani Data Non-Linier (*Kernel Trick*)

Ketika data tidak linier, SVM dapat menanganinya dengan menggunakan *kernel trick*¹². Teknik ini secara implisit mengubah data ke ruang berdimensi lebih tinggi tanpa perlu menghitung koordinat baru secara eksplisit¹³. Kernel-kernel umum seperti Polynomial dan RBF (Radial Basis Function) atau Gaussian memungkinkan model membuat batas keputusan non-linier yang kompleks¹⁴.

Kompleksitas batas keputusan pada RBF kernel dapat dikontrol lebih lanjut oleh parameter gamma (γ)¹⁵. Nilai gamma yang kecil menghasilkan batas yang lebih halus dan umum (*general*)¹⁶, sementara nilai gamma yang besar membuat batas keputusan sangat ketat mengikuti data, yang berpotensi menimbulkan *overfitting*¹⁷.

Regresi dengan SVM

Selain klasifikasi, SVM juga dapat digunakan untuk regresi melalui Support Vector Regression (SVR)¹⁸. Berbeda dengan *classifier* yang mencoba memaksimalkan margin, SVR mencoba menyesuaikan garis regresi yang tetap berada di dalam zona toleransi tertentu (disebut *epsilon-tube*) sambil meminimalkan penyimpangan titik data yang berada di luar zona tersebut¹⁹.