

Chapter 5 : Summary

Diedrick Darrell Darmadi - 1103223031

Bab 5 membahas konsep Support Vector Machines (SVM), sebuah algoritma klasifikasi yang sangat kuat untuk berbagai jenis data. Pembahasan dimulai dengan ide dasar *large margin classification*, yaitu memilih hyperplane yang memisahkan kelas dengan margin terbesar sehingga model memiliki generalisasi yang baik. Contoh dua kelas yang terpisah oleh garis dengan margin sempit versus margin lebar ditampilkan untuk menunjukkan dampak pemilihan hyperplane.

Model kemudian diperjelas dengan konsep *support vectors*, yaitu titik-titik data yang berada paling dekat dengan decision boundary dan menentukan posisi hyperplane. Karena hanya support vectors yang memengaruhi keputusan, SVM tetap efektif meskipun dataset besar namun memiliki dimensi fitur tinggi. Visualisasi titik-titik support vector diberi penanda khusus untuk menunjukkan peran pentingnya.

Setelah itu dijelaskan Soft Margin Classification, yaitu metode yang memperbolehkan beberapa pelanggaran margin agar model tetap fleksibel terhadap data yang tidak sepenuhnya dapat dipisahkan secara linear. Parameter c mengontrol seberapa besar model menghukum pelanggaran tersebut: nilai c besar menghasilkan margin kecil yang ketat, sedangkan nilai c kecil menghasilkan margin lebih lebar namun dengan toleransi lebih tinggi pada misclassifications. Grafik margin untuk berbagai nilai c diperlihatkan untuk memperjelas efek regulasi.

Ketika data tidak linear, SVM dapat menangani kasus tersebut menggunakan *kernel trick*, yaitu transformasi implisit ke ruang berdimensi lebih tinggi tanpa benar-benar menghitung koordinat baru. Kernel seperti Polynomial dan RBF (Gaussian) memungkinkan model membuat boundary non-linear yang kompleks. Visualisasi perbandingan decision boundary linear, polynomial, dan RBF ditampilkan untuk menunjukkan fleksibilitas kernel.

Bab ini juga membahas konsep kontrol kompleksitas melalui parameter gamma, terutama pada RBF kernel. Nilai gamma kecil menghasilkan boundary yang lebih halus dan general, sementara nilai besar membuat boundary mengikuti data secara ketat yang berpotensi menimbulkan overfitting. Contoh grafik untuk berbagai nilai gamma digunakan untuk menunjukkan perubahan bentuk decision boundary.

Selain klasifikasi, SVM juga dapat digunakan untuk regresi melalui *Support Vector Regression* (SVR). Berbeda dengan classifier, SVR mencoba menyesuaikan garis regresi yang tetap berada dalam zona toleransi tertentu (ϵ -tube) sambil meminimalkan penyimpangan titik yang berada di luar zona tersebut. Diagram ϵ -tube ditampilkan untuk menjelaskan prinsip kerja SVR.