

## Support Vector Machines

Bab 5 memperkenalkan Support Vector Machines (SVM)<sup>1</sup>, sebuah algoritma klasifikasi yang sangat kuat dan serbaguna<sup>2</sup>. Konsep inti dari SVM adalah Large Margin Classification<sup>3</sup>, yaitu memilih *hyperplane* yang memisahkan kelas-kelas data dengan margin terbesar untuk memastikan model memiliki generalisasi yang baik<sup>4</sup>.

### Konsep Inti dan Klasifikasi Margin

Keputusan posisi *hyperplane* ini hanya dipengaruhi oleh beberapa titik data yang disebut *support vectors*<sup>5</sup>. Titik-titik ini adalah yang berada paling dekat dengan batas keputusan (*decision boundary*)<sup>6</sup><sup>6</sup><sup>6</sup>. Karena hanya vektor-vektor ini yang penting, SVM tetap efektif bahkan pada *dataset* besar dengan dimensi fitur yang tinggi<sup>7</sup>.

Untuk data yang tidak dapat dipisahkan secara sempurna, SVM menggunakan Soft Margin Classification<sup>8</sup>. Metode ini memperbolehkan beberapa pelanggaran margin untuk meningkatkan fleksibilitas model<sup>9</sup>. Tingkat hukuman terhadap pelanggaran ini dikontrol oleh parameter C<sup>10</sup>. Nilai C yang besar menghasilkan margin yang ketat, sedangkan nilai C yang kecil menghasilkan margin yang lebih lebar dengan toleransi *miscalification* yang lebih tinggi<sup>11</sup>.

### Menangani Data Non-Linier (*Kernel Trick*)

Ketika data tidak linier, SVM dapat menanganinya dengan menggunakan *kernel trick*<sup>12</sup>. Teknik ini secara implisit mengubah data ke ruang berdimensi lebih tinggi tanpa perlu menghitung koordinat baru secara eksplisit<sup>13</sup>. Kernel-kernel umum seperti Polynomial dan RBF (Radial Basis Function) atau Gaussian memungkinkan model membuat batas keputusan non-linier yang kompleks<sup>14</sup>.

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

### Regresi dengan SVM

Selain klasifikasi, SVM juga dapat digunakan untuk regresi melalui Support Vector Regression (SVR)<sup>18</sup>. 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 tersebut<sup>19</sup>.