Optimal Hyperplanes

Linear Hard-Margin Support Vector Machine

Assumption: Training examples are linearly separable.

Optimization Problem (Primal):
\[ \max_{w,b} \frac{1}{2} ||w||^2 \]
\[ \text{s.t. } y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \ldots, n \]

Support Vectors: Examples with minimal distance (i.e. margin).

Non-Separable Training Data

Limitations of hard-margin formulation
- For some training data, there is no separating hyperplane.
- Complete separation (i.e. zero training error) can lead to suboptimal prediction error.

Soft-Margin Separation

Idea: Maximize margin and minimize training error.

Hard-Margin OP (Primal):
\[ \max_{w,b} \frac{1}{2} ||w||^2 \]
\[ \text{s.t. } y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \ldots, n \]

Soft-Margin OP (Primal):
\[ \max_{w,b} \frac{1}{2} ||w||^2 + C \sum_i \xi_i \]
\[ \text{s.t. } y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \]

- Slack variable \( \xi_i \) measures by how much \( y_i(x_i \cdot w + b) \) fails to achieve margin \( \delta \)
- \( \sum \xi_i \) is upper bound on number of training errors
- \( C \) is a parameter that controls trade-off between margin and training error.
Controlling Soft-Margin Separation

- $\sum \xi_i$ is an upper bound on the number of training errors.
- $C$ is a parameter that controls the trade-off between margin and training error.

**Soft-Margin OP (Primal):**

$$
\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum \xi_i \\
\text{subject to } y_i (w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0
$$

**Example Reuters “acq”: Varying C**

![Graph showing the relationship between C and training/testing errors.](image)

**Example: Margin in High-Dimension**

<table>
<thead>
<tr>
<th>Training Sample $S_{train}$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
<th>$x_6$</th>
<th>$x_7$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\langle e_1, y_1 \rangle$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\langle e_2, y_2 \rangle$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\langle e_3, y_3 \rangle$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>$\langle e_4, y_4 \rangle$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hyperplane 1</th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_3$</th>
<th>$w_4$</th>
<th>$w_5$</th>
<th>$w_6$</th>
<th>$w_7$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperplane 1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Hyperplane 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>Hyperplane 3</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hyperplane 4</td>
<td>0.5</td>
<td>-0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hyperplane 5</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hyperplane 6</td>
<td>0.95</td>
<td>-0.95</td>
<td>0</td>
<td>0.05</td>
<td>0.05</td>
<td>-0.05</td>
<td>-0.05</td>
<td>0</td>
</tr>
</tbody>
</table>