Soft-SVM Unconstrained Formulation

Unconstrained Formulation

Let us return to the objective:

$$\min_{\mathbf{w},b} \mathbf{w}^\top \mathbf{w} + C\sum_{i=1}^{n} \xi_i$$

which is optimized subject to the constraints for all i:

$$y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) \geq 1 - \xi_i$$
$$\xi_i \geq 0$$

The second part of the objective minimizes ξ_i as much as possible. For any given i there are two scenarios (assuming $C > 0$).

1. The point $\mathbf{x}_i$ lies on the correct side of the hyperplane, that is $y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) \geq 1$. In this case, the first constraint is automatically satisfied without ξ_i being positive. So, the objective will push it down to $\xi_i = 0$.
2. The point $\mathbf{x}_i$ does not lie on the correct side of the hyperplane, i.e. $y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) < 1$, in this case, we need $\xi_i > 0$ for the first constraint to hold. However, because the objective will try to minimize ξ_i as much as possible, it will set it to the smallest possible value that still satisfies the first constraint, which is $\xi_i = 1 - y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right)$. In other words, the first constraint will be satisfied as an *equality*.

It therefore follows that:

$$\xi_i = \begin{cases} 1 - y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) & \text{if } y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) < 1 \\ 0 & \text{if } y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) \geq 1 \end{cases}$$

These two cases are equivalent to the closed form: $\xi_i = \max\left[1 - y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right), 0\right]$. If we plug this closed form into the objective of our SVM optimization problem, we obtain the following *unconstrained* version as loss function and regularizer:

$$\min_{\mathbf{w},b} \underbrace{\mathbf{w}^\top \mathbf{w}}_{l_2-\text{regularizer}} + C\sum_{i=1}^{n} \underbrace{\max\left[1 - y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right), 0\right]}_{\text{hinge loss}}$$

This formulation allows us to optimize the SVM parameters $\mathbf{w}, b$ just like logistic regression (e.g. through gradient descent). The only difference is that we have the **hinge loss** instead of the **logistic loss**.