On the Uniqueness of the SVM Solution

Hard-Margin SVM

Recall that the hard-margin SVM problem is the following:

$$\begin{array}{lll} \text{minimize}_{w,b} & \|w\|_2^2 & \\ \text{subject to} & y_i(w^T x_i + b) \geq 1 & \text{for all } i = 1, \ldots, n. \end{array}$$

We prove the following theorem.

Theorem 1. *Let $(x_i, y_i) \in \mathbb{R}^d \times \{-1, +1\}$ for $i = 1, \ldots, n$ be our training data, and suppose there are $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that $y_i(w^T x_i + b) > 0$ for all i (i.e., linear separability). Furthermore, suppose there exist i, j with $y_i = +1$ and $y_j = -1$. Then there is a unique minimizer (w^*, b^*) to the hard-margin SVM problem.*

First we establish the following lemma.

Lemma 2. *Consider the optimization problem*

$$\begin{array}{ll} \textit{minimize}_{w \in \mathbb{R}^m, v \in \mathbb{R}^n} & f(w) + g(v) \\ \textit{subject to} & (w, v) \in S, \end{array}$$

where $S \subseteq \mathbb{R}^{m+n}$ is convex, f is strictly convex, and g is convex. If (w_1, v_1) and (w_2, v_2) are both minimizers then $w_1 = w_2$.

Proof. Suppose, for contradiction, that (w_1, v_1) and (w_2, v_2) are minimizers with $w_1 \neq w_2$. Since S is convex, the average $((w_1 + w_2)/2, (v_1 + v_2)/2)$ is also feasible. By strict convexity we have

$$f((w_1 + w_2)/2) < f(w_1)/2 + f(w_2)/2,$$

and by convexity we have

$$g((v_1 + v_2)/2) \leq g(v_1)/2 + g(v_2)/2.$$

Thus

$$f((w_1 + w_2)/2) + g((v_1 + v_2)/2) < \frac{f(w_1) + g(v_1)}{2} + \frac{f(w_2) + g(v_2)}{2} = f(w_1) + g(v_1),$$

with the last equality following since the two minimizers have equal objective values. This contradicts our assumption that (w_1, v_1) is a minimizer, and completes the proof. $\square$

Proof of Theorem 1. First we establish existence. Let w_L, b_L satisfy $y_i(w_L^T x_i + b_L) \geq \epsilon$ for all i and some $\epsilon > 0$ (such w_L, b_L must exist by linear separability). Then we have

$$y_i \left(\frac{w_L}{\epsilon}^T x_i + \frac{b_L}{\epsilon} \right) \geq 1.$$

This shows $(w_L/\epsilon, b_L/\epsilon)$ is in the feasible set. Thus any minimizer (w_*, b_*), if it exists, must have $\|w_*\|_2 \leq \|w_L\|_2/\epsilon$. Furthermore, if $\|w_*\| \leq \|w_L\|_2/\epsilon$ then note that

$$-y_i b \leq y_i w_*^T x_i - 1 \leq |y_i w_*^T x_i| + 1$$

implies that

$$b \leq 1 + \|w_*\|_2 \|x_i\|_2 \leq 1 + \|w_L\|_2 \|x_i\|_2/\epsilon$$

when $y_i = -1$ and

$$-b \leq 1 + \|w_*\|_2 \|x_i\|_2 \leq 1 + \|w_L\|_2 \|x_i\|_2/\epsilon$$

when $y_i = +1$. By assumption, both values of y_i appear in our data set. Thus we obtain

$$|b| \leq 1 + \|w_L\|_2 \max_i \|x_i\|_2/\epsilon.$$

This shows that we are optimizing a continuous function over a non-empty compact region, and thus must have a minimizer.

Next we prove uniqueness. Suppose (w_1, b_1) and (w_2, b_2) are both minimizers. By the lemma we have $w_1 = w_2$ using $f(w) = \|w\|_2^2$ and $g(b) = 0$. To prove $b_1 = b_2$ we use the following fact: at any minimizer (w_*, b_*) there must be i, j with $y_i = +1$, $y_j = -1$, $w_*^T x_i + b_* = 1$ and $w_*^T x_j + b_* = -1$. Geometrically, this says that there must be points from both classes lying on the margin boundaries. Note that this implies $b_1 = b_2$ since increasing b_* makes $w_*^T x_j + b_* > -1$ and decreasing b_* makes $w_*^T x_i + b_* < 1$. Thus what remains is to establish this geometric fact. To prove it, suppose all data points i with $y_i = +1$ have $w_*^T x_i + b > 1$ and let $m = \min_{y_i=+1} w_*^T x_i + b - 1$. Letting $\hat{w} = w_*/(1 + m/2)$ and $\hat{b} = (b_* - m/2)/(1 + m/2)$ we obtain a new feasible point with a strictly lower objective:

$$\hat{w}^T x_i + \hat{b} = \frac{w_*^T x_i + b_* - m/2}{1 + m/2} \geq \frac{1 + m/2}{1 + m/2} = 1 \qquad \text{(if } y_i = +1\text{)},$$
$$\hat{w}^T x_i + \hat{b} = \frac{w_*^T x_i + b_* - m/2}{1 + m/2} \leq \frac{-1 - m/2}{1 + m/2} = -1 \qquad \text{(if } y_i = -1\text{)}.$$

The same argument will apply if we swap the roles of $+1$ and -1, thus proving the geometric fact, and completing our proof. $\square$

Soft-Margin SVM

The soft-margin SVM problem is given by

$$\begin{array}{ll} \text{minimize}_{w,b,\xi} & \|w\|_2^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to} & y_i(w^T x_i + b) \geq 1 - \xi_i \qquad \text{for } i = 1, \ldots, n \\ & \xi_i \geq 0 \qquad \text{for } i = 1, \ldots, n. \end{array}$$

Here $C > 0$ is a given constant, and (x_i, y_i) are as in the hard-margin SVM, but not necessarily linearly separable. Applying the lemma with $f(w) = \|w\|_2^2$ and $g(\xi, b) = C\sum_{i=1}^n \xi_i$ we see that the minimizer w_* is uniquely determined. Unfortunately, b_* is not always uniquely determined. To see how this can happen, suppose

$$\left|\{i \mid y_i = +1 \quad \text{and} \quad y_i(w_*^T x_i + b_*) \leq 1\}\right| = \left|\{i \mid y_i = -1 \quad \text{and} \quad y_i(w_*^T x_i + b_*) < 1\}\right|.$$

Then we can slightly decrease b_* while keeping $\sum_{i=1}^n \xi_i$ constant. This is analogous to the lack of uniqueness that can occur when proving the conditional median minimizes the absolute difference loss. For more, see [1], [2].

References

[1] C. Burges and D. Crisp. Uniqueness of the SVM Solution. NIPS. Vol. 99. 1999.

[2] R. Rifkin, P. Massimiliano, and A. Verri. A note on support vector machine degeneracy. International Conference on Algorithmic Learning Theory. Springer Berlin Heidelberg, 1999.