SVM: RETRAINING WITH JUST THE SUPPORT VECTORS?

DAVID S. ROSENBERG

1. Setup

Consider the following formulation of the SVM objective function:

$$J(w) = \sum_{i=1}^{n} \ell(w^T x_i, y_i) + \lambda \|w\|^2,$$

for $\lambda > 0$ and where the loss function is the hinge loss $\ell(\hat{y}, y) = (1 - \hat{y} y_i)_+$, where $(x)_+ = x1(x \geq 0)$ refers to the "positive part" of x. This differs from our usual objective $J'(w) = \frac{1}{2}||w||^2 + \frac{c}{n} \sum_{i=1}^{n} \ell(w^T x_i, y_i)$, but the two will produce the same set of solutions as we vary the hyperparameters $\lambda, c \in (0, \infty)$.

We know from the duality theory of SVMs that the minimizer of $J(w)$ can be written as

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i,$$

where some subset of the α_i^*'s may be exactly 0. One natural question is, what happens if we drop all the training points corresponding to $\alpha_i^* = 0$ from the objective function and re-fit the model. Does the solution change?

We can't quite show that, but here we show something a bit weaker: if we drop all training points that are on the "good side of the margin", then the solution does not change. In other words, we can drop all training points for which $y_i x_i^T w^* > 1$. This does not include all points for which $\alpha_i^* = 0$, since for those points we only know that $y_i x_i^T w^* \geq 1$. Here's the demonstration of the weaker statement:

Without loss of generality, index the x_i's so $x_{m+1}, \ldots, x_n$ are all on the "good side of the margin". Then we know that $\alpha_{m+1}^*, \ldots, \alpha_n^* = 0$. Let's define

$$J_1(w) = \sum_{i=1}^{m} \ell(w^T x_i, y_i) + \lambda \|w\|^2$$

and let

$$J_2(w) = \sum_{m+1}^{n} \ell(w^T x_i, y_i).$$

The claim is that w^* is also the minimizer of $J_1(w)$. We'll do this with a local analysis of J and J_1 around w^*. The relation $y_i x_i^T w^* > 1$ holds for each $i = m+1, \ldots, n$. Moreover, since $y_i x_i^T w$ is a continuous function of w for each i, these inequality relations will also hold for w in an ε-ball around w^*, for small enough $\varepsilon > 0$. Thus in that ball, $\ell(w^T x_i, y_i) = \left(1 - y_i w^T x_i\right)_+ = 0$, and so $J_2(w) \equiv 0$ for $\|w - w^*\| < \varepsilon$. Thus in that ball, $J_1(w) = J(w)$, and so w^* is a local minimizer of $J_1(w)$. By convexity of $J_1(w)$, w^* is also a global minimizer of J_1, and so the solution is unchanged by dropping the training points on the good side of the margin.

Date: October 31, 2021.

This argument fails if we only have $y_i x_i^T w^* \geq 1$ for some i, and so we have not shown that we can just drop all training points with $\alpha_i^* = 0$.