Jason J. Corso, jcorso@buffalo.edu, SUNY Buffalo
Solutions by David Johnson

Solutions

Problem 1: Recall (2pts) (Answer in one sentence only.)

What is a support vector?

A support vector is a point that lies (approximately) on the margin of an SVM solution. Support vectors are characterized by having α values greater than 0, and essentially define the SVM problem because they determine the margin and are the most difficult points to classify correctly.

Problem 2: Work (8 pts) (Show all derivations/work and explain.) Consider the standard unbiased SVM objective formulation:

$$\min_a \frac{1}{2}||a||^2 \qquad s.t. \frac{z_k a^T y_k}{||a||} \geq b, \forall k \ ,$$

A. What do the variables a, $||a||$, b, z and y represent?

$a =$ the weight vector
$||a||^2 =$ the L2 norm of the weight vector, used to measure its magnitude
$b =$ the margin constraint
$z_k =$ the class label (+1 or -1) of point k
$y_k =$ the data vector of point k

SEE NEXT PAGE

B. Show and explain mathematically why the goal of the SVM is to minimize $\frac{1}{2}||a||^2$.

(Hint 1: The a that satisfies the constraints and yields the minimum value for $||a||^2$ also yields the minimum value for $||a||$.)

(Hint 2: Remember that a can be scaled arbitrarily.)

(Hint 3: The margin to either side of the decision hyperplane can be represented by a pair of parallel hyperplanes defined by the equation $a^T y = \pm b$. How would you compute the distance between them along the axis defined by a?)

The two margin hyperplanes defined by $a^T y = \pm b$ are separated by exactly $2b$ on the axis defined by a. However, the goal of the SVM is to maximize the margin—in other words, to maximize the distance between these two margin hyperplanes ***in the original input space***. Recalling that a (and thus the axis defined by a) can be rescaled arbitrarily, we can see that the true distance between these two planes in the original space is not $2b$, but $\frac{2b}{||a||}$ (i.e. $2b$ normalized by the magnitude of a, as measured via the standard L2 norm).

Our goal, then, is to maximize $\frac{2b}{||a||}$ (subject to the constraints defined by our data), and since b is just an arbitrary input value, this means our optimization problem reduces to minimizing $\frac{1}{2}||a||$. This could be computationally expensive, however, because of the square root needed to compute $||a||$, so we simply substitute $||a||^2$, which will yield the same optimal value of a.

Alternately

If we take b itself to be the margin in the original input space, rather than an arbitrary input, then we must impose a constraint $||a||b = 1$, in order to define the scaling of a (which could still be scaled arbitrarily, otherwise). In this case, it is clear that $b = \frac{1}{||a||}$, so maximizing b is equivalent to minimizing $||a||$, so long as this constraint is enforced.