EECS 127/227AT 11/6/2018
S. Sojoudi, L. El Ghaoui

Quiz

Let $X = [x_1 \ldots, x_m] \in \mathbb{R}^{n \times m}$ represent a matrix of data points, and $y \in \{-1, 1\}^m$ a vector containing labels assigned to them. The Support Vector Machine (SVM) problem has the form

$$\min_{(w,b)} \; L(w, b, X, y) := \sum_{i=1}^{m} (1 - y_i(x_i^T w + b))_+, \tag{1}$$

with variables $w \in \mathbb{R}^n$, $b \in \mathbb{R}$, which corresponds to the classification rule $y = \mathbf{sgn}(w^T x + b)$. Due to over-fitting issues, the problem is often modified to include penalty terms, in order to control for test set errors. In this exercise, we consider robust counterparts to the above.

1. In a first robust version of the problem, we assume that there is a norm-bounded, measurement-wise additive uncertainty around each data point:

$$\mathcal{X} = \{X + \Delta \; : \; \Delta = [\delta_1, \ldots, \delta_m], \;\; \|\delta_i\| \leq \rho, \;\; 1 \leq i \leq m\}, \tag{2}$$

where $\rho > 0$ is a parameter, and $\| \cdot \|$ is a norm. Show that the corresponding robust counterpart to problem (1) can be expressed as

$$\min_{(w,b)} \sum_{i=1}^{m} (1 - y_i(x_i^T w + b) + \rho \|w\|_*)_+,$$

where $\cdot_{-*}$ is the dual norm.

2. We now consider the case when the feature vectors are Boolean, that is, $X \in \{0,1\}^{n\times m}$, and would like to exploit this information, via an uncertainty model that only allows for Boolean matrices. Specifically, we consider the case where the perturbed data points are obtained from the nominal ones by "flipping" at most k of the n entries, from 0 to 1 or vice-versa. Here, $k \in \{1,\ldots,n\}$ is a given integer. For $x \in \{0,1\}^n$, we define

$$\mathcal{D}(x) := \{\delta \ : \ \delta + x \in \{0,1\}^n, \ \ \|\delta\|_1 \leq k\}\,,$$

We assume that all the feature vectors are independently perturbed. Explain why using the following uncertainty set in lieu of (2):

$$\mathcal{X} = \{X + \Delta \ : \ \Delta = [\delta_1,\ldots,\delta_m], \ \ \delta_i \in \mathcal{D}(x_i), \ \ 1 \leq i \leq m\}$$

is consistent with our assumptions.

3. Show that the worst-case loss now expresses as

$$\max_{X\in\mathcal{X}} \ L(w,b,X,y) = \sum_{i=1}^{m}(1 - y_i(x_i^T w + b) + \phi(w,x_i,y_i))_+,$$

where, for given w, x, y:

$$\phi(w,x,y) := \max_{\delta\in\mathbf{Co}\mathcal{D}(x)} \ yw^T\delta,$$

where you can assume without proof that the convex hull of $\mathcal{D}(x)$, $\mathbf{Co}\mathcal{D}(x)$, is given by

$$\mathbf{Co}\mathcal{D}(x) := \{\delta \ : \ 0 \leq \delta + x \leq \mathbf{1}, \ \ \|\delta\|_1 \leq k\}\,.$$

4. Show that the robust counterpart for Boolean perturbations is given by

$$\min_{w,b,(p_i,q_i)_{i=1}^m} \ \sum_{i=1}^{m}(1 - y_i(x_i^T w + b) + k\|y_i w - p_i + q_i\|_\infty + p_i^T(\mathbf{1} - x_i) + q_i^T x_i)_+.$$

5. What is the effect of both the l_1- and l_∞-norms in the above model? Discuss the cases $k = 1$, $k = n$.