1 Nonsmooth Convex Optimization

We are interested in unconstrained minimization of convex, possibly nondifferentiable, $f : \mathbb{R}^n \to \mathbb{R}$

$$\text{minimize}_{x \in \mathbb{R}^n} \ f(x)$$

given first order oracle

1.1 Subgradient Method

Given bounded subgradient $\|g^k\| \leq G$ and bounded domain $\|x^0 - x^*\| \leq R$, subgradient method is in a sense optimal as it achieves the lower bound $\mathcal{O}(\frac{1}{\epsilon^2})$ for this problem class. Subgradient method iteratively updates as follows

$$x^{k+1} = x^k - \alpha_k g^k$$

where $g^k \in \partial f(x^k)$ is *any* subgradient of f. First order optimality condition is now $0 \in \partial f(x^*)$, which is impossible to test for nontrivial function f. Therefore, using $\|g^k\| \leq \epsilon$ is not informative and subgradient method does not really have a stopping criterion.

1.1.1 Solving Support Vector Machine w/ Subgradient Method

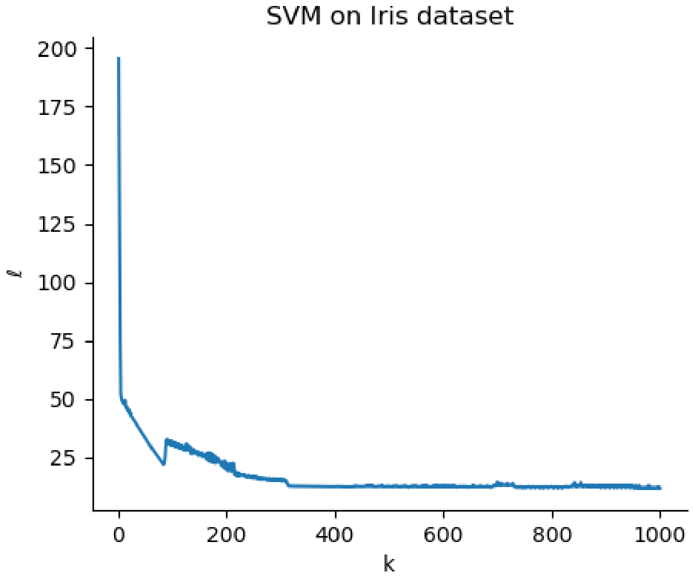
We are given data $\mathcal{D} = \{(x_i, y_i) \mid x_i \in \mathbb{R}^n \ y_i \in \{\pm 1\}\}$, support vector machine is supervised learning model that tries to find $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that the empirical risk and regularizer on w is minimized

$$\text{minimize}_{w,b} \quad \frac{1}{2}\|w\|_2^2 + \lambda \sum_{i=1}^{m} \max\left[0, 1 - y_i(w^T x_i + b)\right] \quad (:= f(w,b))$$

Support vector machines can be solved using subgradient method. We first find a subgradient of f

$$\begin{aligned} g_w^k &= w^k - \lambda \sum_{i \in [m]: y_i(w^T x_i + b) < 1} y_i x_i \\ g_b &= -\lambda \sum_{i \in [m]: y_i(w^T x_i + b) < 1} y_i \end{aligned}$$

where we haved picked $0 \in \partial(\max 0, 1 - y_i(w^T x_i + b))$ when $y_i(w^T x_i + b) = 1$, the only case where the *max term* is non-differentiable. When tested on the Iris dataset, subgradient method worked!


SVM on Iris dataset
ℓ
200
175
150
125
100
75
50
25
0
200
400
600
800
1000
k