School of Electrical and Computer Engineering

Technical University of Crete

Course: Optimization 2022-2023

Exercise 4

Support Vector Machines for Linearly Separable Data

Author: A. P. Liavas

December 20, 2022

In this exercise, we study the application of Support Vector Machines (SVMs) to a simple binary classification problem.

1 Binary classification with linearly separable data

We consider the linearly separable case. That is, we assume that there exist $\mathbf{0} \neq \mathbf{a}_* \in \mathbb{R}^n$ and $b_* \in \mathbb{R}$, such that the hyperplane

$$\mathbb{H}_{\mathbf{a}_*, b_*} := \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{a}_*^T \mathbf{x} = b_*\} \tag{1}$$

separates the data. Thus, for each data pair $(\mathbf{x}_i, y_i)$, for $i = 1, \ldots, N$, we have

$$y_i = \begin{cases} +1, & \text{if } \mathbf{a}_*^T \mathbf{x}_i \geq b_*, \\ +1, & \text{if } \mathbf{a}_*^T \mathbf{x}_i < b_*. \end{cases} \tag{2}$$

2 Distance of a point from a hyperplane

We have proved that the distance of a point $\mathbf{x}$ from the hyperplane $\mathbb{H}_{\mathbf{a},b}$ is given by

$$\mathcal{D}(\mathbf{x}, \mathbb{H}_{\mathbf{a},b}) = \frac{|\mathbf{a}^T \mathbf{x} - b|}{\|\mathbf{a}\|_2}. \tag{3}$$

Thus, the margin, with respect to $\mathbb{H}_{\mathbf{a},b}$, is

$$\min_{i=1\ldots,n} \frac{|\mathbf{a}^T\mathbf{x}_i - b|}{\|\mathbf{a}\|_2}. \tag{4}$$

3 The SVM problem

We consider the problem

$$\begin{aligned} \max_{\mathbf{a},b} & \left\{ \min_{i=1\ldots,n} \frac{|\mathbf{a}^T\mathbf{x}_i - b|}{\|\mathbf{a}\|_2} \right\} \\ \text{s.t.} \quad & \mathbf{a}^T\mathbf{x}_i - b \geq 0, \text{ if } y_i = +1, \\ & \mathbf{a}^T\mathbf{x}_i - b < 0, \text{ if } y_i = -1, \end{aligned} \tag{5}$$

This problem is non-convex. We shall develop a convex reformulation, as follows.

First, we note that if $(\mathbf{a}, b)$ is an optimal solution of (5) and $\alpha \neq 0$, then $(\alpha\mathbf{a}, \alpha b)$ is also optimal for (5) (why?). We can therefore set

$$\min_{i=1\ldots,n} |\mathbf{a}^T\mathbf{x}_i - b| = 1. \tag{6}$$

Then, the problem becomes

$$\begin{aligned} \max_{\mathbf{a},b} & \ \frac{1}{\|\mathbf{a}\|_2} \\ \text{s.t.} \quad & \min_{i=1\ldots,n} |\mathbf{a}^T\mathbf{x}_i - b| = 1, \\ & \mathbf{a}^T\mathbf{x}_i - b \geq 0, \text{ if } y_i = +1, \\ & \mathbf{a}^T\mathbf{x}_i - b < 0, \text{ if } y_i = -1. \end{aligned} \tag{7}$$

This reformulation is equivalent to

$$\begin{aligned} \min_{\mathbf{a},b} \quad & \|\mathbf{a}\|_2^2 \\ \text{s.t.} \quad & \min_{i=1\ldots,n} |\mathbf{a}^T\mathbf{x}_i - b| = 1, \\ & \mathbf{a}^T\mathbf{x}_i - b \geq +1, \text{ if } y_i = +1, \\ & \mathbf{a}^T\mathbf{x}_i - b < -1, \text{ if } y_i = -1, \end{aligned} \tag{8}$$

and, finally (why?), to

$$\begin{aligned} \min_{\mathbf{a},b} \quad & \|\mathbf{a}\|_2^2 \\ \text{s.t.} \quad & \mathbf{a}^T\mathbf{x}_i - b \geq +1, \text{ if } y_i = +1, \\ & \mathbf{a}^T\mathbf{x}_i - b < -1, \text{ if } y_i = -1. \end{aligned} \tag{9}$$

4 A simplified variation of the SVM problem

We use `matlab` notation and define $\mathbf{w} := [b; \mathbf{a}]$. Furthermore, we augment the data as $\bar{\mathbf{x}}_i := [-1; \mathbf{x}_i]$. Then, the separation equation is given by

$$\mathbf{w}^T \bar{\mathbf{x}}_i \gtreqless 0. \tag{10}$$

A slight variation of the SVM problem (9) is[1]

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|_2^2, \quad \text{subject to } y_i \mathbf{w}^T \bar{\mathbf{x}}_i \geq 1,\ i = 1, \dots, n. \tag{11}$$

Problem (11) is equivalent to

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|_2^2, \quad \text{subject to } 1 - y_i \mathbf{w}^T \bar{\mathbf{x}}_i \leq 0,\ i = 1, \dots, n. \tag{12}$$

We define

$$f_0(\mathbf{w}) := \frac{1}{2} \|\mathbf{w}\|_2^2, \tag{13}$$

and, for $i = 1, \dots, n$,

$$f_i(\mathbf{w}) = 1 - y_i \mathbf{w}^T \bar{\mathbf{x}}_i. \tag{14}$$

Thus, the problem (11) can be expressed as

$$\min_{\mathbf{w}} f_0(\mathbf{w}), \quad \text{subject to } f_i(\mathbf{w}) \leq 0,\ i = 1, \dots, n. \tag{15}$$

We have

$$\nabla f_0(\mathbf{w}) = \mathbf{w}, \quad \nabla^2 f_0(\mathbf{w}) = \mathbf{I}, \tag{16}$$

and

$$\nabla f_i(\mathbf{w}) = -y_i \bar{\mathbf{x}}_i, \quad \nabla^2 f_i(\mathbf{w}) = \mathbf{O}, \quad i = 1, \dots, n. \tag{17}$$

5 Barrier function for the simplified SVM problem

In order to be able to apply the barrier method, we define the barrier function

$$\phi(\mathbf{w}) := -\sum_{i=1}^{n} \log(-f_i(\mathbf{w})). \tag{18}$$

[1]Note that, in (11), we regularize b, which does not happen in (9). Usually, this has a small impact on the solution.

We have

$$\nabla \phi(\mathbf{w}) = \sum_{i=1}^{n} \frac{1}{-f_i(\mathbf{w})} \nabla f_i(\mathbf{w}) = \sum_{i=1}^{n} \frac{-y_i}{y_i \bar{\mathbf{x}}_i^T \mathbf{w} - 1} \bar{\mathbf{x}}_i = \sum_{i=1}^{n} \frac{y_i}{1 - y_i \bar{\mathbf{x}}_i^T \mathbf{w}} \bar{\mathbf{x}}_i, \tag{19}$$

and

$$\nabla^2 \phi(\mathbf{w}) = \sum_{i=1}^{n} \frac{1}{f_i^2(\mathbf{w})} \nabla f_i(\mathbf{w}) \nabla f_i(\mathbf{w})^T = \sum_{i=1}^{n} \frac{1}{(y_i \bar{\mathbf{x}}_i^T \mathbf{w} - 1)^2} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T. \tag{20}$$

6 Barrier method for the simplified SVM problem

We shall use the Newton method and minimize the cost functions

$$g_k(\mathbf{w}) := t_k \, f_0(\mathbf{w}) + \phi(\mathbf{w}), \tag{21}$$

for an increasing sequence of coefficients t_k, for $k = 1, \dots, K$.

We have

$$\nabla g_k(\mathbf{w}) = t_k \, \nabla f_0(\mathbf{w}) + \nabla \phi(\mathbf{w}) = t_k \, \mathbf{w} + \sum_{i=1}^{n} \frac{y_i}{1 - y_i \bar{\mathbf{x}}_i^T \mathbf{w}} \bar{\mathbf{x}}_i, \tag{22}$$

and

$$\nabla^2 g_k(\mathbf{w}) = t_k \, \nabla^2 f_0(\mathbf{w}) + \nabla^2 \phi(\mathbf{w}) = t_k \, \mathbf{I} + \sum_{i=1}^{n} \frac{1}{(y_i \bar{\mathbf{x}}_i^T \mathbf{w} - 1)^2} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T. \tag{23}$$

6.1 Newton step

The Newton step for the minimisation of g_k is given by

$$\mathbf{w}^+ = \mathbf{w} - s \left(\nabla^2 g_k(\mathbf{w}) \right)^{-1} \nabla g_k(\mathbf{w}), \tag{24}$$

where s is computed using backtracking for the following reasons:

1. first, we must guarantee that the new point is feasible, that is, $\mathbf{w}^+ \in \mathbf{dom}\, \phi$;

2. second, we must perform backtracking line search for the function g_k.

7 Exercise

In the exercise, I provide the following `matlab` files:

1. `separable_SVMs_interior_point_method.m` (main file)

2. `[X, y] = generate_data(N, a, b)` (data generating function)

3. `flag = data_is_not_separable(X, y, a, b)` (returns 0 if data is separable with respect to $\mathbf{a}$ and b, and 1 otherwise).

You are asked to write the following functions:

1. `flag = point_is_feasible(w_init, X_augm, y)`, which returns 1 if `w_init` $\in \mathbf{dom}\, \phi$ and 0 otherwise.

2. `grad = gradient_SVM_barrier(w(:,inner_iter), X_augm, y, t)`, which computes the gradient of the cost function g_k at the point `w(:,inner_iter)`.

3. `Hess = Hess_SVM_barrier(w(:,inner_iter), X_augm, y, t)`, which computes the Hessian of the cost function g_k at the point `w(:,inner_iter)`.

4. `barrier_SVM_cost_function(w_new, X_augm, y, t)`, which computes the value of g_k at the point `w_new`.