EE-556 Homework 1

Edoardo Debenedetti

October 26, 2019

Problem 1 - Geometric properties of the objective function f

Assuming $\mu = 0$, the objective function f is the smooth Hinge loss function, defined as:

$$f(x) = \ell_{sh}(\mathbf{x}) + \frac{\lambda}{2}\|\mathbf{x}\|^2 \tag{1}$$

where

$$\ell_{sh} = \frac{1}{n}\sum_{i=1}^{n} g_i(\mathbf{x}) \tag{2}$$

and

$$g_i(\mathbf{x}) = \begin{cases} \frac{1}{2} - b_i(\mathbf{a}_i^T\mathbf{x}) & b_i(\mathbf{a}_i^T\mathbf{x}) < 0 \\ \frac{1}{2}(1 - b_i(\mathbf{a}_i^T\mathbf{x}))^2 & 0 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \leq 1 \\ 0 & 1 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \end{cases} \tag{3}$$

(a) Gradient of f

Proof. Since the gradient is a linear operator:

$$\nabla f(\mathbf{x}) = \nabla \ell_{sh} + \nabla \frac{\lambda}{2}\|\mathbf{x}\|^2 \tag{4}$$

We can first compute $\nabla \frac{\lambda}{2}\|\mathbf{x}\|$:

$$\nabla\frac{\lambda}{2}\|\mathbf{x}\|^2 = \frac{\lambda}{2}\nabla\|\mathbf{x}\|^2 = \frac{\lambda}{2}\nabla\sum_{i=1}^{n}|x_i|^2 = \frac{\lambda}{2}\sum_{i=1}^{n}\nabla|x_i|^2 =$$
$$= \frac{\lambda}{2}2\mathbf{x} = \lambda\mathbf{x} \tag{5}$$

Now, let us compute $\nabla\ell_{sh}$:

$$\nabla\ell_{sh} = \nabla\frac{1}{n}\sum_{i=1}^{n}g_i(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\nabla g_i(\mathbf{x}) \tag{6}$$

Where $\nabla g_i(\mathbf{x})$ is the gradient of (3)

$$\nabla g_i(\mathbf{x}) = \begin{cases} \nabla\left[\frac{1}{2} - b_i(\mathbf{a}_i^T\mathbf{x})\right] & b_i(\mathbf{a}_i^T\mathbf{x}) < 0 \\ \nabla\left[\frac{1}{2}(1 - b_i(\mathbf{a}_i^T\mathbf{x}))^2\right] & 0 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \leq 1 \\ \nabla 0 & 1 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \end{cases} \tag{7}$$

The case where $1 \leq b_i(\mathbf{a}_i^T\mathbf{x})$ is trivial, since

$$\nabla 0 = 0 \tag{8}$$

In the case where $b_i(\mathbf{a}_i^T\mathbf{x}) < 0$:

$$\nabla\left[\frac{1}{2} - b_i(\mathbf{a}_i^T\mathbf{x})\right] = \nabla - b_i(\mathbf{a}_i^T\mathbf{x}) = -b_i\mathbf{a}_i \tag{9}$$

Next, in the case where $0 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \leq 1$:

$$\nabla\left[\frac{1}{2}(1 - b_i(\mathbf{a}_i^T\mathbf{x}))^2\right] = -\frac{1}{2}2b_i\mathbf{a}_i(1 - b_i(\mathbf{a}_i^T\mathbf{x})) =$$
$$-b_i\mathbf{a}_i(1 - b_i(\mathbf{a}_i^T\mathbf{x})) = b_i\mathbf{a}_i(b_i(\mathbf{a}_i^T\mathbf{x}) - 1) \tag{10}$$

Finally, combining (8), (9) and (10), we get:

$$\nabla g_i(\mathbf{x}) = \begin{cases} -b_i\mathbf{a}_i & b_i(\mathbf{a}_i^T\mathbf{x}) < 0 \\ b_i\mathbf{a}_i(b_i(\mathbf{a}_i^T\mathbf{x}) - 1) & 0 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \leq 1 \\ 0 & 1 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \end{cases} \tag{11}$$

Now, let us define, as in the problem statement, $\tilde{\mathbf{A}} := [b_1\mathbf{a}_1, ..., b_n\mathbf{a}_n]^T$, and $\mathbf{I}_L, \mathbf{I}_Q$ as the diagonal $n \times n$ matrices such that $\mathbf{I}_L(i,i) = 1$ if $b_i(\mathbf{a}_i^T\mathbf{x}) < 0$ and $\mathbf{I}_Q(i,i) = 1$ if $0 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \leq 1$, and 0 otherwise.

We can observe that $\tilde{\mathbf{A}}^T\mathbf{I}$ is the matrix whose i-th column is $b_i\mathbf{a}_i$. Instead, $\tilde{\mathbf{A}}^T\mathbf{I}_L$'s i-th columns will be non-zero only in the case where $b_i(\mathbf{a}_i^T\mathbf{x}) < 0$. Then it is possible to represent this case of $\nabla g_i(\mathbf{x})$ where $b_i(\mathbf{a}_i^T\mathbf{x}) < 0$ as

$$-\frac{1}{n}\tilde{\mathbf{A}}^T\mathbf{I}_L\mathbf{1} \tag{12}$$

since multiplying $\tilde{\mathbf{A}}^T\mathbf{I}_L$ by $\mathbf{1}$ will give as result the vector containing the sum of the elements of each column, which means the element-wise sum of the different j-th components of the i-th gradients relative to each $g_i(\mathbf{x})$. Each j-th component can be written as

$$\sum_{i\in\{i|b_i(\mathbf{a}_i^T\mathbf{x})<0\}}^{n} a_{i,j}b_j$$

In a similar fashion, $\tilde{\mathbf{A}}^T\mathbf{I}_Q$ is the matrix whose i-th column is $\mathbf{a}_ib_i$ only if i is such that $0 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \leq 1$. Moreover, $\tilde{\mathbf{A}}\mathbf{x}$ is the vector such that $[\tilde{\mathbf{A}}\mathbf{x}]_n = \sum_{i=1}^n b_i\mathbf{a}_i\mathbf{x}$. Consequently, $\tilde{\mathbf{A}}\mathbf{I}_Q[\tilde{\mathbf{A}}\mathbf{x} - 1]$ is the vector whose j-th component is

$$[\tilde{\mathbf{A}}^T\mathbf{I}_Q[\tilde{\mathbf{A}}\mathbf{x} - 1]]_j = \sum_{i\in\{i|0\leq b_i(\mathbf{a}_i^T\mathbf{x})\leq 1\}}^{n} b_ja_{i,j}(b_j(\mathbf{a}_i^T\mathbf{x}) - 1)$$

if $0 \leq b_i(\mathbf{a}_i^T\mathbf{x}) \leq 1$. Then, with

$$\frac{1}{n}\tilde{\mathbf{A}}^T\mathbf{I}_Q[\tilde{\mathbf{A}}\mathbf{x} - 1] \tag{13}$$

we can represent the components of $\nabla g_i(\mathbf{x})$ in the aforementioned case.

Combining (12) and (13), it is proven that

$$\nabla \ell_{sh} = \frac{1}{n}(\tilde{\mathbf{A}}^T\mathbf{I}_Q[\tilde{\mathbf{A}}\mathbf{x} - 1] - \tilde{\mathbf{A}}^T\mathbf{I}_L\mathbf{1}) \tag{14}$$

Finally, combining (5) and (14) we get the final result

$$\nabla f(\mathbf{x}) = \lambda\mathbf{x} + \frac{1}{n}\tilde{\mathbf{A}}^T\mathbf{I}_Q[\tilde{\mathbf{A}}\mathbf{x} - 1] - \frac{1}{n}\tilde{\mathbf{A}}^T\mathbf{I}_L\mathbf{1} \tag{15}$$

$\square$

(b) Hessian of f

Proof. Cool proof $\square$

(c) Strong convexity of f

Proof. (Type your proof here.) $\square$