Final Project of MATP6610/4820

(Due on April 28, 2019)

Instructions

For each student, choose one of the two topics below. You should write a report to show how you derive your solver, all results you obtain, and all observations you make. To earn full credit, you need include in your report all results required below each topic, and also, you need make sure your code works in a right way and produce correct solutions. Please pack your report and code in a single .zip file and send it to the email address `optimization.rpi@gmail.com`

Bonus credit: If you develop both ALM and ADMM solvers for topic I OR if you do both topic I and topic II, you will get up to 20% credit.

Topic I: Support vector machine

Description. Given a set of data $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ with $\mathbf{x}_i \in \mathbb{R}^p$ and the label $y_i \in \{+1, -1\}$ for each i, the linear support vector machine (SVM) aims at finding a hyperplane $\mathbf{w}^\top \mathbf{x} + b = 1$ to separate the two classes of data points. Suppose these points can be strictly separately, namely, there exists $(\mathbf{w}, b)$ such that $\mathbf{w}^\top \mathbf{x}_i + b \geq 1$ if $y_i = +1$ and $\mathbf{w}^\top \mathbf{x}_i + b \leq -1$ if $y_i = -1$, or equivalently $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$ for all i. To find a hyperplane with the maximum margin between the two classes, we solve the following problem:

$$\underset{\mathbf{w},b}{\text{maximize}} \frac{1}{\|\mathbf{w}\|}, \text{ subject to } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \forall i = 1, \ldots, N, \tag{1}$$

or equivalently

$$\underset{\mathbf{w},b}{\text{minimize}} \frac{1}{2}\|\mathbf{w}\|^2, \text{ subject to } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \forall i = 1, \ldots, N. \tag{2}$$

While the feasibility of the above problems depends on the separability of the data, a soft-margin SVM is employed if the data points cannot be strictly separated by solving the problem:

$$\underset{\mathbf{w},b,\mathbf{t}}{\text{minimize}} \sum_{i=1}^N t_i + \frac{\lambda}{2}\|\mathbf{w}\|^2, \text{ subject to } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - t_i,\ t_i \geq 0, \forall i = 1, \ldots, N, \tag{3}$$

where $\lambda > 0$ is a given parameter. Note that (3) is more general than (2) because if the data points can be strictly separated, then $t_i = 0$ at the optimal solution for each $i = 1, \ldots, N$. In addition, note that at an optimal solution $(\bar{\mathbf{w}}, \bar{b}, \bar{\mathbf{t}})$, we must have $\bar{t}_i = \max\big(0, 1 - y_i(\bar{\mathbf{w}}^\top \mathbf{x}_i + \bar{b})\big)$. Hence, (3) is equivalent to

$$\underset{\mathbf{w},b}{\text{minimize}} \sum_{i=1}^{N} \max\big(0, 1 - y_i(\bar{\mathbf{w}}^\top \mathbf{x}_i + \bar{b})\big) + \frac{\lambda}{2}\|\mathbf{w}\|^2. \tag{4}$$

Requirements

1. Develop a solver for (3) by the augmented Lagrangian method (ALM) OR a solver for (4) by the alternating direction method of multipliers (ADMM). Your solver should be able to accept $(\mathbf{X}, \mathbf{y}, \lambda)$ and also a few optional parameters $(\mathbf{z}^0, \text{tol}, \text{maxit})$ as the input. Here, $\mathbf{X} \in \mathbb{R}^{p \times N}$ and $\mathbf{y} \in \mathbb{R}^N$ with the i-th column of $\mathbf{X}$ being the i-th data point, and $\mathbf{z}^0$ is the initial iterate.

2. If you use ALM for (3), in your report, explicitly formulate the subproblem that is solved within each outer iteration and explain how you solve the subproblem, including the optimization method you use, the iterative update formula, and the stopping condition. If you use ADMM for (4), in your report, explicitly formulate the two subproblems that are solved at each iteration and explain how you solve the subproblems (provide the solution formula if they have closed-form solutions).

3. Report the stopping condition you use to terminate your ALM or ADMM solver.

4. For the training data sets provided by the instructor, run your solver. Save and report (by figure or table) the violation of primal feasibility and also dual feasibility at each outer iteration.

5. Based on the output solution, do the classification on the testing data and report the accuracy you obtain. To have a higher accuracy, you can pass a smaller tolerance to your solver and/or tune the model parameter λ. You should report the accuracy and the corresponding value of λ you use.

Topic II: Neyman-Pearson Classification

Given data points $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ with $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \in \{+1, -1\}$ for each i, let $\mathcal{N}_+ = \{i \in [N] : y_i = 1\}$ and $\mathcal{N}_- = \{i \in [N] : y_i = -1\}$ be the positive and negative sets of points respectively. The Neyman-Pearson classification (NPC) minimizes the false negative error while controlling the level $\alpha > 0$ of false positive error through solving the problem:

$$\underset{\mathbf{w},b}{\text{minimize}}\ \frac{1}{N_+}\sum_{i\in\mathcal{N}_+}\ell(\mathbf{w},b;\mathbf{x}_i,y_i),\ \text{s.t.}\ \frac{1}{N_-}\sum_{i\in\mathcal{N}_-}\ell(\mathbf{w},b;\mathbf{x}_i,y_i)\le\alpha, \tag{5}$$

where N_+ and N_- denote the cardinalities of $\mathcal{N}_+$ and $\mathcal{N}_-$, and ℓ is a certain error function. For this project, we let ℓ be the error function used in logistic regression, i.e.,

$$\ell(\mathbf{w},b;\mathbf{x}_i,y_i) = \log\left(1+\exp\left(-y_i(\mathbf{w}^\top\mathbf{x}_i+b)\right)\right). \tag{6}$$

Requirements

1. Develop a solver for (5) with ℓ given in (6) by the augmented Lagrangian method (ALM). Your solver should be able to accept $(\mathbf{X},\mathbf{y},\lambda)$ and also a few optional parameters $(\mathbf{z}^0,\text{tol},\text{maxit})$ as the input. Here, $\mathbf{X}\in\mathbb{R}^{p\times N}$ and $\mathbf{y}\in\mathbb{R}^N$ with the i-th column of $\mathbf{X}$ being the i-th data point, and $\mathbf{z}^0$ is the initial iterate.

2. In your report, explicitly formulate the subproblem that is solved within each outer iteration and explain how you solve the subproblem, including the optimization method you choose, the iterative update formula, and the stopping condition.

3. Report the stopping condition you use to terminate your ALM solver.

4. For the training data sets provided by the instructor, run your ALM solver. Save and report (by figure or table) the violation of primal feasibility and also dual feasibility at each outer iteration.

5. Based on the output solution, do the classification on the testing data and report the false positive and false negative errors you obtain. You can tune the model parameter α. You should report the errors and the corresponding value of α you use.