Homework #4: Machine Learning

Kyler Little

March 25, 2018

Problem #1

Exercise 8.5
Show that the matrix Q described in the linear hard-margin SVM algorithm above is positive semi-definite (that is $u^T Q u \geq 0$ for any u).
From the problem statement, Q is defined to be:

$$\mathrm{Q} = \begin{bmatrix} 0 & \mathbf{0}_d^T \\ \mathbf{0}_d & I_d \end{bmatrix}$$

Let $u = \{u_0\} \times x \in R^{d+1}$, where I augment an arbitrary element $u \in R$ to $x \in R^d$. This simplifies notation a little bit. Then:

$$\begin{aligned} u^T \mathrm{Q} u &= [u_0 \; x_1 \cdots \; x_d] Q [u_0 \; x_1 \cdots \; x_d]^T \\ &= [0 \; x_1 \cdots \; x_d][u_0 \; x_1 \cdots \; x_d]^T \\ &= ||x||^2 \geq 0 \end{aligned}$$

Thus, Q is positive semi-definite.

Problem #2

Exercise 8.11
(a) Show that the problem in (8.21) is a standard QP-problem:

$$\begin{aligned} \underset{\alpha \in R^N}{\text{minimize}} \quad & \frac{1}{2}\alpha^T \mathrm{Q}_D \alpha - 1_N^T \alpha \\ \text{subject to} \quad & A_D \alpha \geq 0_{N+2} \end{aligned}$$

where Q_D and A_D (D for the dual) are given by:

$$Q_D = \begin{bmatrix} y_1 y_1 x_1^T x_1 & \ldots & y_1 y_N x_1^T x_N \\ y_2 y_1 x_2^T x_1 & \ldots & y_2 y_N x_2^T x_N \\ \vdots & \vdots & \vdots \\ y_N y_1 x_N^T x_1 & \ldots & y_N y_N x_N^T x_N \end{bmatrix} \text{ and } A_D = \begin{bmatrix} y^T \\ -y^T \\ I_{NxN} \end{bmatrix}$$

Starting from the standard QP-problem, I will derive the original problem (8.21) to show equivalence.
Firstly, it is obvious that $-\sum_{i=1}^{N}\alpha_i = [-1 \ \cdots \ -1][\alpha_1 \ \cdots \alpha_N]^T$. Next, we will show equivalence of the first terms.

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i^T x_j &= \alpha^T \mathrm{Q}_D \alpha \\
&= [\alpha_1 \ \cdots \ \alpha_N]\mathrm{Q}_D[\alpha_1 \ \cdots \ \alpha_N]^T \\
&= [\sum_{i=1}^{N}\alpha_i y_i y_1 x_i^T x_1 \ \cdots \ \sum_{i=1}^{N}\alpha_i y_i y_N x_i^T x_N][\alpha_1 \ \cdots \ \alpha_N]^T \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i^T x_j
\end{aligned}
$$

Lastly, I'll show equivalence of the constraints.

$$
A_D\alpha = \begin{bmatrix} y^T \\ -y^T \\ I_{NxN} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i\alpha_i \\ -\sum_{i=1}^{N} y_i\alpha_i \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} \geq 0
$$

This directly implies that $\alpha_i \geq 0, \ \forall i \in \{1, \ldots, N\}$. And then only way for $\sum_{i=1}^{N} y_i\alpha_i \geq 0$ and $-\sum_{i=1}^{N} y_i\alpha_i \geq 0$ is if $\sum_{i=1}^{N} y_i\alpha_i = 0$. Thus, the problems are equivalent.
(b) The matrix Q_d of quadratic coefficients is $[Q_d]_{mn} = y_m y_n x_m^T x_n$. Show that $Q_d = X_s X_s^T$, where X_s is the 'signed data matrix',

$$
X_s = \begin{bmatrix} y_1 x_1^T \\ y_2 x_2^T \\ \vdots \\ y_N x_N^T \end{bmatrix}
$$

Hence, show that Q_D is positive semi-definite.
First, note that $X_s^T = [y_1 x_1 \ \cdots \ y_N x_N]$. From there, it is pretty easy to see that

$$
X_s X_s^T = \begin{bmatrix} y_1 x_1^T \\ y_2 x_2^T \\ \vdots \\ y_N x_N^T \end{bmatrix} [y_1 x_1 \ \cdots \ y_N x_N] = \begin{bmatrix} y_1 y_1 x_1^T x_1 & \ldots & y_1 y_N x_1^T x_N \\ y_2 y_1 x_2^T x_1 & \ldots & y_2 y_N x_2^T x_N \\ \vdots & \vdots & \vdots \\ y_N y_1 x_N^T x_1 & \ldots & y_N y_N x_N^T x_N \end{bmatrix}
$$

using basic matrix multiplication. To show Q_D is positive semidefinite, we first define arbitrary vector $y \in R^N$. Then, we have:

$$
\begin{aligned}
y^T \mathrm{Q}_D y &= y^T X_s X_s^T y \\
&= (X_s^T y)^T X_s^T y \\
&= ||X_s^T y||^2 \geq 0
\end{aligned}
$$

Thus, Q is positive semi-definite.

Problem #3

Exercise 8.13

KKT complementary slackness gives that if $\alpha_n^* > 0$, then (x_n, y_n) is on the boundary of the optimal fat-hyperplane and $y_n(w^{*T}x_n + b^*) = 1$. Show that the reverse is not true. Namely, it is possible that $\alpha_n^* = 0$ and yet (x_n, y_n) is on the boundary satisfying $y_n(w^{*T}x_n + b^*) = 1$. [Hint: Consider a toy data set with two positive examples at (0, 0) and (1, 0), and one negative example at (0, 1).]

Assuming we have that $y_n(w^{*T}x_n + b^*) = 1$, we need to show that it's possible for $\alpha_n^* = 0$. We can do so by using the toy data set described in the hint. In that problem, the optimal hyperplane is $w = [0\ \ -2]^T$ where $b = 1$. For this case, the class labels would have to be $+1$ and -1. It's easy to see that the point (0, 1) is not a support vector. If we remove it, the optimal hyperplane wouldn't change. Thus, $\alpha_n^* = 0$ for that data point. All that's left to show is that $y_n(w^{*T}x_n + b^*) = 1$ for all three data points. If this is the case, we have all data points lying on the boundary but one point exists with $\alpha_n^* = 0$. For (0, 0), we have $+1((0 * 0 + -2 * 0) + 1) = 1$. For (1, 0), we have $-1((0 * 0 + -2 * 1) + 1) = 1$. Lastly, for (1, 0), we have $-1((0 * 1 + -2 * 0) + 1) = 1$. Thus, there exists a counterexample, so the statement cannot be true.

Problem #4

Problem 8.1

Consider a data set with two data points $x_\pm \in R^d$ having class ± 1 respectively. Manually solve (8.4) by explicitly minimizing $||w||^2$ subject to the two separation constraints. Compute the optimal (maximum margin) hyperplane (b^*, w^*) and its margin. Compare with your solution to Exercise 8.1.

The two separation constraints are:

$$\begin{aligned} w^T x_+ + b &\geq 1 \\ -w^T x_- - b &\geq 1 \end{aligned}$$

Combining these two constraints yields: $w^T(x_+ - x_-) \geq 2$. This can be equivalently written as:

$$||w|| \frac{w^T(x_+ - x_-)}{||w||} \geq 2$$

Note that in the problem statement we are trying to explicitly minimize $||w||^2$; this is equivalent to minimizing $||w||$. Thus, we need to maximize $1/||w||$, or equivalently $\frac{w^T(x_+ - x_-)}{||w||}$. Since w^T and $(x_+ - x_-)$ are vectors, we need to maximize their dot product. This is only possible if w and $(x_+ - x_-)$ have zero angle between them. This would mean that their dot product is: $w^T(x_+ - x_-) = ||w|| * ||x_+ - x_-|| \cos(0)$. Since w must be in the same direction as $(x_+ - x_-)$, we know then $w = \lambda(x_+ - x_-)$, where λ is just some scalar. Subbing this in yields $\lambda \geq \frac{2}{||x_+ - x_-||^2}$. Since we are minimizing w and λ is directly proportional to w, we must minimize λ as well. This results in the value of w being:

$$w = 2\frac{(x_+ - x_-)}{||x_+ - x_-||^2}$$

The magnitude of w is: $||w|| = w^T w = \frac{2}{||x_+ - x_-||}$; this means that the margin is $\frac{||x_+ - x_-||}{2}$, which makes sense. Lastly, b is easily solved for by using either of the constraints:

$$b = 1 - \frac{2||x_+||^2 - 2x_-^T x_+}{||x_+ - x_-||^2}$$

Problem #5

Problem 8.2
Consider a data set with three data points in R^2:

$$X = \begin{bmatrix} 0 & 0 \\ 0 & -1 \\ -2 & 0 \end{bmatrix} \text{ and } y = \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix}$$

Manually solve (8.4) to get the optimal hyperplane (b^*, w^*) and its margin.
The three constraints produced are:

$$\begin{aligned} -w^T x_1 - b \geq 1 &\leftrightarrow b \leq -1 \\ -w^T x_2 - b \geq 1 &\leftrightarrow w_2 - b \geq 1 \\ w^T x_3 + b \geq 1 &\leftrightarrow -2w_1 + b \geq 1 \end{aligned}$$

At first glance, it seems we need to make b as small as possible, but this will actually make w quite large. Instead, we choose $b = -1$ and sub in accordingly to get w_1 and w_2. This results in the optimal w and b. The values are:

$$(b^*, w^*) = (-1, [-1\ 0]^T)$$

Problem #6

Problem 8.4
Set up the dual problem for the toy data set in Exercise 8.2. Then, solve the dual problem and compute α^*, the optimal Lagrange multipliers.
The toy data set is:

$$X = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \end{bmatrix}, y = \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix}, w = \begin{bmatrix} 1.2 \\ -3.2 \end{bmatrix}, \text{ and } b = -0.5$$

The dual problem in QP form is:

$$\begin{aligned} \underset{\alpha \in R^N}{\text{minimize}} \quad & \frac{1}{2}\alpha^T \mathrm{Q}_D \alpha - 1_3^T \alpha \\ \text{subject to} \quad & A_D \alpha \geq 0_5 \end{aligned}$$

where Q_D and A_D (D for the dual) are given by:

$$Q_D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 8 & -4 \\ 0 & -4 & 4 \end{bmatrix} \text{ and } A_D = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Plugging in Q_D and A_D leads us to

$$\begin{aligned} \underset{\alpha \geq 0}{\text{minimize}} \quad & 4\alpha_2^2 - 2\alpha_2\alpha_3 + 2\alpha_3^2 - \alpha_1 - \alpha_2 - \alpha_3 \\ \text{subject to} \quad & -\alpha_1 - \alpha_2 + \alpha_3 \geq 0 \\ & \alpha_1 + \alpha_2 - \alpha_3 \geq 0 \end{aligned}$$

The constraints actually directly imply that $-\alpha_1 - \alpha_2 + \alpha_3 = 0$. I'll solve for α_3 and eliminate the α_3's from the minimization problem. Once we substitute in, we get $8\alpha_2^2 + 2\alpha_1\alpha_2 + 2\alpha_1^2 - 2\alpha_1 - 2\alpha_2$. Since we are minimizing this, we simply can take the partial derivatives and set them to zero. This gives us the values $\alpha_1 = \frac{7}{15}$ and $\alpha_2 = \frac{1}{15}$. Using the constraint yields $\alpha_3 = \frac{8}{15}$. We then plug this into $w = \sum_{i=1}^{3} \alpha_i y_i x_i$ to obtain w. This gives us:

$$w = \begin{bmatrix} -\frac{14}{15} \\ -\frac{2}{15} \end{bmatrix}.$$