CS-419m: Practice question set 3

1. A popular loss function for regression is square error For the following alternative error functions discuss pros and cons on account of accuracy, sensitivity to noisy training data, and training time.

   (a) Absolute loss: $|y^i - \mathbf{w}^T\mathbf{x}^i|$ This loss function is better behaved than the Square loss because it is less sensitive to outliers in the training data. The training time will be higher because the function to be optimized is not differentiable. However, the training objective with the above loss is convex and can be solved using constrained convex optimization techniques.

   (b) Clipped loss: $\min(|y^i - \mathbf{w}^T\mathbf{x}^i|, T)$ where T is a user specified threshold. This is even more suited to handle outliers in training data because the loss will not continue to increase as we clip it at T. However, the training objective in this case, is neither differentiable nor convex. Therefore, it is difficult to guarantee a globally optimal value for $\mathbf{w}$. ..4

   (c) Suppose we wish to train a square error regression function with a $L2$ regularizer. That is the training objective is $\min_{\mathbf{w}} F(\mathbf{w})$ where $F(\mathbf{w}) = \sum_i (y_i - \mathbf{w}.\mathbf{x}^i)^2 + ||\mathbf{w}||^2$. Suppose, we decide to use a descent-based algorithm as follows:

   $t = 0$, $\mathbf{w}^t = 0$
   while $||\nabla F(\mathbf{w}^t)|| \neq 0$ **do**
   Choose a direction $\mathbf{e}^t$.
   $g_t(\lambda) = F(\mathbf{w}^t + \lambda\mathbf{e}^t) - F(\mathbf{w}^t)$
   $\lambda^* = \min_\lambda g_t(\lambda)$
   $\mathbf{w}^{t+1} = \mathbf{w}^t + \lambda^*\mathbf{e}^t$
   $t = t + 1$
   end while

   Let the labeled data be as shown in the table below.

x_1	x_2	y
2	8	1
4	7	1
1	5	-1
7	2	-1
8	3	-1

   i. Suppose you are allowed to add an $N+1$th training point, what would you choose so that the algorithm terminates without entering the while loop?
   The algorithm won't enter the loop if $||\nabla F(\mathbf{w}^0)|| = 0$
   $||\nabla F(\mathbf{w}^0)|| = \sum_i 2(y_i - \mathbf{w}^{0^T}\mathbf{x}^i)(-\mathbf{x}^i) + 2\mathbf{w}^0 = [0\ \ 0]^T$
   $\sum_i y_i\mathbf{x}^i = [0\ \ 0]^T$
   $\sum_i y_i x_1^i = 0$ and $\sum_i y_i x_2^i = 0$
   For $y_6 = 1$,
   $x_1^6 = -\sum_i y_i x_1^i$ and $x_2^6 = -\sum_i y_i x_2^i$, i from 1 to 5
   which gives $x_1^6 = 10$ and $x_2^6 = -5$

   ..2

   ii. For the N points in the table, suppose at $t = 0$ $\mathbf{e}^t = [1\ 1]'$, what is $g_0(\lambda)$. (Make sure $g_0(\lambda)$ contains no variable other than λ).

$g_0(\lambda) = F(\mathbf{w}^0 + \lambda[1\ \ 1]^T) - F(\mathbf{w}^0)$
$g_0(\lambda) = F([\lambda\ \ \lambda]^T) - F([0\ \ 0]^T)$
$g_0(\lambda) = (1-10\lambda)^2 + (1-11\lambda)^2 + (-1-6\lambda)^2 + (-1-9\lambda)^2 + (-1-11\lambda)^2 + 2\lambda^2 - 5$
$g_0(\lambda) = 461\lambda^2 + 10\lambda$

..2

iii. Find optimal λ^* for g_0.
$\lambda^* = \min_\lambda g_0(\lambda)$
$g_0(\lambda)^{'} = 922\lambda + 10 = 0$
$g_0(\lambda)^{''} = 922 > 0$
$\lambda^* = -5/461$

..1

iv. Now suppose we solve for the $\mathbf{w}$ that minimizes $F(\mathbf{w})$ directly in closed form. Show that the optimal $\mathbf{w}^*$ can be written as $\sum_{i=1}^{N} \alpha_i y_i \mathbf{x}^i$ and identify the optimal value of α_i.
$||\nabla F(\mathbf{w})|| = \sum_i 2(y_i - \mathbf{w}^T\mathbf{x}^i)(-\mathbf{x}^i) + 2\mathbf{w} = [0\ \ 0]^T$
$-\sum_i y_i\mathbf{x}^i + \mathbf{x}^i(\mathbf{x}^{iT}\mathbf{w}) + \mathbf{w} = [0\ \ 0]^T$
$(\mathbf{x}^i\mathbf{x}^{iT} + I)\mathbf{w} = \sum_i y_i\mathbf{x}^i$
$\mathbf{w} = (\mathbf{x}^i\mathbf{x}^{iT} + I)^{-1}\sum_i y_i\mathbf{x}^i$

..3