EE 559 **Homework 9** Posted: Thur., 3/7/2019
Jenkins Due: Thur., 3/21/2019, 2:00 PM

1. In the parts below, you will set up the primal Lagrangian and derive the dual Lagrangian, for support vector machine classifier, for the case of data that is separable in expanded feature space.

   For the SVM learning, we stated the optimization problem for the linearly separable (in $\underline{u}$-space) case, as:

$$\text{Minimize } J(\underline{w}) = \tfrac{1}{2}\|\underline{w}\|^2$$
$$\text{s.t. } z_i\left(\underline{w}^T\underline{u}_i + w_0\right) - 1 \geq 0 \quad \forall i$$

   Assume our data is linearly separable in the expanded feature space. Also, nonaugmented notation is used throughout this problem.

(a) If the above set of constraints (second line of equations above) is satisfied, will all the training data be correctly classified?

(b) Write the Lagrangian function $L(\underline{w}, w_0, \underline{\lambda})$ for the minimization problem stated above. Use $\lambda_i,\ i = 1, 2, \cdots, N$ for the Lagrange multipliers. Also state the KKT conditions. **(Hint:** there are 3 KKT conditions).

(c) Derive the dual Lagrangian L_D, by proceeding as follows:

   (i) Minimize L w.r.t. the weights.
   Hint: solve $\nabla_{\underline{w}} L = \underline{0}$ for the optimal weight $\underline{w}^*$ (in terms of λ_i and other variables); and set $\dfrac{\partial L}{\partial w_0} = 0$ and simplify.

   (ii) Substitute your expressions from part (i) into L, and use your expression from $\dfrac{\partial L}{\partial w_0} = 0$ as a new constraint, to derive L_D as:

$$L_D(\underline{\lambda}) = -\tfrac{1}{2}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j z_i z_j \underline{u}_i^T\underline{u}_j\right] + \sum_{i=1}^{N}\lambda_i$$

   subject to the (new) constraint: $\sum_{i=1}^{N}\lambda_i z_i = 0$. Also give the other two KKT conditions on λ_i, which carry over from the primal form.

2. In this problem you will do SVM learning on 2 data points, using the result from Problem 1 above.

You are given a training data set consisting of 2 samples, in a 2-class problem. In the expanded feature space, the samples are:

$$\underline{u}_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \in S_1; \quad \underline{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in S_2.$$

(a) Solve, by hand, for the SVM decision boundary and regions. To do this, use Lagrangian techniques to optimize L_D w.r.t. $\underline{\lambda}$ subject to its equality constraint $\sum_{i=1}^{N} \lambda_i z_i = 0$, in order to find the optimal weight vector $\underline{w}^*$, and also find the optimal w_0. Plot the training samples and the decision boundary in 2D (nonaugmented) $\underline{u}$-space, and show which side of the boundary is positive (Γ_1).

Hints: Start from:

$$L_D{}'(\underline{\lambda}, \mu) = \sum_{i=1}^{N} \lambda_i - \tfrac{1}{2}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j z_i z_j \underline{u}_i{}^T \underline{u}_j\right] + \mu\left(\sum_{i=1}^{N} z_i \lambda_i\right)$$

in which the last term has been added to incorporate the equality constraint stated above. Once finished, then check that the resulting $\underline{\lambda}$ satisfies the KKT conditions on $\underline{\lambda}$ that you stated in Problem 1(c)(ii); one of these can also help you find $w_0{}^*$.

(b) Use the data points and solution you found in part (a), and call its solution decision boundary $\mathbf{H}$. Calculate in 2D $\underline{u}-$space the distance between $\underline{u}_1$ and $\mathbf{H}$, and the distance between $\underline{u}_2$ and $\mathbf{H}$. Is there any other possible linear boundary in $\underline{u}-$space that would give larger values for both distances ("margins") than $\mathbf{H}$ gives? (No proof required.)

3. In this problem you will derive the dual Lagrangian for the case of data that is not linearly separable in $\underline{u}$-space. From class, the optimization problem can be stated as:

$$\text{Minimize } J(\underline{w}) = \tfrac{1}{2}\|\underline{w}\|^2 + C\sum_{i=1}^{N} \xi_i$$

$$\text{s.t. } z_i\left(\underline{w}^T \underline{u}_i + w_0\right) \geq 1 - \xi_i \quad \forall i$$

$$\xi_i \geq 0 \quad \forall i$$

Tip: Write clearly so that μ_i and $\underline{u}_i$ are always distinguishable.

(a) Write the Lagrangian function $L\left(\underline{w}, w_0, \underline{\xi}, \underline{\lambda}, \underline{\mu}\right)$ for the minimization problem stated above. Use $\lambda_i, i = 1, 2, \cdots, N$ for the Lagrange multipliers relating to the first set of constraints, and use $\mu_i, i = 1, 2, \cdots, N$ for the Lagrange multipliers relating to the second set of constraints State the KKT conditions. **Hint:** there are 6 KKT conditions: 4 sets of restrictions on the λ_i and $\underline{u}_i$, and the 2 given sets of constraints.

(b) Derive the dual Lagrangian L_D, by minimizing L w.r.t. $\underline{w}$, w_0 , and $\underline{\xi}$, and give the formula for $\underline{w}^*$. Also list the restrictions and constraints on the Lagrange multipliers ($\underline{\lambda}$ and $\underline{\mu}$): there are 5 sets of them.

Hints:

(1) Procedure is similar to the separable-data case.

(2) Each ξ_i can be treated as another independent variable, so that $\dfrac{\partial \xi_i}{\partial \underline{w}} = 0$.

(3) To minimize L w.r.t. $\underline{\xi}$, set $\dfrac{\delta L}{\delta \xi_i} = 0$. This gives an equation that can be combined with restrictions on μ_i and λ_i to yield a revised restriction on λ_i.