ECE523: Engineering Applications of Machine Learning and Data Analytics
Due 03/18/2020 @ 11:59PM (D2L)

Name: Zhaoyang Bu

Signature: Zhaoyang Bu

Date: 3/23/2020

Instructions: There are three problems. Partial credit is given for answers that are partially correct. No credit is given for answers that are wrong or illegible. All work must be supported and code must be submitted for credit.

Theory: ____________

Practice: ____________

Total: ____________

Part A: Theory (20pts)

(10pts) Support Vector Machines

In class, we discussed that if our data is not linearly separable, then we need to modify our optimization problem to include slack variables. The formulation that was used is known as the ℓ_1-norm soft marging SVM. Now consider the formulation of the ℓ_2-norm soft margin SVM, which squares the slack variables within the sum. Notice that non-negativity of the slack variables has been removed.

$$\arg\min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2$$
$$\text{s.t. } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i \qquad \forall i \in [n]$$

Derive the dual form expression along with any constraints. Work must be shown. *Hints*: Refer to the methodology that was used in class to derive the dual form. The solution is given by:

$$\arg\max_{\alpha} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j - \frac{1}{2C}\sum_{i=1}^{n}\alpha_i^2$$
$$\text{s.t. } \alpha_i \geq 0 \quad \forall i \in [n] \quad \text{and} \quad \sum_{i=1}^{n}\alpha_i y_i = 0$$

(10pts) Support Vector Machines (Revisited)

We now look at a different type of SVM that is designed for domain adaptation and optimizes the hyperplanes given by $\mathbf{w}_S$ (source hyperplane) before optimizing $\mathbf{w}_T$ (target hyperplane). The process begins by training a support vector machine on source data then once data from the target are available, train a new SVM using the hyperplane from the first SVM and the data from the target to solve for a new "domain adaptation" SVM.

The primal optimization problem is given by

$$\begin{aligned}
\min_{\mathbf{w}_T,\xi} \quad & \frac{1}{2}\|\mathbf{w}_T\|^2 + C\sum_{i=1}^{n}\xi_i - B\mathbf{w}_T^T\mathbf{w}_S \\
\text{s.t.} \quad & y_i(\mathbf{w}_T^T\mathbf{x}_i + b) \geq 1 - \xi_i & \forall i \in \{1,\dots,n\} \\
& \xi_i \geq 0 & \forall i \in \{1,\dots,n\}
\end{aligned}$$

where $\mathbf{w}_S$ is hyperplane trained on the source data (*assumed to be known*), $\mathbf{w}_T$ is hyperplane for the target, $y_i \in \{\pm 1\}$ is the label for instance $\mathbf{x}_i$, C & B are regularization parameters defined by the user and ξ_i is a slack variable for instance $\mathbf{x}_i$. The problem becomes finding a hyperplane, $\mathbf{w}_T$, that minimizes the above objective function subject to the constraints. Solve/derive the dual optimization problem.

Part A: Practice (40pts)

(10pts) Multi-Layer Perceptron

(200pts) Support Vector Machines

Generate 2D Gaussian data with *at least* two components (i.e., you need at least one for the positive class and one for the negative class). Train and test a classifier of two disjoint data sets generated from the Gaussian components you described above and your selection of kernel parameters. Plot the training data and indicate Compute the classifier error, plot the training data, and plot the test data with the predicted class labels. Use at least two different kernels, and submit your results & code. One of the kernels should be the RBF kernel and you must show the impact the the RBF free parameter has on the decision boundary.

(20pts) Support Vector Machines (Revisited)

Implement the domain adaptation SVM from the first problem in the theory section. A data set for the source and target domains (both training and testing) have been uploaded to D2L. There are several ways to implement this algorithm. If I were doing this for an assignment, I would implement the SVM directly using quadratic programming. For example, see CVX (https://goo.gl/3f7StQ), but there are other tools available. You do not need to build the classifier (i.e., solve for the bias term); however, you will need to find $\mathbf{w}_T$ and $\mathbf{w}_S$.

ECE 523
HW#3
Zhaoyang Bu

1) $\arg\min_{w,b,\xi} \frac{1}{2}\|W\|_2^2 + \frac{C}{2}\sum_{i=1}^n \xi_i^2$, s.t. $y_i(W^T x_i + b) \geq 1 - \xi_i$, $\forall i \in [n]$

$L(w,b,\xi,\alpha) = \frac{1}{2}\|W\|_2^2 + \frac{C}{2}\sum_{i=1}^n \xi_i^2 - \sum \alpha_i [y_i(W^T x_i + b) - 1 + \xi_i]$

$\frac{\partial L}{\partial w} = 0 \Rightarrow w + 0 - 0 - \sum_{i=1}^n \alpha_i y_i x_i = 0$

$W = \sum_{i=1}^n \alpha_i y_i x_i$

$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0$

$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C\xi_i = \alpha_i$

$\xi_i = \frac{\alpha_i}{C}$

$\Rightarrow L = \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j + \frac{C}{2}\sum_{i=1}^n \frac{\alpha_i^2}{C^2} - \sum_{i=1}^n \alpha_i y_i (W^T x_i + b) + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i$

$= -\frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j + \frac{1}{2C}\sum_{i=1}^n \alpha_i^2 - b\sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \frac{\alpha_i^2}{C}$

$\Rightarrow \arg\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j - \frac{1}{2C}\sum_{i=1}^n \alpha_i^2$

s.t. $\alpha_i \geq 0$ $\forall i \in [n]$

$\sum_{i=1}^n \alpha_i y_i = 0$

2) $\min_{w_T, \xi} \frac{1}{2}\|W_T\|_2^2 + C\sum_{i=1}^n \xi_i - B W_T^T W_S$

s.t. $\begin{cases} y_i(W_T^T x_i + b) \geq 1 - \xi_i & \forall i \in \{1, \dots, n\} \\ \xi_i \geq 0 & \forall i \in \{1, \dots, n\} \end{cases}$

$L = \frac{1}{2}\|W_T\|_2^2 + C\sum_{i=1}^n \xi_i - B W_T^T W_S - \sum_{i=1}^n \alpha_i [y_i(W_T^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^n \mu_i \xi_i$

$\frac{d_L}{d w_T} = 0 \Rightarrow W_T = BW_S + \sum \alpha_i y_i x_i$

$\frac{d_L}{d_b} = 0 \Rightarrow \sum \alpha_i y_i = 0$

$\frac{d_L}{d \xi} = 0 \Rightarrow C = \alpha_i + \mu_i$

$\Rightarrow$

$L = \frac{1}{2}(BW_S + \sum_{i=1}^n \alpha_i y_i x_i)^2 + \cancel{C\sum_{i=1}^n \xi_i} - BW_S(BW_S + \sum_{i=1}^n \alpha_i y_i x_i)^T - \sum_{i=1}^n \alpha_i y_i [(BW_S + \sum_{i=1}^n \alpha_i y_i x_i)^T x_i + b] + \sum_{i=1}^n \alpha_i - \cancel{\sum_{i=1}^n \alpha_i \xi_i} - \cancel{\sum_{i=1}^n \mu_i \xi_i}$

$= \frac{1}{2}(BW_S^T + \sum_{i=1}^n \alpha_i y_i x_i^T)(BW_S + \sum_{i=1}^n \alpha_i y_i x_i) - BW_S(BW_S^T + \sum_{i=1}^n \alpha_i y_i x_i^T) - \sum_{i=1}^n \alpha_i y_i [(BW_S^T + \sum_{i=1}^n \alpha_i y_i x_i^T) x_i + b] + \sum_{i=1}^n \alpha_i$

$= \frac{1}{2}(BW_S)^2 + \frac{1}{2}BW_S^T \sum_{i=1}^n \alpha_i y_i x_i + \frac{1}{2}BW_S \sum \alpha_i y_i x_i^T + \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j - (BW_S)^2 - BW_S \sum_{i=1}^n \alpha_i y_i x_i^T - BW_S^T \sum_{i=1}^n \alpha_i y_i x_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j - b\sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i$

$= \sum_{i=1}^n \alpha_i - \frac{1}{2}(BW_S)^2 - \frac{1}{2}BW_S^T \sum_{i=1}^n \alpha_i y_i x_i - \frac{1}{2}BW_S \sum \alpha_i y_i x_i^T - \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j$

$\Rightarrow \arg\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2}(BW_S)^2 - \frac{1}{2}BW_S^T \sum_{i=1}^n \alpha_i y_i x_i - \frac{1}{2}BW_S \sum_{i=1}^n \alpha_i y_i x_i^T - \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j$

s.t. $\begin{cases} 0 \leq \alpha_i \leq C & \forall i \in \{1, \dots, n\} \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases}$