CS 5691: Pattern Recognition and Machine Learning
Worksheet 3
Instructor: Harish Guruprasad Ramaswamy – BSB350 – hariguru@cse.iitm.ac.in

1. Consider the following 2-dimensional binary classification dataset with 8 points given by

$$X^\top = \begin{bmatrix} -2 & -2 & -1 & -1 & 1 & 1 & 2 & 3 \\ -1 & 2 & 1 & 2 & 1 & 3 & 3 & 2 \end{bmatrix}$$
$$\mathbf{y}^\top = [+1, +1, +1, -1, -1, +1, -1, -1]$$

Run one iteration of gradient descent with the logistic regression objective by hand. No bias required, only the 2-dimensional weight vector is to be optimised. Choose the step size $\eta = 1$. Initialise at $\mathbf{w} = [0, 0]^\top$.

2. Consider a regression problem with training data $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ with $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$. The Support Vector Regression algorithm, essentially solves the below problem:

$$\min_{\mathbf{w}, b} \frac{1}{2}||\mathbf{w}||^2$$
$$\text{s.t.} \mathbf{w}^\top \mathbf{x}_i + b \le y_i + \epsilon$$
$$\mathbf{w}^\top \mathbf{x}_i + b \ge y_i - \epsilon$$

for some fixed $\epsilon > 0$. Derive the Lagrangian dual optimisation problem to the above problem.

3. Consider the following 2-dimensional binary classification dataset with 8 points given by

$$X^\top = \begin{bmatrix} 1 & 1 & 2 & 4 & 5 \\ 1 & 0 & 5 & 4 & 2 \end{bmatrix}$$
$$\mathbf{y}^\top = [+1, +1, -1, -1, -1]$$

Consider the hard margin SVM problem with linear kernel $k(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\top \mathbf{v}$.

(a) Give the support vectors just by looking at the data. Give reasons.
(b) Give the dual solution $\boldsymbol{\alpha}^*$ using the answer to the above part.
(c) Check if the entire solution got above is the right answer using KKT conditions. (Thus also checking the first part guessed by "eyeballing".)
(d) Derive the primal solution $\mathbf{w}^*, b^*$ from the dual solution $\boldsymbol{\alpha}^*$ and draw a figure illustrating the final solution.

4. Let $\mathbf{u} \in \mathbb{R}^d$ be a point. Let $\mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}$ and the hyperplane given by $\mathbf{w}, b$ is $\{\mathbf{x} \in \mathbb{R}^d : \mathbf{w}^\top \mathbf{x} + b = 0\}$. Consider the following problem of projection of the point $\mathbf{u}$ on to a (hyper)plane given by $\mathbf{w}, b$.

$$\min_{\mathbf{v} \in \mathbb{R}^d} \frac{1}{2}||\mathbf{v} - \mathbf{u}||^2$$
$$\text{s.t. } \mathbf{w}^\top \mathbf{v} + b = 0$$

Derive the solution to the above problem via solving the Lagrangian dual (which is an unconstrained quadratic problem, and hence can be easily solved). Show that distance of the point $\mathbf{u}$ to the hyperplane given by $\mathbf{w}, b$ is $\frac{|\mathbf{w}^\top v + b|}{||\mathbf{w}||}$.

5. Let $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ be a binary classification dataset. Let $\mathbf{w}^*, b^*$ be any solution to the problem below:

$$\max_{\mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{||\mathbf{w}||}$$
$$\text{s.t. } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \ge 1$$

Show that $\min_{i \in [n]} y_i(\mathbf{w}^\top \mathbf{x}_i + b) = 1$.

6. Consider the following 2-dimensional binary classification dataset with 8 points given by

$$X^\top = \begin{bmatrix} 1 & 1 & 2 & 2 & 4 & 4 & 5 & 5 \\ 0 & 1 & 0 & 1 & 3 & 4 & 3 & 4 \end{bmatrix}$$
$$\mathbf{y}^\top = [-1, -1, -1, -1, +1, +1, +1, +1]$$

Consider the hard-margin linear SVM problem. Evaluate the following $\mathbf{w}, b$. By evaluate, check if it satisfies feasibility, and give the objective value.

(a) $\mathbf{w} = [\frac{1}{2}, 0], b = \frac{-3}{2}$
(b) $\mathbf{w} = [1, 0], b = -3$
(c) $\mathbf{w} = [2, 0], b = -6$
(d) $\mathbf{w} = [1, 0], b = -4$
(e) $\mathbf{w} = [1, 0], b = -5$
(f) $\mathbf{w} = [\frac{1}{4}, \frac{1}{4}], b = \frac{-5}{4}$
(g) $\mathbf{w} = [\frac{1}{2}, \frac{1}{2}], b = \frac{-5}{2}$
(h) $\mathbf{w} = [1, 1], b = -5$
(i) $\mathbf{w} = [1, 1], b = -6$
(j) $\mathbf{w} = [1, 1], b = -7$

7. Consider the following 2-dimensional binary classification dataset with 12 points given by

$$X^\top = \begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 4 & 4 & 5 & 5 & 6 & 6 \\ -1 & 0 & 1 & -1 & 0 & 1 & 3 & 4 & 3 & 4 & 3 & 4 \end{bmatrix}$$
$$\mathbf{y}^\top = [-1, -1, -1, -1, +1, -1, +1, -1, +1, +1, +1, +1]$$

Consider the soft-margin linear SVM problem with $C = 0.1, 1, 10, 100$. For each C evaluate the following $\mathbf{w}, b$. By evaluate, I mean you should give a the slack variables ξ that make the $\mathbf{w}, b, \xi$ feasible, and also give the value of the objective.

(a) $\mathbf{w} = [\frac{1}{2}, 0], b = \frac{-3}{2}$
(b) $\mathbf{w} = [1, 0], b = -3$
(c) $\mathbf{w} = [2, 0], b = -6$
(d) $\mathbf{w} = [1, 0], b = -4$
(e) $\mathbf{w} = [1, 0], b = -5$
(f) $\mathbf{w} = [\frac{1}{4}, \frac{1}{4}], b = \frac{-5}{4}$
(g) $\mathbf{w} = [\frac{1}{2}, \frac{1}{2}], b = \frac{-5}{2}$
(h) $\mathbf{w} = [1, 1], b = -5$
(i) $\mathbf{w} = [1, 1], b = -6$
(j) $\mathbf{w} = [1, 1], b = -7$

8. Consider the following 2-dimensional binary classification dataset with 10 points given by

$$X^\top = \begin{bmatrix} 1 & 1 & 2 & 2 & 4 & 4 & 5 & 5 & 2.9 & 3.1 \\ 0 & 1 & 0 & 1 & 3 & 4 & 3 & 4 & 6 & 6 \end{bmatrix}$$
$$\mathbf{y}^\top = [-1, -1, -1, -1, +1, +1, +1, +1, -1, +1]$$

Consider the soft-margin linear SVM problem with $C = 0.1, 1, 10, 100$. For each C evaluate the following $\mathbf{w}, b$. By evaluate, I mean you should give a the slack variables ξ that make the $\mathbf{w}, b, \xi$ feasible, and also give the value of the objective.

(a) $\mathbf{w} = [\frac{1}{2}, 0], b = \frac{-3}{2}$
(b) $\mathbf{w} = [1, 0], b = -3$
(c) $\mathbf{w} = [4, 0], b = -12$
(d) $\mathbf{w} = [16, 0], b = -48$
(e) $\mathbf{w} = [64, 0], b = -192$
(f) $\mathbf{w} = [\frac{1}{4}, \frac{1}{4}], b = \frac{-5}{4}$
(g) $\mathbf{w} = [\frac{1}{2}, \frac{1}{2}], b = \frac{-5}{2}$
(h) $\mathbf{w} = [1, 1], b = -5$
(i) $\mathbf{w} = [2, 2], b = -10$
(j) $\mathbf{w} = [4, 4], b = -20$

9. Consider the following 2-dimensional binary classification dataset with 10 points given by

$$X^\top = \begin{bmatrix} 1 & 1 & 2 & 2 & 4 & 4 & 5 & 5 & 2.9 & 3.1 \\ 0 & 1 & 0 & 1 & 3 & 4 & 3 & 4 & 6 & 6 \end{bmatrix}$$
$$\mathbf{y}^\top = [-1, -1, -1, -1, +1, +1, +1, +1, -1, +1]$$

Consider the hard-margin linear SVM problem. Guess the support vectors, and use it to solve the dual SVM problem to get $\boldsymbol{\alpha}^*$. Verify your guess by checking if the $\boldsymbol{\alpha}^*$ satisfies the KKT conditions of the dual. Use $\boldsymbol{\alpha}^*$ to get the primal optimal solutions $\mathbf{w}^*, b^*$.

10. Consider the following 2-dimensional binary classification dataset with 12 points given by

$$X^\top = \begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 4 & 4 & 5 & 5 & 6 & 6 \\ -1 & 0 & 1 & -1 & 0 & 1 & 3 & 4 & 3 & 4 & 3 & 4 \end{bmatrix}$$
$$\mathbf{y}^\top = [-1, -1, -1, -1, +1, -1, +1, -1, +1, +1, +1, +1]$$

Consider the soft-margin linear SVM problem with $C = 0.1, 1, 10, 100$. Use software to just get the points i, where $\alpha_i^* = 0$ or $\alpha_i^* = C$ for the above problem (for each value of C). And use just that to derive the dual optimal solution $\boldsymbol{\alpha}^*$ and the primal optimal solution $\mathbf{w}^*, b^*$. (In a test you can reasonably expect that the support vectors will be given to you, and you can just do the rest.)

11. Consider the following 2-dimensional binary classification dataset with 10 points given by

$$X^\top = \begin{bmatrix} 1 & 1 & 2 & 2 & 4 & 4 & 5 & 5 & 2.9 & 3.1 \\ 0 & 1 & 0 & 1 & 3 & 4 & 3 & 4 & 6 & 6 \end{bmatrix}$$
$$\mathbf{y}^\top = [-1, -1, -1, -1, +1, +1, +1, +1, -1, +1]$$

Consider the soft-margin linear SVM problem with $C = 0.01, 0.1, 1, 10, 100$. Use software to just get the points i, where $\alpha_i^* = 0$ or $\alpha_i^* = C$ for the above problem (for each value of C). And use just that to derive the dual optimal solution $\boldsymbol{\alpha}^*$ and the primal optimal solution $\mathbf{w}^*, b^*$. Try and explain why the optimal $\mathbf{w}^*, b^*$ changes with C as above.

12. Consider the following 1-dimensional binary classification dataset with 8 points given by

$$X^\top = [1, 2, 4, 5, 6, 7, 9, 10]$$
$$\mathbf{y}^\top = [+1, +1, -1, -1, -1, -1, +1, +1]$$

Solve the Kernel hard margin SVM problem with the kernel $k(u, v) = \exp(-\gamma(u - v)^2)$. Try $\gamma = 0.1$ and $\gamma = 10$. For each γ, use software to just get the support vectors for the above problem. And use just that to derive the dual optimal solution $\boldsymbol{\alpha}^*$. Give the primal solution corresponding to b^*. (The solution part corresponding to $\mathbf{w}^*$ is infinte dimensional). Draw a plot of the decision function given by

$$\hat{y}(x) = \boldsymbol{\phi}(x)^\top \mathbf{w}^* + b^* = \sum_{i=1}^{8} \alpha_i^* k(x, x_i) + b^*$$

13. Consider the following 2-dimensional binary classification dataset with 9 points given by

$$X^\top = \begin{bmatrix} 0 & 0 & 1 & 1 & 3 & 2 & 3 & 4 & 5 \\ 0 & 1 & 0 & 1 & 3 & 5 & 4 & 3 & 2 \end{bmatrix}$$
$$\mathbf{y}^\top = [+1, +1, +1, +1, -1, -1, -1, -1, -1]$$

Consider the hard margin linear SVM problem. Argue what would the support vectors be. Use this to derive the dual optimal $\boldsymbol{\alpha}^*$, and hence the primal optimal $\mathbf{w}^*, b^*$.

Repeat the above if the point $\mathbf{x}_5 = [3, 3]$ is removed.

Hint : The dual optimal $\boldsymbol{\alpha}^*$ need not be unique, even if the primal optimal $\mathbf{w}^*, b^*$ are. Bonus points for arguing why this can be so for the Hard-margin SVM solution.

14. Consider a soft margin SVM problem with C set to some constant. Let $\boldsymbol{\alpha}^*$ be the dual solution, and let $\mathbf{w}^*, b^*$ be the dual solution. Let the dataset be $(\mathbf{x}_i, y_i)$ with i ranging from 1 to n.

   (a) If $\alpha_i^* = 0$ what are the possible range of values of $(\mathbf{w}^*)^\top \mathbf{x}_i + b^*$?

   (b) If $0 < \alpha_i^* < C$ what are the possible range of values of $(\mathbf{w}^*)^\top \mathbf{x}_i + b^*$?

   (c) If $\alpha_i^* = C$ what are the possible range of values of $(\mathbf{w}^*)^\top \mathbf{x}_i + b^*$?