Problem Set – SVM -Key

1. Using the following 2-D dataset (x1 and x2 are the attributes and y is the class variable), find the linear SVM classifier. Do your optimization using the dual problem. Namely, provide an explicit expression for the dual optimization problem, solve it (compute the values of the various α_i's) and use the solution to compute the weights attached to the two attributes as well as the bias term.

Dataset:

X1	X2	Y
-1	1	+
1	-1	-

Since there are two points there will be two Lagragian multipliers one associated with each point.

Recall the dual formula

$$L = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j \, y_i \, y_j \, x_i \, x_j$$

Calculate the dot product of all points. You will be doing vector multiplication

$$\begin{matrix} -1 & 1 \end{matrix} * \begin{matrix} -1 \\ 1 \end{matrix} = -1 * -1 + 1 * 1 = 2$$

	P1	P2
P1	2	-2
P2	-2	2

Substituting the dot products and y- values we get..

$$L = \alpha_1 + \alpha_2 - \tfrac{1}{2}\,\alpha_1^2\,(1)(\,2) - \tfrac{1}{2}\alpha_1\alpha_2(-1)(-2) - \tfrac{1}{2}\alpha_1\alpha_2(-1)(-2) - \tfrac{1}{2}\,\alpha_2^2\,(1)(\,2)$$

$$L = \alpha_1 + \alpha_2 - \alpha_1^2 - 2\alpha_1\alpha_2 - \alpha_2^2$$

We already know what we get when we differentiate L w.r.t to w and b

$$w = \sum_i \alpha_i y_i x_i$$

$$\sum_i \alpha_i y_i = 0$$

By using the second equation we get $\alpha_1 - \alpha_2 = 0 => \alpha_1 = \alpha_2$

By differentiating L w.r.t to α_1 and α_2, we get

$$1 - 2\alpha_1 - 2\alpha_2 = 0 \ \ and\ 1 - 2\alpha_1 - 2\alpha_2 = 0$$

By using $\alpha_1 = \alpha_2$ in any one of the above equations we get $1 - 4\alpha_2 = 0 => \alpha_2 = \frac{1}{4} = \alpha_1$

By substituting the α_1 *and* α_2 in w equation we get

$$w = \begin{matrix} \frac{1}{4}(+1)(-1) + \frac{1}{4}(-1)(1) \\ \frac{1}{4}(+1)(1) + \frac{1}{4}(-1)(-1) \end{matrix} = \begin{matrix} -\frac{1}{2} \\ \frac{1}{2} \end{matrix}$$

Substitute w in any one of the initial conditions i.e.

$w.x + b = +1\ for\ positive\ support\ vector\ and\ w.x + b = -1\ for\ negative$

we will get b = 0.

2. A SVM is trained with the following data:

x_1	x_2	class
-1	-1	-1
1	1	1
0	2	1

Let α_1, α_2 *and* α_3 be the lagrangian multipliers for the three data points.

a. Using polynomial kernel of degree 2 what (dual) optimization problem needs to be solved in terms of the lagrangian multipliers in order to determine their values? The polynomial kernel of degree d is given by the equation

$K(x_i, x_j) = (1 + x_i^T x_j)^d$

$$where\ x_i\ and\ x_j\ are\ input\ vectors$$

$$maximize \sum_{i=1}^{3} \alpha_i - \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} \alpha_i \alpha_j y_i y_j (1 + x_i^T x_j)^2$$

subject to the constraint : $\alpha_1, \alpha_2, \alpha_3 \geq 0$ and $-\alpha_1 + \alpha_2 + \alpha_3 = 0$. The quantity $y_i y_j (1 + x_i^T x_j)^2$ for different values of i and j is given by the cells of the following matrix

$$\begin{pmatrix} 9 & -1 & -1 \\ -1 & 9 & 9 \\ -1 & 9 & 25 \end{pmatrix}$$

b. Let us say that we have solved the optimization problem and found that $\alpha_1 = \alpha_2 = \frac{1}{8} and\ \alpha_3 = 0$. Moreover b = 0. Can you tell me which of the data points are support vectors? Explain your answer.

The support vector are points 1 and 2 because α_1 and α_2 are greater than zero.

c. Assuming $\alpha_1 = \alpha_2 = \frac{1}{8}$, $\alpha_3 = 0$ and b = 0, how will the SVM classify the point (x_1 = -1, x_2 = 0)? Explain your answer?

The dot product of the kernel with the test point is (4, 0).

$$-\frac{1}{8}(4) + \frac{1}{8}(0) < 0$$

Therefore, the class is −1.

d. Assuming $\alpha_1 = \alpha_2 = \frac{1}{8}$, $\alpha_3 = 0$ and b = 0, how will the SVM classify the point (x_1 = 1, x_2 = 0)? Explain your answer?

The dot product of the kernel with the test point is (0, 4).

$$-\frac{1}{8}(0) + \frac{1}{8}(4) > 0$$

Therefore, the class is +1.

3. Consider the training data given below (Y is the class variable)

X	Y
-2	1
-1	-1
1	-1
2	1

a. Assume that you are using linear SVM. Let $\alpha_1, \alpha_2, \alpha_3, and\ \alpha_4$ be the lagrangian multipliers for the four data points. Write the precise expression for the lagrangian dual optimization problem that needs to be solved in order to compute the values of $\alpha_1, \dots \dots \alpha_4$ for the data set given above.

Recall the dual formula

$$L = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j$$

Calculate the dot product of all points. You will be doing vector multiplication
The quantity $y_i y_j x_i^T x_j$ for different values of i and j is given by the cells of the following matrix

$$\begin{pmatrix} 4 & -2 & 2 & -4 \\ -2 & 1 & -1 & 2 \\ 2 & -1 & 1 & -2 \\ -4 & 2 & -2 & 4 \end{pmatrix}$$

b. Do you think, you will get zero training error on this dataset if you use linear SVM? Explain your answer?

No, because the data is not linearly separable. Plot it on a line and see for yourself.

c. Now assume that you are using a quadratic kernel $\left(1 + x_i^T x_j\right)^2$. Again , let $\alpha_1, \alpha_2, \alpha_3, and\ \alpha_4$ be the lagrangian multipliers for the four data points. Write the precise expression for the lagrangian dual optimization problem that needs to be solved in order to compute the values of $\alpha_1, \dots \dots \alpha_4$ for the data set and the quadratic kernel given above.

$$maximize \sum_{i=1}^{4} \alpha_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j \left(1 + x_i^T x_j\right)^2$$

subject to the constraint : $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \geq 0$ and $\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 0$. The quantity $y_i y_j \left(1 + x_i^T x_j\right)^2$ for different values of i and j is given by the cells of the following matrix

$$\begin{pmatrix} 25 & -9 & -1 & 9 \\ -9 & 4 & 0 & -1 \\ -1 & 0 & 4 & -9 \\ 9 & -1 & -9 & 25 \end{pmatrix}$$

d. Do you think you will get zero training error on this data set with quadratic kernel, explain your answer?

Yes, the quadratic kernel will correctly classify this data because it will add the feature x^2 to the dataset. You can plot the points in 2-D with this added dimension and see for yourself.

4. Using the following 2-D dataset (x1 and x2 are the attributes and y is the class variable), find the linear SVM classifier. Do your optimization using the dual problem. Namely, provide an explicit expression for the dual optimization problem, solve it (compute the values of the various α_i's) and use the solution to compute the weights attached to the two attributes as well as the bias term.

Dataset:

X1	X2	Y

1	0	+
-1	2	-
0	-1	+

Since there are two points there will be two Lagragian multipliers one associated with each point.

Recall the dual formula

$$L = \sum_i \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i \alpha_j \, y_i \, y_j \, x_i \, x_j$$

Calculate the dot product of all points. You will be doing vector multiplication

$$1 \quad 0 * \begin{matrix} 1 \\ 1 \end{matrix} = 1 * 1 + 0 * 1 = 1$$

	P1	P2	P3
P1	1	-1	0
P2	-1	5	-2
P3	0	-2	1

Substituting the dot products and y- values we get..

$$L = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\alpha_1^2 (1)(1) - \frac{1}{2}\alpha_1\alpha_2(-1)(-1) - \frac{1}{2}\alpha_1\alpha_3(1)(0) - \frac{1}{2}\alpha_1\alpha_2(-1)(-1) - \frac{1}{2}\alpha_2^2 (1)(5) - \frac{1}{2}\alpha_2\alpha_3(-1)(-2) - \frac{1}{2}\alpha_3\alpha_1(1)(0) - \frac{1}{2}\alpha_3\alpha_2(-1)(-2) - \frac{1}{2}\alpha_3^2 (1)(1)$$

$$L = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\alpha_1^2 - \alpha_1\alpha_2 - \frac{5}{2}\alpha_2^2 - 2\alpha_2\alpha_3 - \frac{1}{2}\alpha_3^2$$

We already know what we get when we differentiate L w.r.t to w and b

$$w = \sum_i \alpha_i y_i x_i$$

$$\sum_i \alpha_i y_i = 0$$

By using the second equation we get $\alpha_1 - \alpha_2 + \alpha_3 = 0$, call it eq(1).

By differentiating L w.r.t to α_1, α_2 and α_3, we get following three equations

$$1 - \alpha_1 - \alpha_2 = 0$$

$$1 - \alpha_1 - 5\alpha_2 - 2\alpha_3 = 0$$

$$1 - 2\alpha_2 - \alpha_3 = 0$$

Call it eq(2), eq(3) and eq(4).

eq(2) and eq(3) contradict each other so we will use only eq(1), eq(2) and eq(4) to solve for all αs. so we

$$\alpha_1 = \frac{1}{2} \quad \alpha_2 = \frac{1}{2} \quad \alpha_3 = 0$$

By substituting the α_1 *and* α_2 in w equation we get

$$w = \begin{matrix} \frac{1}{2}(+1)(1) + \frac{1}{2}(-1)(-1) \\ \frac{1}{2}(+1)(0) + \frac{1}{2}(-1)(2) \end{matrix} = \begin{matrix} 1 \\ -1 \end{matrix}$$

Substitute w in any one of the initial conditions i.e.

$w.x + b = +1$ *for positive support vector and* $w.x + b = -1$ *for negative*

we will get b = 0 and b = 2 so average is b=1.

5. You are trying to use SVMs to build a classifier for a dataset. In the dataset, there are only a few positive training examples and a large number of negative examples. You have to modify the basic SVM dual problem such that none of the positive examples are misclassified but it is ok to misclassify few negative points. Introduce additional parameters and / or constraints in order to achieve this.

   Add the constraint

$$0 \leq \alpha_i \leq C$$

   for negative points only.