Problem 1 (Fitting an SVM by hand, 10pts)

For this problem you will solve an SVM by hand, relying on principled rules and SVM properties. For making plots, however, you are allowed to use a computer or other graphical tools.

Consider a dataset with the following 7 data points each with $x \in \mathbb{R}$ and $y \in \{-1, +1\}$:

$$\{(x_i, y_i)\}_{i=1}^{7} = \{(-3, +1), (-2, +1), (-1, -1), (0, +1), (1, -1), (2, +1), (3, +1)\}$$

Consider mapping these points to 2 dimensions using the feature vector $\phi(x) = (x, -\frac{8}{3}x^2 + \frac{2}{3}x^4)$. The hard margin classifier training problem is:

$$\begin{aligned} &\min_{\mathbf{w}, w_0} \frac{1}{2}\|\mathbf{w}\|_2^2 \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \phi(x_i) + w_0) \geq 1, \ \forall i \in \{1, \ldots, n\} \end{aligned}$$

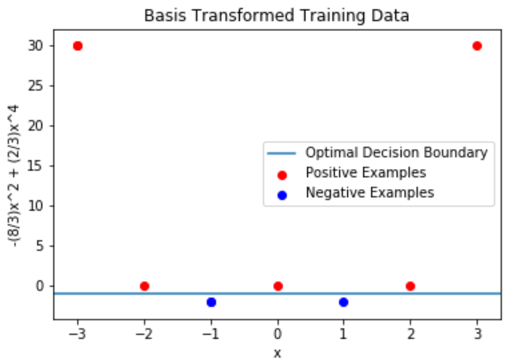
Make sure to follow the logical structure of the questions below when composing your answers, and to justify each step.

1. Plot the transformed training data in $\mathbb{R}^2$ and draw the optimal decision boundary of the max margin classifier. You can determine this by inspection (i.e. by hand, without actually doing any calculations).
2. What is the value of the margin achieved by the optimal decision boundary found in Part 1?
3. Identify a unit vector that is orthogonal to the decision boundary.
4. Considering the discriminant $h(\phi(x); \mathbf{w}, w_0) = \mathbf{w}^\top \phi(x) + w_0$, give an expression for *all possible* $(\mathbf{w}, w_0)$ that define the optimal decision boundary from 1.1. Justify your answer.

   Hint: The boundary is where the discriminant is equal to 0. Use what you know from 1.1 and 1.3 to solve for $\mathbf{w}$ in terms of w_0. (If you solve this problem in this way, then w_0 corresponds to your free parameter to describe the set of all possible $(\mathbf{w}, w_0)$.)
5. Consider now the training problem for this dataset. Using your answers so far, what particular solution to $\mathbf{w}$ will be optimal for the optimization problem?
6. What is the corresponding optimal value of w_0 for the $\mathbf{w}$ found in Part 5 (use your result from Part 4 as guidance)? Substitute in these optimal values and write out the discriminant function $h(\phi(x); \mathbf{w}, w_0)$ in terms of the variable x .
7. What are the support vectors of the classifier? Confirm that your solution in Part 6 makes the constraints above tight—that is, met with equality—for these support vectors.

Solution

1. From the plot we see that our data is linearly separable.



   The optimal decision boundary of $-\frac{8}{3}x^2 + \frac{2}{3}x = -1$ can be determined by inspection, as this decision boundary is the midpoint between the positive examples $\{x_i, y_i\} = (-2, +1), (0, +1), (2, +1)$ that are nearest to the negative examples $\{x_i, y_i\} = (-1, -1), (1, -1)$.

2. The value of the margin achieved by the optimal decision boundary is 1. This is because the distance from the optimal decision boundary to the positive examples $\{x_i, y_i\} = (-2, +1), (0, +1), (2, +1)$ that are nearest to the negative examples is 1. Similarly, the distance from the optimal decision boundary to the negative examples $\{x_i, y_i\} = (-1, -1), (1, -1)$ is 1.

3. An orthogonal unit vector would be $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. This vector is indeed orthogonal to the optimal decision boundary $-\frac{8}{3}x^2 + \frac{2}{3}x = -1$, because it is a vector that points vertically (and no deviations horizontally). It is a unit vector because its dimensions sum to 1.

4. Following the hint, we set the discriminant equal to 0:

$$h(\phi(x); \mathbf{w}, w_0) = \mathbf{w}^\top \phi(x) + w_0 = 0$$

   Rearranging:

$$\mathbf{w}^\top \phi(x) = -w_0$$

   In the basis space, the decision boundary is where $x_1 = 0$ and $x_2 = -\frac{8}{3}x^2 + \frac{2}{3}x^4 = -1$, where $\phi(x) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Our weight vector has two entries $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$. We also know that the orthogonal unit vector to the decision boundary is $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Since the weight vector $\mathbf{w}$ is also orthogonal to the decision boundary, we must have that $w_1 = 0$, and our weight matrix is now $\mathbf{w} = \begin{pmatrix} 0 \\ w_2 \end{pmatrix}$. Given that, we can directly compute the left hand side as:

$$\mathbf{w}^\top \phi(x) = \begin{pmatrix} 0 \\ w_2 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \end{pmatrix} = -w_2$$

   Substituting, we have

$$-w_2 = -w_0$$

   Therefore we have that $w_1 = 0$ and $w_2 = w_0$, or that $\mathbf{w} = \begin{pmatrix} 0 \\ w_0 \end{pmatrix}$.

5. Our feature vector, transformed by the basis function, is

$$\{\phi(x_i)\}_{i=1}^{7} = \{(-3, 30), (-2, 0), (-1, -2), (0, 0), (1, -2), (2, 0), (3, 30)\}$$

The optimal setting of $\mathbf{w}$ will minimize the margin from the decision boundary to the nearest correctly classified point. We know from part (1) that the data is linearly separable in the basis space. Using the hard margin classifier training problem:

$$\begin{aligned} &\min_{\mathbf{w}, w_0} \frac{1}{2}\|\mathbf{w}\|_2^2 \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \phi(x_i) + w_0) \geq 1, \ \forall i \in \{1, \ldots, n\} \end{aligned}$$

In part (4), we found that for $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 \\ w_0 \end{pmatrix}$, we have $w_1 = 0$ and $w_2 = w_0$. We can substitute to write our optimization problem in terms of one variable, w_0:

$$\begin{aligned} &\min_{w_0} \frac{1}{2}\| \begin{pmatrix} 0 \\ w_0 \end{pmatrix} \|_2^2 \\ \text{s.t.} \quad & y_i(\begin{pmatrix} 0 \\ w_0 \end{pmatrix} \phi(x_i) + w_0) \geq 1, \ \forall i \in \{1, \ldots, n\} \end{aligned}$$

or equivalently, for $\phi(x_i) = \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}$:

$$\begin{aligned} &\min_{w_0} \frac{1}{2} w_0^2 \\ \text{s.t.} \quad & y_i(w_0 x_{i2} + w_0) \geq 1, \ \forall i \in \{1, \ldots, n\} \end{aligned}$$

Rather than computing the constraint for all 7 data points, we can use our plot from part (1) to see that only five points will lie on our hard margin boundary. There are three that have the label $y_i = +1$: $(-2, 0), (0, 0)$, and $(2, 0)$. There are two that have the label $y_i = -1$: $(-1, -2)$ and $(1, -2)$. Observe that the datapoints on the hard margin either have $x_{i2} = 0$ for labels with $y_i = +1$, and $x_{i2} = -2$ for labels with $y_i = -1$. Therefore, to find the optimal $\mathbf{w}$, it suffices to determine the optimal w_0 since $w_2 = w$ and $w_1 = 0$. Moreover, we only need to consider $x_{i2} = 0$ and $x_{i2} = -2$, since our optimization problem (as expressed above) is independent of x_{i2}. We know that for these points, because they lie on the hard margin boundary, the constraint in the optimization problem will hold for equality. First, let's consider the points where $x_{i2} = 0$ and $y_i = +1$, and solve for the optimal w_0.

$$y_i(w_0 x_{i2} + w_0) = 1$$

$$1(w_0(0) + w_0) = 1 \rightarrow w_0 = 1$$

We find the same solution when we consider the points where $x_{i2} = -2$ and $y_i = -1$.

$$y_i(w_0 x_{i2} + w_0) = 1$$

$$-1(w_0(-2) + w_0) = 1 \rightarrow w_0 = 1$$

Therefore, our three optimal parameter values are $w_0 = 1$, $w_1 = 0$, and $w_2 = 1$, such that our optimal solution to $\mathbf{w}$ is $\mathbf{w} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

6. As we showed in part (4) and part (5), we found that the optimal $w_0 = w_2$, where w_2 is the second entry in the optimal solution to $\mathbf{w}$. We found $w_2 = 1$ in part (5), so therefore the corresponding

optimal value of w_0 is $w_0 = 1$. Substituting the optimal values of $w_0 = 1$, $w_1 = 0$, and $w_2 = 1$ into the discriminant $h(\phi(x); \mathbf{w}, w_0)$:

$$h(\phi(x); \mathbf{w}, w_0) = \mathbf{w}^\top \phi(x) + w_0 = w_1 x_1 + w_2 x_2 + w_0$$

$$h(\phi(x); \mathbf{w}, w_0) = x_2 + 1$$

Finally, we know that in our basis function, $x_2 = -\frac{8}{3}x^2 + \frac{2}{3}x^4$, so we may write the discriminant function in terms of the variable x as:

$$h(\phi(x); \mathbf{w}, w_0) = 1 - \frac{8}{3}x^2 + \frac{2}{3}x^4$$

7. As mentioned in part (5), and shown on the plot in part (1), there are five points that lie on the hard margin boundary: $\phi(x) = \{(2,0), (-1,-2), (0,0), (1,-2), (2,0)\}$. Or equivalently, these five points are $\{(x_i, y_i)\} = \{(-2,+1), (-1,-1), (0,+1), (1,-1), (2,+1)\}$. The constraint in the hard margin classifier training problem is:

$$y_i(\mathbf{w}^\top \phi(x_i) + w_0) \geq 1, \forall i \in \{1, \dots, n\}$$

Each of the five points that lie on the hard margin boundary are candidates as support vectors. For all five points, they have a margin of 1, which is equivalent to:

$$\frac{1}{||\mathbf{w}||_2} = \frac{1}{\sqrt{0^2 + 1^2}} = \frac{1}{1} = 1$$

Therefore, to check if each of these five points are support vectors of the classifier, we check that these point will lead the hard margin constraint to bind, such that

$$y_i(\mathbf{w}^\top \phi(x_i) + w_0) = 1$$

for a point i on the hard margin boundary. Observe that our determinant is $h(\phi(x); \mathbf{w}, w_0) = \mathbf{w}^\top \phi(x_i) + w_0$, which we expressed in part (6) as $h(\phi(x); \mathbf{w}, w_0) = 1 - \frac{8}{3}x^2 + \frac{2}{3}x^4$. Therefore, we can rewrite the hard margin constraint in terms of x:

$$y_i(1 - \frac{8}{3}x^2 + \frac{2}{3}x^4) = 1$$

- Checking $(x_i, y_i) = (-2, +1)$:

$$y_i(1 - \frac{8}{3}x^2 + \frac{2}{3}x^4) = 1(1 - \frac{8}{3}(-2)^2 + \frac{2}{3}(-2)^4) = 1(1 - 0) = 1$$

  Therefore the point $(x_i, y_i) = (-2, +1)$ is a support vector of the classifier.

- Checking $(x_i, y_i) = (0, +1)$:

$$y_i(1 - \frac{8}{3}x^2 + \frac{2}{3}x^4) = 1(1 - \frac{8}{3}0^2 + \frac{2}{3}0^4) = 1(1 - 0) = 1$$

  Therefore the point $(x_i, y_i) = (0, +1)$ is a support vector of the classifier.

- Checking $(x_i, y_i) = (2, +1)$:

$$y_i(1 - \frac{8}{3}x^2 + \frac{2}{3}x^4) = 1(1 - \frac{8}{3}2^2 + \frac{2}{3}2^4) = 1(1 - 0) = 1$$

  Therefore the point $(x_i, y_i) = (2, +1)$ is a support vector of the classifier.

- Checking $(x_i, y_i) = (-1, -1)$:

$$y_i(1 - \frac{8}{3}x^2 + \frac{2}{3}x^4) = -1(1 - \frac{8}{3}(-1)^2 + \frac{2}{3}(-1)^4) = -1(-1) = 1$$

  Therefore the point $(x_i, y_i) = (-1, -1)$ is a support vector of the classifier.

- Checking $(x_i, y_i) = (1, -1)$:

$$y_i(1 - \frac{8}{3}x^2 + \frac{2}{3}x^4) = -1(1 - \frac{8}{3}1^2 + \frac{2}{3}1^4) = -1(-1) = 1$$

  Therefore the point $(x_i, y_i) = (1, -1)$ is a support vector of the classifier.

We find that all five points that lie on the hard margin boundary are the support vectors of the classifier, that is $i \in n$ such that $\{(x_i, y_i)\} = \{(-2, +1), (-1, -1), (0, +1), (1, -1), (2, +1)\}$.