Machine Learning (SS 23)
Assignment 09: Support Vector Machine

Team Members:

- Likhit Jain, 3678905, M.Sc. Computer Science
- Tareq Abu El Komboz, 3405686, M.Sc. Informatik
- Serge Kotchourko, 3309449, M.Sc. Informatik

1. Support Vector Machine

Question 1 Which of the decision boundaries (a), (b) or (c) shown on the figure would be the resulting decision boundary of linear SVM? Show your calculations. When answering this question, no need to solve by optimizing the SVM objective function.

Answer We are looking for the hyperplane that maximizes the minimal distance to each class.

- $\delta(h, \begin{pmatrix} 0 \\ 0 \end{pmatrix}) = \delta(h, \begin{pmatrix} 4 \\ 0 \end{pmatrix}) = 1.5 = \delta(h, \begin{pmatrix} -1 \\ -3 \end{pmatrix})$, the other point $\begin{pmatrix} -1 \\ -4 \end{pmatrix}$ is furhter away (visually) and therfore we did not calculate the distance as it is not relevant
- $\delta(h, \begin{pmatrix} 0 \\ 0 \end{pmatrix})) = 1.5 = \delta(h, \begin{pmatrix} -1 \\ -3 \end{pmatrix}) = \delta(h, \begin{pmatrix} -1 \\ -4 \end{pmatrix})$, the other point $\begin{pmatrix} 4 \\ 0 \end{pmatrix}$ is furhter away (visually) and therfore we did not calculate the distance as it is not relevant
- $\delta(h, \begin{pmatrix} 0 \\ 0 \end{pmatrix})) = \sqrt{0.5^2 + 1.5^2}) = 1.58 = \delta(h, \begin{pmatrix} -1 \\ -3 \end{pmatrix})$, the other two points $\begin{pmatrix} 4 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ -4 \end{pmatrix}$ are furhter away (visually) and therfore we did not calculate the distance as it is not relevant

Question 2 What are the support vectors based on your answer in (1.)?

Answer In this case every class has one support vector. For the circle class, the support vector is $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and for the cross class it is $\begin{pmatrix} -1 \\ -3 \end{pmatrix}$.

Question 3 How would adding a training point in location $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ to the dataset that belongs to the circle class change the decision boundary?

Answer Adding a training point in location $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ to the dataset that belongs to the circle class would not change the decision boundary because $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is further away from all three decision boundaries than the current support vector of the circle class (namely $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$).

2. Perceptron

Question For the data given below, apply the perceptron algorithm (slide 10) to find the weight vector w of the decision boundary. Show the output of each iteration till convergence. Assume that the weight vector is initialized $w^{(0)} = \begin{pmatrix}1\\3\end{pmatrix}$ with $\alpha = 0.1$ (there is no bias).
For class cross, $t = 1$
For class circle, $t = -1$

Answer We have to calculate $\hat{f}(x_i)y_i$ and check whether there is an i for which the expression is less than zero.
Iteration 1 - $w^{(0)} = \begin{pmatrix}1\\3\end{pmatrix}$

$$\hat{f}\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = w^T\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = \begin{pmatrix}1\\3\end{pmatrix}^T\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = (1\cdot(-3)+3\cdot3)\cdot(-1) = (-3+9)\cdot(-1) = -6 < 0$$

$$\hat{f}\begin{pmatrix}-4\\1\end{pmatrix}\cdot(-1) = w^T\begin{pmatrix}-4\\1\end{pmatrix}\cdot(-1) = \begin{pmatrix}1\\3\end{pmatrix}^T\begin{pmatrix}-4\\1\end{pmatrix}\cdot(-1) = (1\cdot(-4)+3\cdot1)\cdot(-1) = (-4+3)\cdot(-1) = 1 > 0$$

$$\hat{f}\begin{pmatrix}2\\3\end{pmatrix}\cdot(+1) = w^T\begin{pmatrix}2\\3\end{pmatrix} = \begin{pmatrix}1\\3\end{pmatrix}^T\begin{pmatrix}2\\3\end{pmatrix} = (1\cdot2+3\cdot3) = (2+9) = 11 > 0$$

$$\hat{f}\begin{pmatrix}4\\1\end{pmatrix} = w^T\begin{pmatrix}4\\1\end{pmatrix}\cdot = \begin{pmatrix}1\\3\end{pmatrix}^T\begin{pmatrix}4\\1\end{pmatrix} = (1\cdot4+3\cdot1) = (4+3) = 7 > 0$$

$\hat{f}(x_i)y_i$ is smaller than zero for $x_i = \begin{pmatrix}-3\\3\end{pmatrix}$ and $y_i = -1$

$$w^{(1)} = w^{(0)} + \alpha x_i y_i = \begin{pmatrix}1\\3\end{pmatrix} + 0.1\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = \begin{pmatrix}1\\3\end{pmatrix} - \begin{pmatrix}-0.3\\0.3\end{pmatrix} = \begin{pmatrix}1.3\\2.7\end{pmatrix}$$

Iteration 2 - $w^{(1)} = \begin{pmatrix}1.3\\2.7\end{pmatrix}$

$$\hat{f}\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = w^T\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = \begin{pmatrix}1.3\\2.7\end{pmatrix}^T\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = (1.3\cdot(-3)+2.7\cdot3)\cdot(-1) = -4.2 < 0$$

$$\hat{f}\begin{pmatrix}-4\\1\end{pmatrix}\cdot(-1) = w^T\begin{pmatrix}-4\\1\end{pmatrix}\cdot(-1) = \begin{pmatrix}1.3\\2.7\end{pmatrix}^T\begin{pmatrix}-4\\1\end{pmatrix}\cdot(-1) = (1.3\cdot(-4)+2.7\cdot1)\cdot(-1) = 2.5 > 0$$

$$\hat{f}\begin{pmatrix}2\\3\end{pmatrix}\cdot(+1) = w^T\begin{pmatrix}2\\3\end{pmatrix} = \begin{pmatrix}1.3\\2.7\end{pmatrix}^T\begin{pmatrix}2\\3\end{pmatrix} = (1.3\cdot2+2.7\cdot3) = 10.7 > 0$$

$$\hat{f}\begin{pmatrix}4\\1\end{pmatrix} = w^T\begin{pmatrix}4\\1\end{pmatrix}\cdot = \begin{pmatrix}1.3\\2.7\end{pmatrix}^T\begin{pmatrix}4\\1\end{pmatrix} = (1.3\cdot4+2.7\cdot1) = 7.9 > 0$$

$\hat{f}(x_i)y_i$ is smaller than zero for $x_i = \begin{pmatrix}-3\\3\end{pmatrix}$ and $y_i = -1$

$$w^{(2)} = w^{(1)} + \alpha x_i y_i = \begin{pmatrix}1.3\\2.7\end{pmatrix} + 0.1\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = \begin{pmatrix}1.3\\2.7\end{pmatrix} - \begin{pmatrix}-0.3\\0.3\end{pmatrix} = \begin{pmatrix}1.6\\2.4\end{pmatrix}$$

Iteration 3 - $w^{(2)} = \begin{pmatrix} 1.6 \\ 2.4 \end{pmatrix}$

$$\hat{f}\begin{pmatrix} -3 \\ 3 \end{pmatrix} \cdot (-1) = w^T \begin{pmatrix} -3 \\ 3 \end{pmatrix} \cdot (-1) = \begin{pmatrix} 1.6 \\ 2.4 \end{pmatrix}^T \begin{pmatrix} -3 \\ 3 \end{pmatrix} \cdot (-1) = (1.6 \cdot (-3) + 2.4 \cdot 3) \cdot (-1) = -2.4 < 0$$

$$\hat{f}\begin{pmatrix} -4 \\ 1 \end{pmatrix} \cdot (-1) = w^T \begin{pmatrix} -4 \\ 1 \end{pmatrix} \cdot (-1) = \begin{pmatrix} 1.6 \\ 2.4 \end{pmatrix}^T \begin{pmatrix} -4 \\ 1 \end{pmatrix} \cdot (-1) = (1.6 \cdot (-4) + 2.4 \cdot 1) \cdot (-1) = 4 > 0$$

$$\hat{f}\begin{pmatrix} 2 \\ 3 \end{pmatrix} \cdot (+1) = w^T \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1.6 \\ 2.4 \end{pmatrix}^T \begin{pmatrix} 2 \\ 3 \end{pmatrix} = (1.6 \cdot 2 + 2.4 \cdot 3) = 10.4 > 0$$

$$\hat{f}\begin{pmatrix} 4 \\ 1 \end{pmatrix} = w^T \begin{pmatrix} 4 \\ 1 \end{pmatrix} \cdot = \begin{pmatrix} 1.6 \\ 2.4 \end{pmatrix}^T \begin{pmatrix} 4 \\ 1 \end{pmatrix} = (1.6 \cdot 4 + 2.4 \cdot 1) = 8.8 > 0$$

$\hat{f}(x_i)y_i$ is smaller than zero for $x_i = \begin{pmatrix} -3 \\ 3 \end{pmatrix}$ and $y_i = -1$

$$w^{(3)} = w^{(2)} + \alpha x_i y_i = \begin{pmatrix} 1.6 \\ 2.4 \end{pmatrix} + 0.1 \begin{pmatrix} -3 \\ 3 \end{pmatrix} \cdot (-1) = \begin{pmatrix} 1.6 \\ 2.4 \end{pmatrix} - \begin{pmatrix} -0.3 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 1.3 \\ 2.7 \end{pmatrix}$$

Iteration 4 - $w^{(3)} = \begin{pmatrix} 1.9 \\ 2.1 \end{pmatrix}$

$$\hat{f}\begin{pmatrix} -3 \\ 3 \end{pmatrix} \cdot (-1) = w^T \begin{pmatrix} -3 \\ 3 \end{pmatrix} \cdot (-1) = \begin{pmatrix} 1.9 \\ 2.1 \end{pmatrix}^T \begin{pmatrix} -3 \\ 3 \end{pmatrix} \cdot (-1) = (1.9 \cdot (-3) + 2.1 \cdot 3) \cdot (-1) = -0.6 < 0$$

$$\hat{f}\begin{pmatrix} -4 \\ 1 \end{pmatrix} \cdot (-1) = w^T \begin{pmatrix} -4 \\ 1 \end{pmatrix} \cdot (-1) = \begin{pmatrix} 1.9 \\ 2.1 \end{pmatrix}^T \begin{pmatrix} -4 \\ 1 \end{pmatrix} \cdot (-1) = (1.9 \cdot (-4) + 2.1 \cdot 1) \cdot (-1) = 5.5 > 0$$

$$\hat{f}\begin{pmatrix} 2 \\ 3 \end{pmatrix} \cdot (+1) = w^T \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1.9 \\ 2.1 \end{pmatrix}^T \begin{pmatrix} 2 \\ 3 \end{pmatrix} = (1.9 \cdot 2 + 2.1 \cdot 3) = 10.1 > 0$$

$$\hat{f}\begin{pmatrix} 4 \\ 1 \end{pmatrix} = w^T \begin{pmatrix} 4 \\ 1 \end{pmatrix} \cdot = \begin{pmatrix} 1.9 \\ 2.1 \end{pmatrix}^T \begin{pmatrix} 4 \\ 1 \end{pmatrix} = (1.9 \cdot 4 + 2.1 \cdot 1) = 9.7 > 0$$

$\hat{f}(x_i)y_i$ is smaller than zero for $x_i = \begin{pmatrix} -3 \\ 3 \end{pmatrix}$ and $y_i = -1$

$$w^{(4)} = w^{(3)} + \alpha x_i y_i = \begin{pmatrix} 1.9 \\ 2.1 \end{pmatrix} + 0.1 \begin{pmatrix} -3 \\ 3 \end{pmatrix} \cdot (-1) = \begin{pmatrix} 1.9 \\ 2.1 \end{pmatrix} - \begin{pmatrix} -0.3 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 1.3 \\ 2.7 \end{pmatrix}$$

Iteration 5 - $w^{(4)} = \begin{pmatrix} 2.2 \\ 1.8 \end{pmatrix}$

$$\hat{f}\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = w^T\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = \begin{pmatrix}2.2\\1.8\end{pmatrix}^T\begin{pmatrix}-3\\3\end{pmatrix}\cdot(-1) = (2.2\cdot(-3)+1.8\cdot 3)\cdot(-1) = 1.2 > 0$$
$$\hat{f}\begin{pmatrix}-4\\1\end{pmatrix}\cdot(-1) = w^T\begin{pmatrix}-4\\1\end{pmatrix}\cdot(-1) = \begin{pmatrix}2.2\\1.8\end{pmatrix}^T\begin{pmatrix}-4\\1\end{pmatrix}\cdot(-1) = (2.2\cdot(-4)+1.8\cdot 1)\cdot(-1) = 7 > 0$$
$$\hat{f}\begin{pmatrix}2\\3\end{pmatrix}\cdot(+1) = w^T\begin{pmatrix}2\\3\end{pmatrix} = \begin{pmatrix}2.2\\1.8\end{pmatrix}^T\begin{pmatrix}2\\3\end{pmatrix} = (2.2\cdot 2+1.8\cdot 3) = 9.8 > 0$$
$$\hat{f}\begin{pmatrix}4\\1\end{pmatrix} = w^T\begin{pmatrix}4\\1\end{pmatrix}\cdot = \begin{pmatrix}2.2\\1.8\end{pmatrix}^T\begin{pmatrix}4\\1\end{pmatrix} = (2.2\cdot 4+1.8\cdot 1) = (8.8+1.8) = 10.6 > 0$$

$$\forall i : \hat{f}(x_i)y_i > 0 \rightarrow \hat{f}(x_i) = w^T x_i = \begin{pmatrix}2.2\\1.8)\end{pmatrix}^T x_i$$

3. Polynomial Kernel

Question The second-order polynomial kernel for a two-dimensional vector $x_i = \begin{pmatrix}x_{i1}\\x_{i2}\end{pmatrix}^T$ is defined as:
$$\phi(x_i) = \begin{pmatrix}x_{i1}^2\\\sqrt{2}x_{i1}x_{i2}\\x_{i2}^2\end{pmatrix}$$
Show that the mapping of the two-dimensional vector to three dimensions is not necessary for calculating the scalar product $\langle\phi(x_i),\phi(x_j)\rangle$. (Note: Transform the equation such that it only uses the scalar product of two-dimensional vectors.)

Answer

$$\begin{aligned}
\langle\phi(x_i),\phi(x_j)\rangle &= \langle\begin{pmatrix}x_{i1}^2\\\sqrt{2}x_{i1}x_{i2}\\x_{i2}^2\end{pmatrix},\begin{pmatrix}x_{j1}^2\\\sqrt{2}x_{j1}x_{j2}\\x_{j2}^2\end{pmatrix}\rangle\\
&= x_{i1}^2\cdot x_{j1}^2 + 2x_{i1}x_{i2}\cdot x_{j1}x_{j2} + x_{i2}^2\cdot x_{j2}^2\\
&= (x_{i1}\cdot x_{j1} + x_{i2}\cdot x_{j2})(x_{i1}\cdot x_{j1} + x_{i2}\cdot x_{j2})\\
&= (x_{i1}\cdot x_{j1} + x_{i2}\cdot x_{j2})\\
&= \langle\begin{pmatrix}x_{i1}\\x_{i2}\end{pmatrix},\begin{pmatrix}x_{j1}\\x_{j2}\end{pmatrix}\rangle\langle\begin{pmatrix}x_{i1}\\x_{i2}\end{pmatrix},\begin{pmatrix}x_{j1}\\x_{j2}\end{pmatrix}\rangle\\
&= \langle x_i,x_j\rangle\langle x_i,x_j\rangle
\end{aligned}$$

4. Gaussian Kernel

Question Slide 69 mentions that the Gaussian kernel, also called Radial Basis Function (RBF), projects to an infinite dimensional feature space. Give an intuition on why this is the case and prove it. (Note: Use the Taylor expansion over e^x to show that the Gaussian kernel is an infinite sum over the polynomial kernels.)

Answer The Gaussian kernel, also known as the Radial Basis Function (RBF) kernel, exhibits the property of projecting the input data into an infinite-dimensional feature space.

The RBF kernel defines the similarity between two points in the input space based on their Euclidean distance. It assigns a higher similarity value when the points are closer and decreases as the distance increases. This behavior is captured by the Gaussian function, which forms the basis of the RBF kernel.

The Gaussian function, defined as $e^{-\gamma||x-x'||^2}$, where γ is a positive parameter, has a bell-shaped curve. This curve can be thought of as an infinite sum of polynomial terms. Each term of the sum represents a polynomial kernel of a different degree, and together, they form the infinite-dimensional feature space.

$$
\begin{aligned}
k(x, x') &= e^{-\frac{||x-x'||^2}{2\sigma^2}} \\
&= \sum_{n=0}^{\infty} \frac{-(\frac{||x-x'||^2}{2\sigma^2})^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{(||x-x'||^2)^n}{n! \cdot (-2\sigma^2)^n} \\
&= \sum_{n=0}^{\infty} \frac{(x^T x - x'^T x')^n}{n! \cdot (-2\sigma^2)^n} \\
&= 1 - \frac{x^T x - x'^T x'}{2\sigma^2} + \frac{(x^T x - x'^T x')^2}{8\sigma^4} - \frac{(x^T x - x'^T x')^3}{48\sigma^6} + \dots
\end{aligned}
$$

As we can see, each value of n generates a polynomial kernel with a different degree. The infinite sum of these polynomial kernels forms the Gaussian kernel.