Mustafa Darras 50106030
Ekrem Yilmaz 501108034
Alden Shin-Culhane 501036054

Lab7 - Support Vector Machine

Tak1:

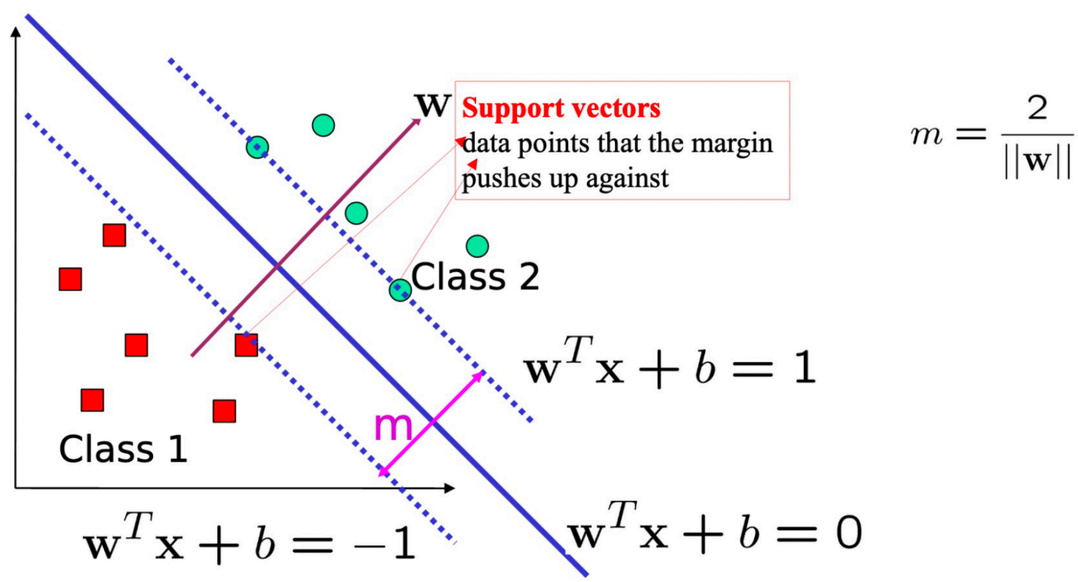
Question 1.1: In SVM what is the meaning of margin? Which are the equations of the two margin hyperplans H+ and H- ? (1 Mark)

Margin: In Support Vector Machines (SVM), the distance between the line and the closest data points is referred to as the margin. A larger margin typically implies better generalization to unseen data.

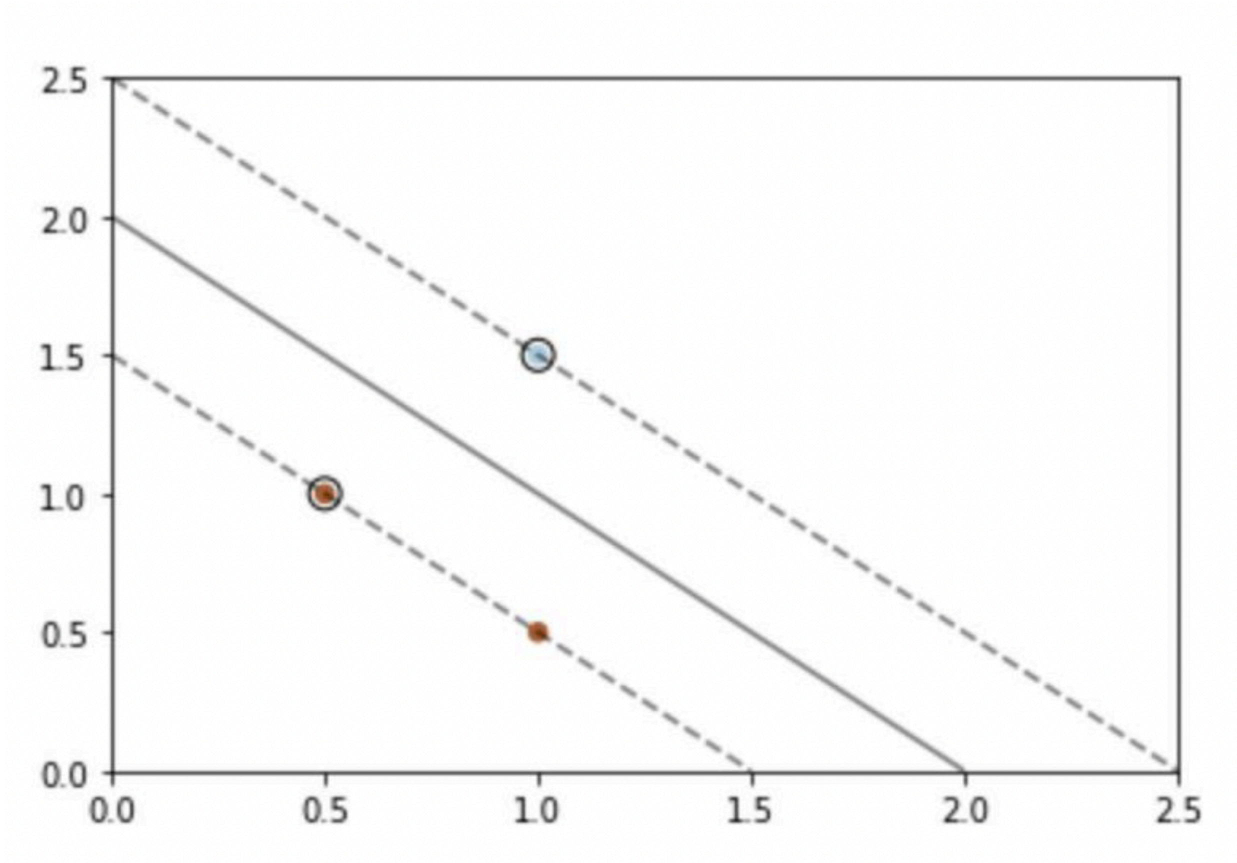
Equations of the margin hyperplanes as shown in the lecture slide:

H^+: $w^T x + b = 1$

H^-: $w^T x + b = -1$



Question 1.2: Consider the three linearly separable two-dimensional input vectors in the following figure. Find the linear SVM that optimally separates the classes by maximizing the margin. (1 Mark)



Given three linearly separable points, the SVM will find the hyperplane that maximizes the margin between the two classes.
As in the lecture slide, the optimal hyperplane can be found by solving the optimization problem:

Optimal Hyperplane (3)

- Finally, the problem of maximizing **M** is equivalent to the problem of minimizing a function $L(\beta)$ subject to some constraints. The constraints model the requirement for the hyperplane to classify correctly all the training examples x_i

$$\min_{\beta,\beta_0} L(\beta) = \frac{1}{2}||\beta||^2 \text{ subject to } y_i(\beta^T x_i + \beta_0) \geq 1 \ \forall i,$$

- Where y_i represents each of the labels of the training examples

Top circle $x(1) = (1.0, 1.5)$, label $y(1) = +1$.
Bottom circle $x(2) = (1.0, 0.5)$, label $y(2) = -1$.
Middle circle $x(3) = (0.5, 1.0)$, label $y(3) = -1$.

First we write the Margin Constraints

For the positive support vector:
$w1(1.0) + w2(1.5) + b = +1 \quad (1)$

For the negative support vector:
$w1(1.0) + w2(0.5) + b = -1 \quad (2)$

Subtract (2) from (1):
$(1.5 - 0.5)w2 = 2 \Longrightarrow w2 = 2$.

Substitute $w2 = 2$ into (1):
$w1 + 1.5(2) + b = 1 \Longrightarrow w1 + 3 + b = 1$,

which gives:
$w1 + b = -2. \quad (3)$

For the middle point $(0.5, 1.0)$ with $y = -1$:
$0.5w1 + 2(1.0) + b = -1$.

Substitute $b = -2 - w1$ from (3):
$0.5w1 + 2 - 2 - w1 = -1 \Longrightarrow -0.5w1 = -1$,

so
$w1 = 2$.

Then, from (3), $b = -2 - 2 = -4$.

The Final SVM parameters are:
- Weight vector: $w = (2, 2)$
- Bias: $b = -4$

The decision boundary is:
$2x1 + 2x2 - 4 = 0 \Longrightarrow x1 + x2 = 2$.

The margin hyperplanes are:
- $2x1 + 2x2 - 4 = +1 \Rightarrow x1 + x2 = 2.5$,
- $2x1 + 2x2 - 4 = -1 \Rightarrow x1 + x2 = 1.5$.

Question 1.3: What is a kernel function? (1 Mark)

Kernel Function: A kernel function is used in SVM to transform the input data into a higher-dimensional space where it becomes linearly separable. The kernel function computes the inner product of two vectors in this transformed space without explicitly computing the transformation. Common kernel functions include:

The dot-product is called the kernel and can be re-written as: K(x, xi) = sum(x * xi)

The equation for making a prediction for a new input using the dot product between the input (x) and each support vector (xi)is calculated as follows: f(x) = B0 + sum(ai * (x,xi))

Task 2:

Compare Neural Network and SVM in Classification of heart disease data set in Python language. You can use the sklearn Python library to implement both Neural Networks and SVM. For SVM, build the model by changing the different kernels such as Linear, Gaussian and Sigmoid and note down the model accuracy. Similarly, use Stochastic Gradient Descent and Adam Gradient Descent to build the multi-layer Neural Network and note down the model accuracy for each. Finally, tell us which model performs better and why? (5 Marks)

We conducted an experiment to compare SVM and Artificial Neural Networks (ANNs) in classifying heart disease patients using different configurations.

Experimental Setup:

- Dataset: Heart disease dataset (preprocessed: missing values removed, categorical data converted to numerical, and features normalized).
- Models Used:
  - SVM with three different kernels: Linear, Gaussian (RBF), and Sigmoid.
  - ANN with two optimizers: Stochastic Gradient Descent (SGD) and Adam.
- Performance Metric: Accuracy Score

Results:

Model Type	Configuration	Accuracy
SVM	Linear Kernel	**61.67%**
SVM	Gaussian Kernel	58.33%
SVM	Sigmoid Kernel	58.33%
ANN	SGD Optimizer	**61.67%**
ANN	Adam Optimizer	58.33%

Observations:

- SVM with a linear kernel and ANN with SGD performed the best (61.67%).
- Gaussian and Sigmoid kernels in SVM, and Adam optimizer in ANN, underperformed slightly (58.33%).
- When data is linearly separable, SVM with a linear kernel is generally the best choice.
- ANNs can generalize better with more data and hyperparameter tuning.

Conclusion:

Both SVM (Linear Kernel) and ANN (SGD Optimizer) performed equally well in this task. If the dataset had non-linearly separable patterns, a Gaussian (RBF) kernel in SVM or a deeper ANN with fine-tuned hyperparameters might have provided better results.