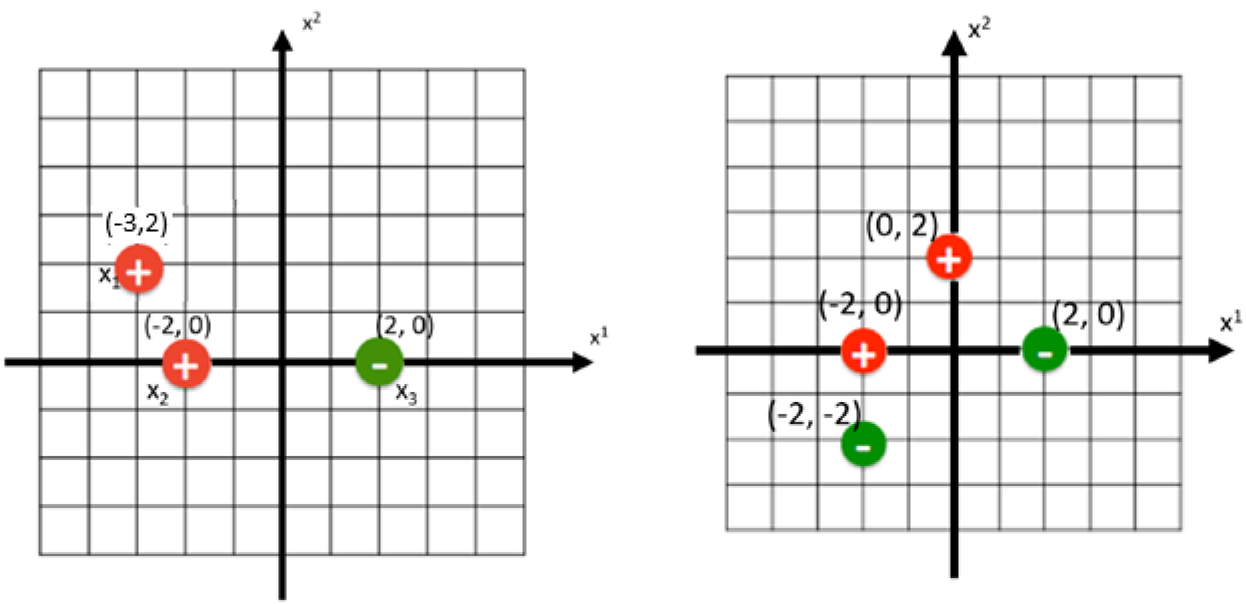
CSE575 HW02, Friday, 10/27/2023, Due: Wednesday, 11/08/2023

1 Support Vector Machine [20 points]

In Figure 1a and Figure 1b, we are given some data points in 2-D space and we aim to train a hard margin linear SVM classifier. The positive data points (with "+" sign in the figure) are labeled as 1. The negative data points (with "-" sign in the figure) are labeled as -1. The length of each cell in the grid is 1.



(a) Part A

(b) Part B

Figure 1: Problem 1: SVM

1) [8 points] For each case in Figure 1, circle the support vectors of your SVM classifier. What is the size of the margin of your SVM classifier in each case?

1.1) ANS:

a) The SVM classifier would likely use the points x2 (-2, 0) and x3 (2, 0) as support vectors, as they would be closest to the decision boundary. The point (-3, 2) would not be a support vector as it is further away from the decision boundary.

Hence, Margin is 2.

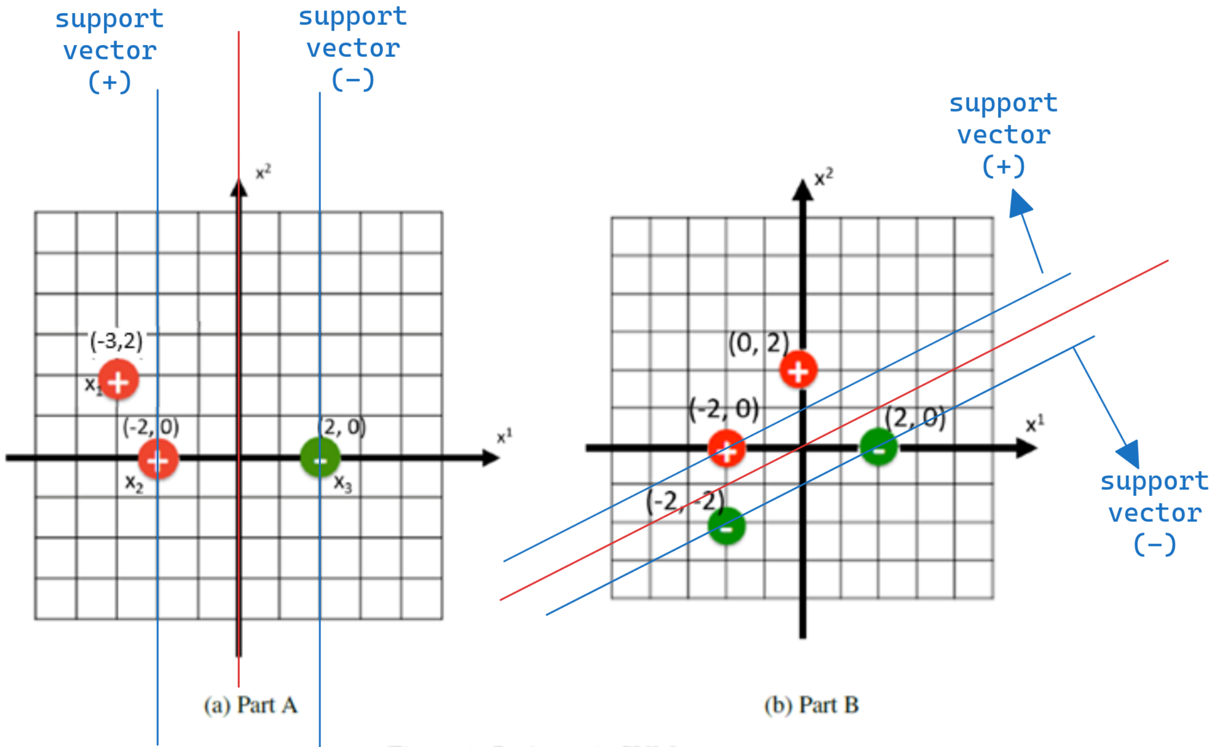


Figure 1: Problem 1: SVM

b) The SVM classifier would likely use the points (-2, -2) and (2, 0) as support vectors for negative sign and (-2, 0) as positive sign, as they would be closest to the decision boundary. The margin is $\frac{2}{\sqrt{5}}$.

---

2) [6 points] Suppose we consider the data points shown in Figure 1a, but point x_1 is moved from (-3,2) to (-2,2). Write down the (convex) primal optimization formulation for this problem. Write down the dual optimization formulation for this problem.

1.2) ANS:

a) **For (convex) primal optimization formulation,**
   we will use following problem:
   $min_{\{w,b\}} \frac{1}{2} ||w||^2 \ s.t. \ t_n(W^T \phi(X_n) + b) \geq 1, \ n = 1, 2, ..., N$
   we will put following coordinate in above equation in $W = [w_1, w_2]^T$ format, which is
   $x_1 = [-2, 2]^T$

$x_2 = [-2, 0]^T$

$x_3 = [2, 0]^T$

So, (convex) primal optimization formulation for given problem would look like

$min\ \frac{1}{2}(w_1 + w_2)^2$

$s.t.$

$$-2w_1 + 2w_2 + b \geq 1$$
$$-2w_1 + b \geq 1$$
$$-2w_1 - b \geq 1$$

b) For dual optimization formulation

we will use lagrangian function which given as follow:

$$L(w, b, a) = \frac{1}{2}||w||^2 - \sum_{n=1}^{N} a_n\{t_n(w^T\phi(x_n) + b) - 1\}$$

Let's put value in above formula

$$L(w, b, a) = \frac{1}{2}(w_1^2 + w_2^2) - a_1(-2w_1 + 2w_2 + b - 1) - a_2(-2w_1 + b - 1)$$

Put training coordinates to dual problem formula:

$$L(a) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m \phi^T(X_n)\phi(X_m)$$

$s.t.\ a_n > 0;\ n = 1, 2, 3,\ldots N$

$$\sum_{n=1}^{N} a_n t_n = 0$$

So, max value become

$$-\frac{1}{2}(8a_1^2 + 4a_2^2 + 4a_3^2 + 8a_1a_2 + 8a_1a_2 + 8a_2a_3) + a_1 + a_2 + a_3$$

$s.t.\ a_1 \geq 0,\ a_2 \geq 0,\ a_3 \geq 0$

$-a_1 - a_2 + a_3 = 0$

3) [6 points] Consider data points shown in Figure 2a, can you draw the decision boundary of your SVM classifier in this case? Suppose we add a new data sample at position (-3, 5) (as shown in Figure 2b). How will the decision boundary change if we add this new data point?

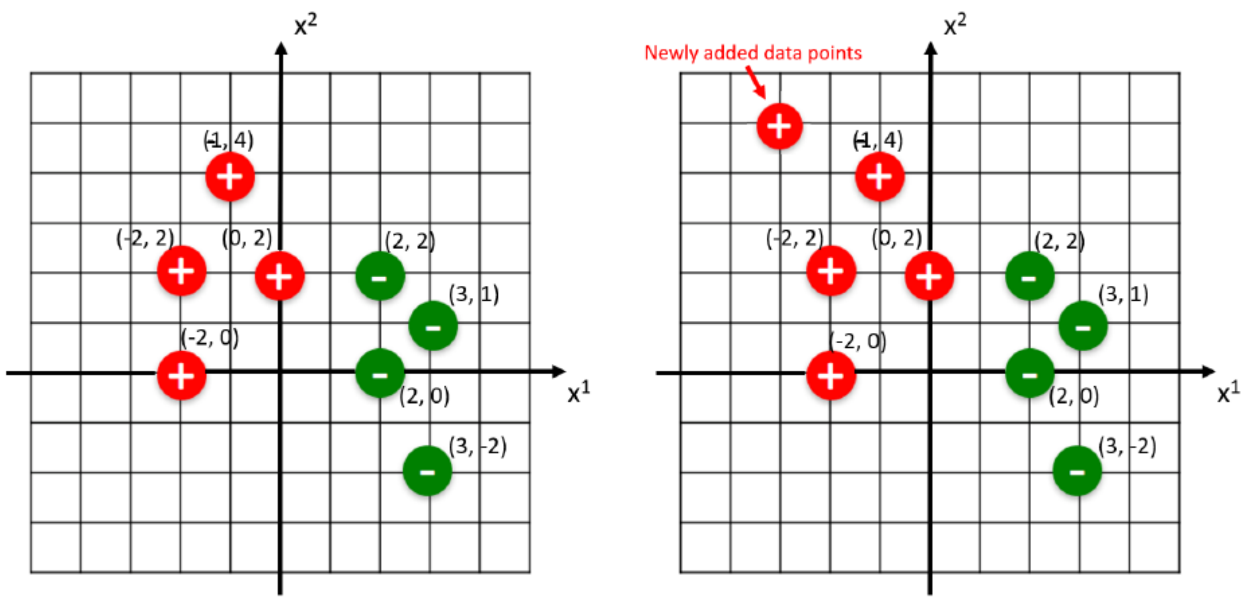
1.3) ANS:

The decision boundary is defined as x_1=1, indicating that data points with $x < 1$ are assigned label 1, while those with $x > 1$ receive label -1.

The addition of the new data point at (-3, 5) does not alter the decision boundary, as it resides on the left side of the margin.

---

2 K-means [40 points]



(a) before adding new data point

(b) after adding new data points

Figure 2: Problem 1: Adding a new point

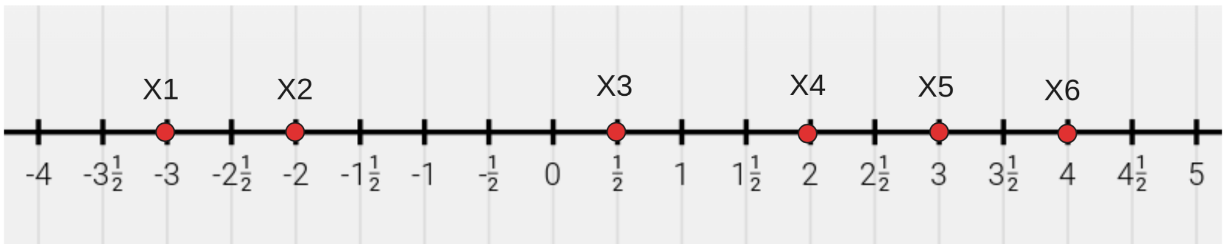
Given N data points x_i $(i = 1, 2, ...N)$, K-means will group them into K clusters by minimizing the loss/cost/distortion function $J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2$, where μ_k is the cluster center of the kth cluster, and $r_{nk} = 1$ if x_n belongs to the kth cluster and $r_{nk} = 0$ otherwise. In this question, we will use the following iterative procedure.

- Initialize the cluster centers μ_k, $k = 1, 2, ...K$

- Iterate until convergence
  - Step 1: Update the cluster assignments r_{nk} for each data point x_n
  - Step 2: Update the cluster center μ_k for each cluster k

1) [5 points] Given 6 data points in 1-D space $x_1 =- 3,\ x_2 =- 1,\ x_3 = 0.5,\ x_4 = 2,\ x_5 = 3,\ x_6 = 4$. Plot these six data points in 1-D space.

2.1) ANS:



---

2) [10 points] Suppose the initial cluster centers are $\mu_1 =- 1$ and $\mu_2 = 3$. If we only run K-means for one iteration on the above dataset (i.e., $K = 2$), what is the cluster assignment for each data point after Step 1 (5 points)? What are the updated cluster centers after Step 2 (5 points)?

2.2) ANS:

Let say K=2 and starting with cluster centers are $\mu_1 =- 1$ and $\mu_2 = 3$

After 1 iteration on above dataset,

- points x_1, x_2 and x_3 are in cluster 1 ($\mu_1 =- 1$)
- points x_4, x_5 and x_6 are in cluster 2 ($\mu_2 = 3$)

For 2 iteration,

we will following formula to update cluster center

$$\mu_k = \frac{\sum_n r_{n,k} X_n}{\sum_n r_{n,k}}$$

$\mu_1 = \frac{-3-1+0.5}{3} = \frac{-3.5}{3} = = -1.167$

$\mu_2 = \frac{2+3+4}{3} = \frac{9}{3} = 3$

---

3) [15 points] Suppose the initial cluster centers are $\mu_1 =- 3$, $\mu_2 = 0$, $\mu_3 = 3$ on the above dataset (i.e., $K = 3$). If we only run K-means for one iteration, what is the cluster assignment for each data point after Step 1 (5 points)? What are the updated cluster centers after Step 2 (5 points)? What is the value of the loss function J after this iteration (5 points)?

2.3) ANS:

After Step 1,

- cluster 1 ($\mu_1 =- 3$) will have points x_1
- cluster 2 ($\mu_2 = 0$) will have points $x_2,\ x_3$
- cluster 3 ($\mu_3 = 3$) will have points $x_4, x_5,\ and\ x_6$

After Step 2

- $\mu_1 = -3$
- $\mu_2 = \frac{-1+0.05}{2} = -0.25$
- $\mu_3 = \frac{2+3+4}{3} = 3$

Value of the loss function J after Step 2

$$J = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}||x_n - \mu_k||^2$$

$$J = (-3+3)^2 + (-1+0.25)^2 + (0.05+0.25)^2 + (-2+3)^2 + (3+3)^2 + (-3-4)^2$$

$J = 3.125$

---

4) [10 points] If we run K-means on the above dataset to find six clusters (i.e., $K = 6$). What is the **optimal** cluster assignment for each data point? (5 points)? What is the corresponding value of the loss function J for the optimal clustering assignment (5 points)?

2.4) ANS:

- Optimal cluster assignment for 6, each point will be in a separate cluster.
- loss function J will be 0; $J = 0$

---

3 Unconstrained Optimization [20 points]

We want to minimize a function

$$f(x) = x_1^2 + 5x_2^2$$

in a two-dimensional Euclidean space where $x = [x_1, x_2]^T$. Answer the following questions.

1) [5 points] What is the gradient at point $\bar{x} = [5, 1]^T$?

3.1) ANS:

$$\nabla f(x) = \begin{vmatrix} 2.x_1 \\ 10.x_2 \end{vmatrix}$$

$$\nabla f(x') = \begin{vmatrix} 10 \\ 10 \end{vmatrix}$$

---

2) [5 points] What is the steepest descent direction at point $\bar{x}$?

3.2) ANS:

$$-\nabla f(x') = \begin{vmatrix} -10 \\ -10 \end{vmatrix}$$

---

3) [5 points] Suppose we want to perform the backtrack line search at point $\bar{x}$ along direction $-\nabla f(\bar{x})$, using $\alpha = 0.4$ and $\beta = 0.9$. What is the computed step-size?

3.3) ANS:

to find computed step-size we need find largest value of t such that

$$f(x' - t\nabla f(x')) \leq f(x') - 0.4t\,||\nabla f(x')||^2$$

for $t = 1$, then $t \rightarrow t_\beta$

t in $\{1,\ 0.9,\ 0.9^2,\ 0.9^3,$ ………………..$\}$

∴ Computed step size = $0.9^{16} \approx 0.1850$

---

4) [5 points] Suppose we want to perform the exact line search at point $\bar{x}$ along direction $-\nabla f(\bar{x})$ What is the computed step-size?

3.4) ANS:

We need to find value using following formula:

$min\ f(x' - t\nabla f(x'))$

= $min\ 600t^2 - 200t + 30$

computed step size = $\frac{1}{6}$

---

4 The Decision Boundary for 1NN (i.e., 1-Nearest Neighbors Classifier) (20 points)

For each of the following 4 figures, we are given a few data points in the 2-D space, each of which is labeled as either '+' or '-'. Draw the decision boundary for 1NN, assuming we use L_2 (Euclidean) distance. **4) ANS: —> Drawn Blue lines.**

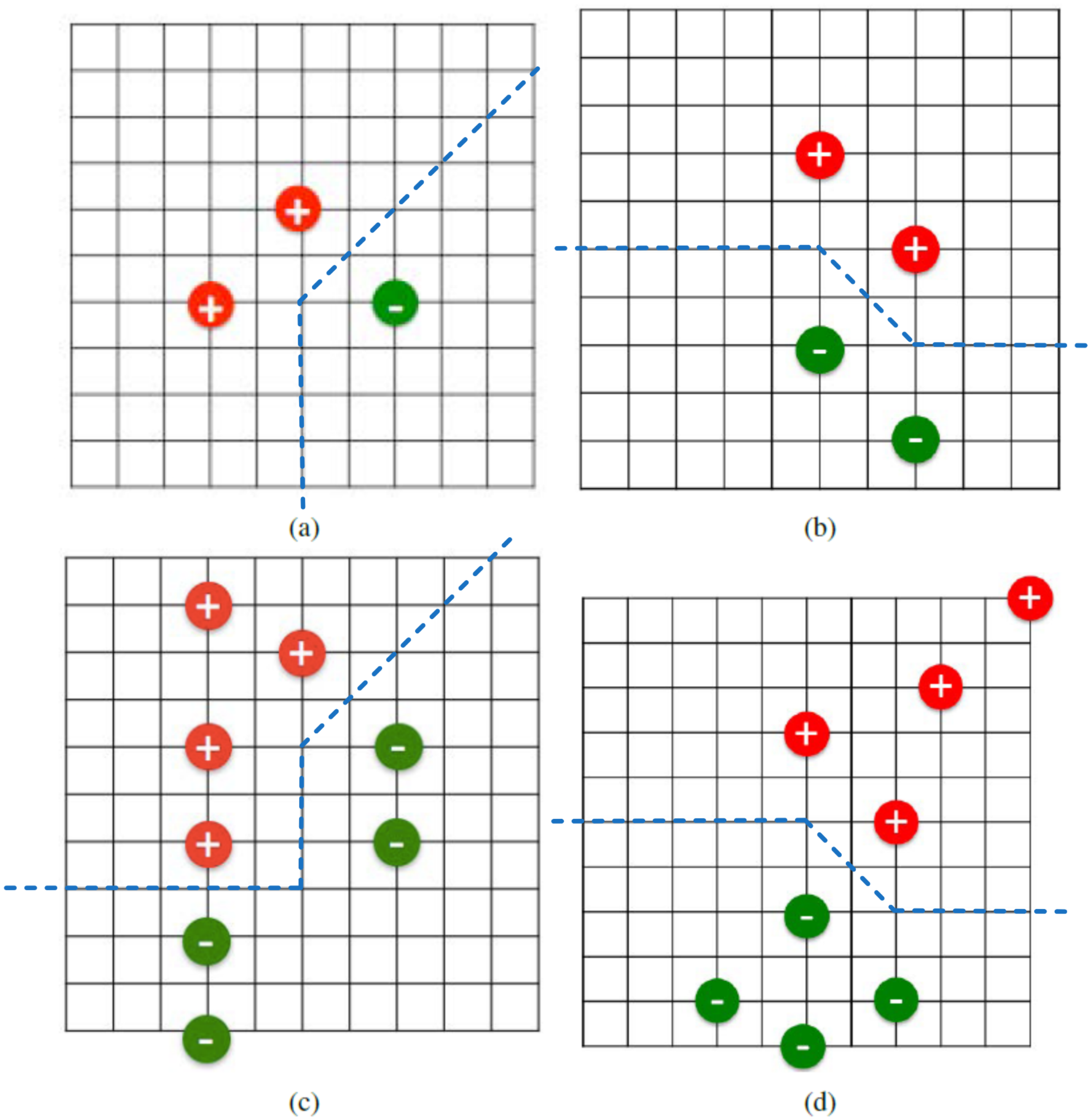


Figure 3: Problem 4: Training Data Set for 1NN Classifiers