

ECE 569 HW 2

Brandyn Tucknott

Due 28 October 2025

1. (Geometry of SVM) Consider the training samples $\{x_i, y_i\}$. Suppose that the training samples are from two classes. Let us denote the samples as $\mathcal{A} \subseteq \mathbb{R}^n$ (class label +1) and $\mathcal{B} \subseteq \mathbb{R}^n$ (class label -1) where $\mathcal{A} \cup \mathcal{B} = \{x_1, \dots, x_n\}$. SVM aims to find a hyperplane that can separate the convex hulls of $\mathcal{A} = \{a_1, \dots, a_m\}$ and $\mathcal{B} = \{b_1, \dots, b_k\}$ in the n -dimensional space.

- (a) [**Existence**] When does such a hyperplane exist; i.e., under what geometry of the training set does the SVM classifier exist? (use convex theory to support your answer)

Answer: Such a hyperplane exists when $\text{conv}\mathcal{A}$ and $\text{conv}\mathcal{B}$ are disjoint.

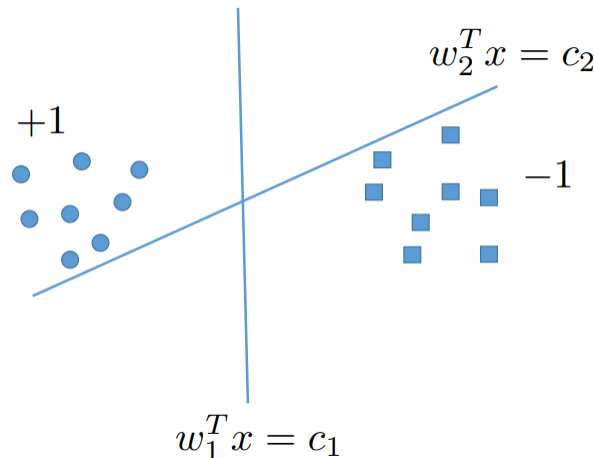
- (b) [**Generalization**] After a hyperplane is learned, say $w^T x = c$, which satisfies

$$w^T a_i \leq c \text{ for all } a_i \in \mathcal{A} \quad w^T b_k > c \text{ for all } b_k \in \mathcal{B}.$$

The prediction is as follows:

$$\hat{y}_{\text{new}} = \text{sign}(c - w^T x_{\text{new}}),$$

where x_{new} is unseen test data and \hat{y}_{new} is the predicted label. Consider the two hyperplanes in the figure. Which one do you prefer to have better performance in the prediction stage? Explain.



Answer: Call the hyperplanes $H_1 := w_1^T x = c_1$ and $H_2 := w_2^T x = c_2$. For this particular example, we prefer to use the classifier given by H_1 since if we use H_2 it is possible to have a point near the +1 cluster and far from the -1 cluster still get classified as -1. In a slightly more mathematical sense, H_1 has a larger margin, so it has a larger room for error and is therefore the better choice.

2. (Problem Formulation of SVM) The rationale behind finding a good separating hyperplane for \mathcal{A} and \mathcal{B} is as follows:

1. Find convex hulls of \mathcal{A} and \mathcal{B} .
2. Find points c^* and d^* from $\text{conv } \mathcal{A}$ and $\text{conv } \mathcal{B}$ respectively, and they are the closest to each other.
3. Construct a line segment between the two points.
4. Find a hyperplane orthogonal to the line segment and passing through its midpoint.

(a) Express this as a convex optimization problem.

Answer:

$$\begin{aligned} & \text{minimize } \|Au - Bv\|_2 \\ & \text{subject to } \mathbf{1}^T u = 1 \\ & \quad \mathbf{1}^T v = 1 \\ & \quad u \geq 0 \\ & \quad v \geq 0 \end{aligned}$$

All constraints of the above optimization problem are affine, and since all norms in \mathbb{R}^n are convex, we conclude that it is a convex optimization problem.

(b) Express the classifier hyperplane following steps 1-4.

Proof. We assume c^* and d^* are from $\text{conv } \mathcal{A}$ and $\text{conv } \mathcal{B}$ respectively and admit the smallest distance between the two convex hulls. To derive a hyperplane which is an orthogonal bisector to the line segment from c^* to d^* , we wish for the following to be true:

$$\text{for any } x \text{ on our hyperplane, } \|x - c^*\|_2 = \|x - d^*\|_2.$$

We can manipulate this equation to end up with a hyperplane.

$$\begin{aligned} \|x - c^*\|_2 &= \|x - d^*\|_2 \\ \|x - c^*\|_2^2 &= \|x - d^*\|_2^2 \\ x^T x - 2x^T c^* + c^{*T} c^* &= x^T x - 2x^T d^* + d^{*T} d^* \\ 2x^T (d^* - c^*) &= \|d^*\|_2^2 - \|c^*\|_2^2 \\ (d^* - c^*)^T x &= \frac{\|d^*\|_2^2 - \|c^*\|_2^2}{2}. \end{aligned}$$

This is exactly the equation of a hyperplane, so we are done. □

3. (Alternative Problem Formulation for SVM) Instead of first finding two points and constructing the separating hyperplane, we can formulate the problem directly.

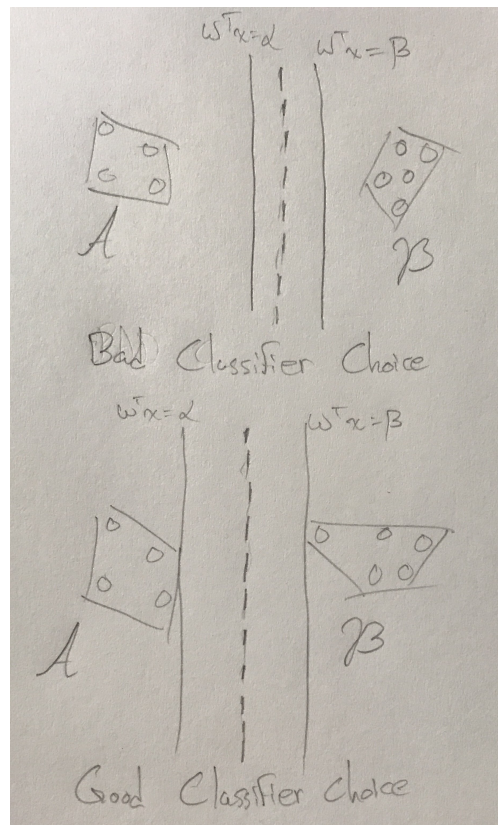
(a) What is the distance between the following two hyperplanes:

$$w^T x = \alpha \quad w^T x = \beta.$$

Answer: Notice these hyperplanes are parallel because they have the same normal vector, and in HW 1 we proved that the distance between them is

$$d = \frac{|\alpha - \beta|}{\|w\|_2}.$$

(b) 2-D plot of intuition.



- (c) Formulate $w^T x = \alpha$ and $w^T x = \beta$ as a convex optimization problem.

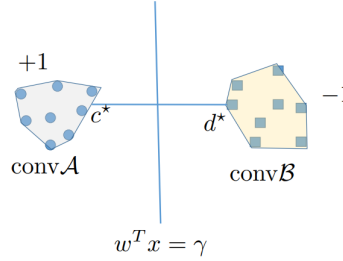
Answer: If we assume that $\alpha \neq \beta$, then we know that $\alpha < \beta$. We wish to maximize the margin between the supporting hyper planes, which we know to be $|\alpha - \beta| / \|w\|_2$. If we fix the magnitude of w to be 1, we fix the scaling and can Instead focus on maximizing the distance itself, which is now $\beta - \alpha$. This gives us the following convex problem:

$$\begin{aligned} & \text{maximize } \beta - \alpha \\ & \text{subject to } \|w\|_2 = 1 \\ & \quad w^T a_i \leq \alpha \\ & \quad w^T b_j \geq \beta \end{aligned}$$

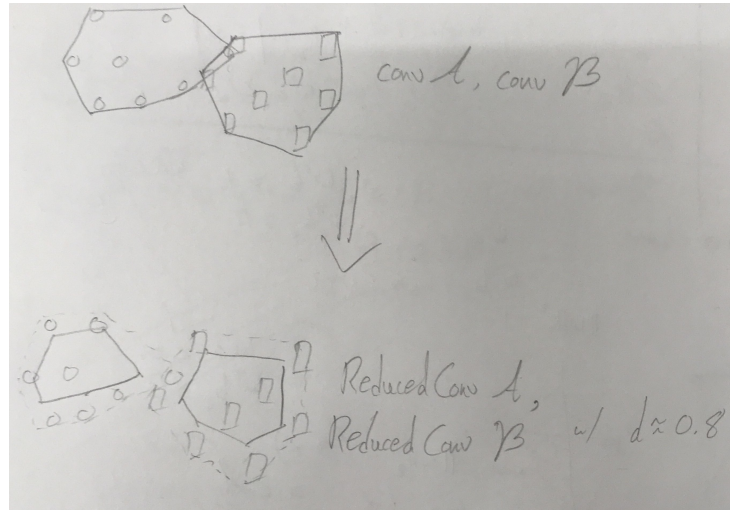
4. (Robustification of SVM) Consider when $\text{conv } \mathcal{A}$ and $\text{conv } \mathcal{B}$ overlap. Define a **Reduced Convex Hull** as follows:

$$\text{ReducedConv}\mathcal{A} = \{z \in \mathbb{R}^n : z = Au, \mathbf{1}^T u = 1, u \geq 0, u_i \leq d, \forall i = 1, \dots, m, d \in (0, 1)\}.$$

- (a) Given the following two training sets, draw the reduced convex hull sets with $d \approx 0.8$. Explain why using reduced convex hulls help in the overlapping case.



Answer:



The reduced convex hull is useful because with it we can define two convex subsets $\text{ReducedConv}\mathcal{A} \subset \text{conv}\mathcal{A}$ and $\text{ReducedConv}\mathcal{B} \subset \text{conv}\mathcal{B}$ such that $\text{ReducedConv}\mathcal{A}, \text{ReducedConv}\mathcal{B}$ are disjoint (just as in **Q2**), allowing us to use our previous formulations. Of course this won't yield a classifier which works for all datapoints, but it should be "good enough".

- (b) If the goal of the first step of the robustified SVM is to find two points from the reduced convex hulls of \mathcal{A} and \mathcal{B} (under prespecified d) such that the two points have the smallest distance, how do we formulate the problem? Write out your answer. Is your formulation convex?

Answer: Our answer is nearly identical to **Q2.a**, but we add the constraint $u_i, v_i \leq d$ to account for the reduced convex hull.

$$\begin{aligned} & \text{minimize } \|Au - Bv\|_2 \\ & \text{subject to } \mathbf{1}^T u = 1 \\ & \quad \mathbf{1}^T v = 1 \\ & \quad 0 \leq u \leq d \\ & \quad 0 \leq v \leq d \end{aligned}$$