ECE595 / STAT598: Machine Learning I
Lecture 6.3: Linear Separability

Spring 2020

Stanley Chan

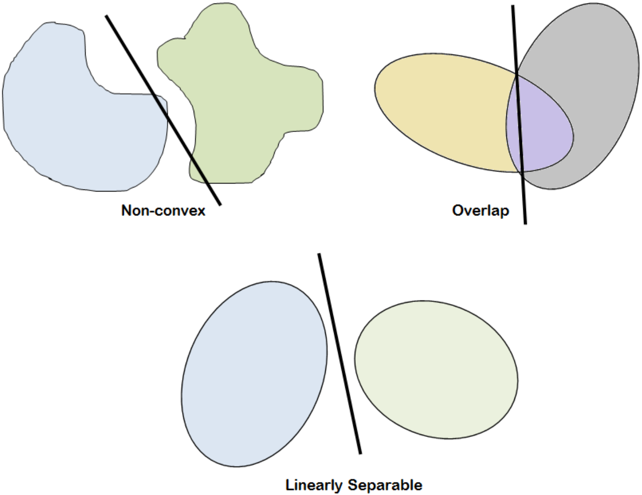
School of Electrical and Computer Engineering
Purdue University

Outline

Goal: Understand the geometry of linear separability.

- Notations
  - Input Space, Output Space, Hypothesis
  - Discriminant Function
- Geometry of Discriminant Function
  - Separating Hyperplane
  - Normal Vector
  - Distance from Point to Plane
- Linear Separability
  - Which set is linearly separable?
  - Separating Hyperplane Theorem
  - What if theorem fails?

Which one is Linearly Separable? Which one is Not?



Separating Hyperplane Theorem

Can we always find a separating hyperplane?

- No.
- Unless the classes are linearly separable.
- If convex and not overlapping, then yes.

Theorem (Separating Hyperplane Theorem)

Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two closed convex sets such that $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$. Then, there exists a linear function

$$g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0,$$

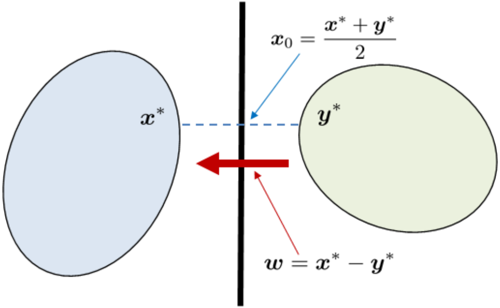
such that $g(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{C}_1$ and $g(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \in \mathcal{C}_2$.

Remark: The theorem above provides sufficiency but not necessity for linearly separability.

Separating Hyperplane Theorem

Pictorial "proof":

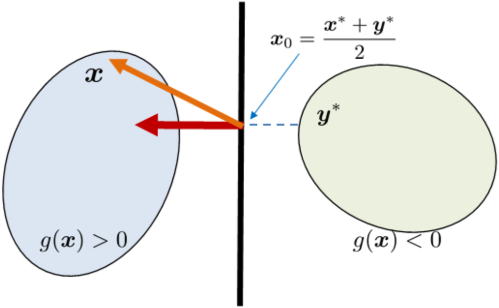
- Pick two points $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ s.t. the distance between the sets is minimized.
- Define the mid-point as $\boldsymbol{x}_0 = (\boldsymbol{x}^* + \boldsymbol{y}^*)/2$.
- Draw the separating hyperplane with normal $\boldsymbol{w} = \boldsymbol{x}^* - \boldsymbol{y}^*$
- Convexity implies any inner product must be positive.



Separating Hyperplane Theorem

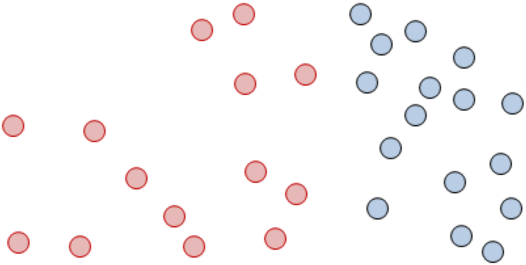
Pictorial "proof":

- Pick two points $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ s.t. the distance between the sets is minimized.
- Define the mid-point as $\boldsymbol{x}_0 = (\boldsymbol{x}^* + \boldsymbol{y}^*)/2$.
- Draw the separating hyperplane with normal $\boldsymbol{w} = \boldsymbol{x}^* - \boldsymbol{y}^*$
- Convexity implies any inner product must be positive.



Linearly Separable?

- I have data $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$.
- Closed. Convex. Non-overlapping.
- Separating hyperplane theorem: I can find a line.
- Victory?
- Not quite.



When Theory Fails

Theorem (Separating Hyperplane Theorem)

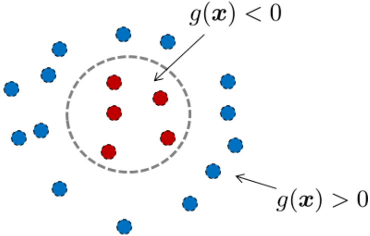
Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two closed convex sets such that $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$. Then, there exists a linear function

$$g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0,$$

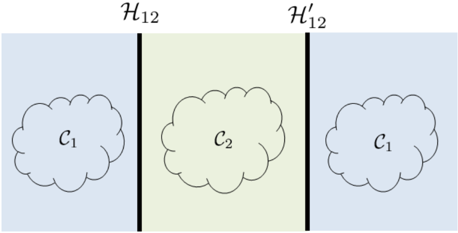
such that $g(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{C}_1$ and $g(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \in \mathcal{C}_2$.

- Finding a separating hyperplane for **training set** does not imply it will work for the **testing set**.
- Separating hyperplane theorem is more often used in **theoretical analysis** by assuming properties of the testing set.
- If a dataset is linearly separable, then you are guaranteed to find a perfect classifier. Then you can say how good is the classifier you designed compared to the perfect one.

Linear Classifiers Do Not Work



Example 1



Example 2

- **Intrinsic geometry** of the two classes could be bad.
- The training set could be **lack of training samples**.
- Solution 1: Use non-linear classifiers, e.g., $g(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{W} \boldsymbol{x} + \boldsymbol{w}^T \boldsymbol{x} + \omega_0$.
- Solution 2: Kernel method, e.g., Radial basis function.
- Solution 3: Extract features, e.g., $g(\boldsymbol{x}) = \boldsymbol{w}^T \phi(\boldsymbol{x})$.

Reading List

Separating Hyperplane:

- Duda, Hart and Stork's *Pattern Classification*, Chapter 5.1 and 5.2.
- Princeton ORFE-523, Lecture 5 on Separating hyperplane http://www.princeton.edu/~amirali/Public/Teaching/ORF523/S16/ORF523_S16_Lec5_gh.pdf
- Cornell ORIE-6300, Lecture 6 on Separating hyperplane https://people.orie.cornell.edu/dpw/orie6300/fall2008/Lectures/lec06.pdf
- Caltech, Lecture Note http://www.its.caltech.edu/~kcborder/Notes/SeparatingHyperplane.pdf