Support vector machines and Radon's theorem

Brittany Carr, with Henry Adams and Elly Farnell

Department of Mathematics, Colorado State University

Goal

Find and classify all possible support vector configurations for a set of points in "general position".

Support Vector Machines

Support vector machines (SVM) are an algorithm which, when given a set of linearly separable points in $\mathbb{R}^n$, will find the separating hyperplane with the widest margin of separation between the two classes. Specifically, this work focuses on hard margin SVM where

- No points inside the margin
- No points misclassified

Classifier

The classifier is given by $f(x) = w^T x + b$ where

- x: The data points
- b: Shift of the hyperplane away from the origin
- w: The normal vector defining the hyperplane
- $y_x \in \{-1, 1\}$: Labels for our data

Support vectors are the points which determine the edges of the separating hyperplane and have margin exactly 1.

Example

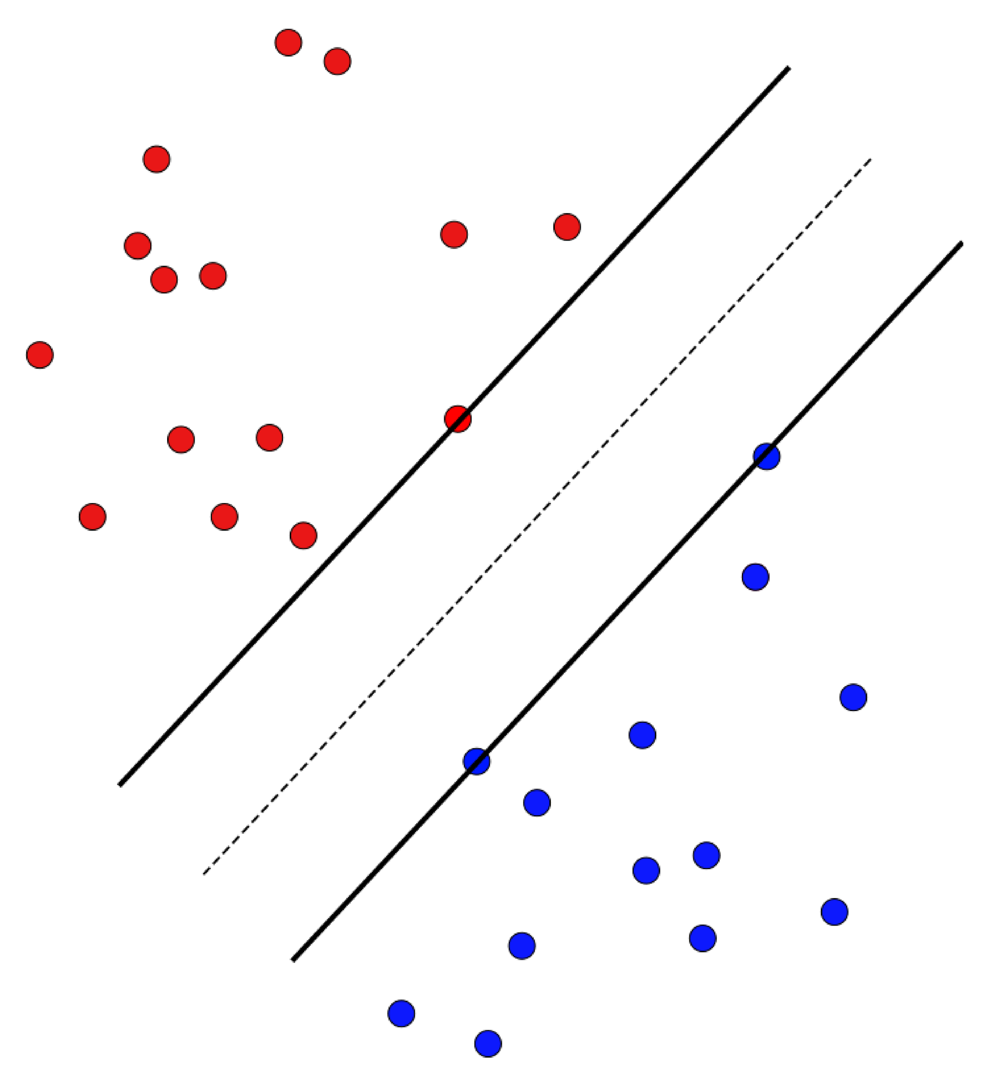


Figure: An example of a SVM in $\mathbb{R}^2$

Optimization Problem

SVM is an optimization problem. Given a set of linearly separable data, the optimization problem is,

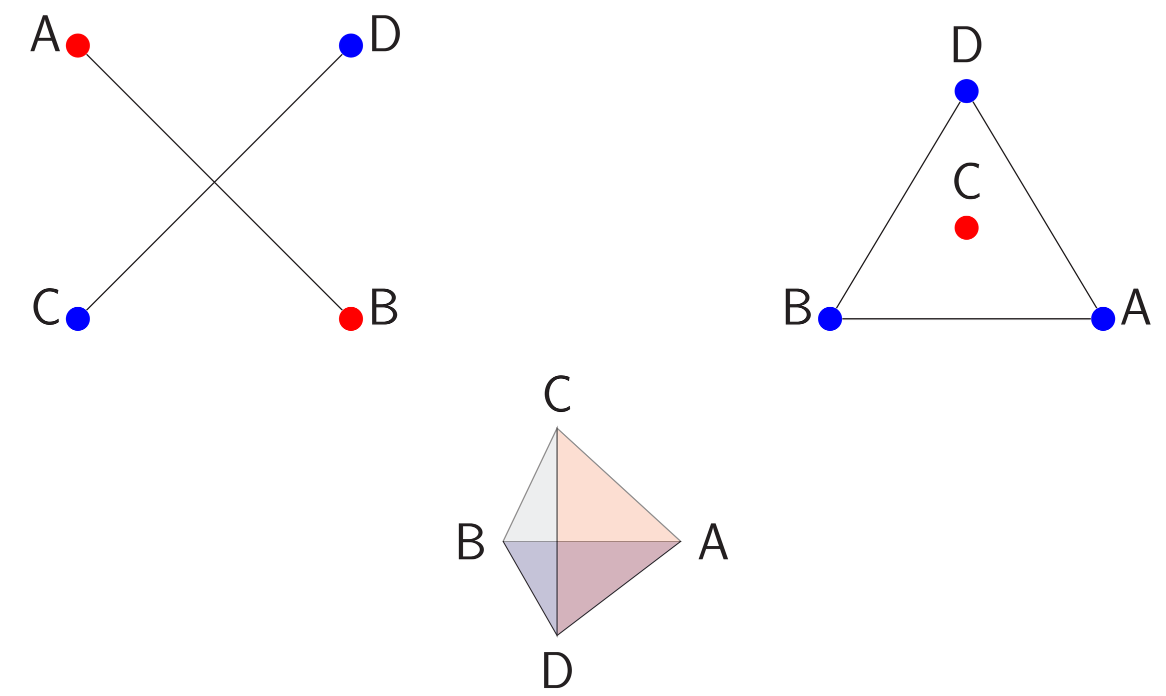
$$\arg\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \text{ subject to } y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1 \text{ for all } i.$$

Using the Karush Kuhn Tucker conditions, this becomes a problem which can be solved in different ways.

Radon's Theorem

If T is a set of k points in Euclidean n-dimensional space $\mathbb{R}^n$ with $k \geq n+2$, then there exist disjoint sets T_1 and T_2 with $T = T_1 \cup T_2$ and $\operatorname{conv}(T_1) \cap \operatorname{conv}(T_2) \neq \varnothing$.

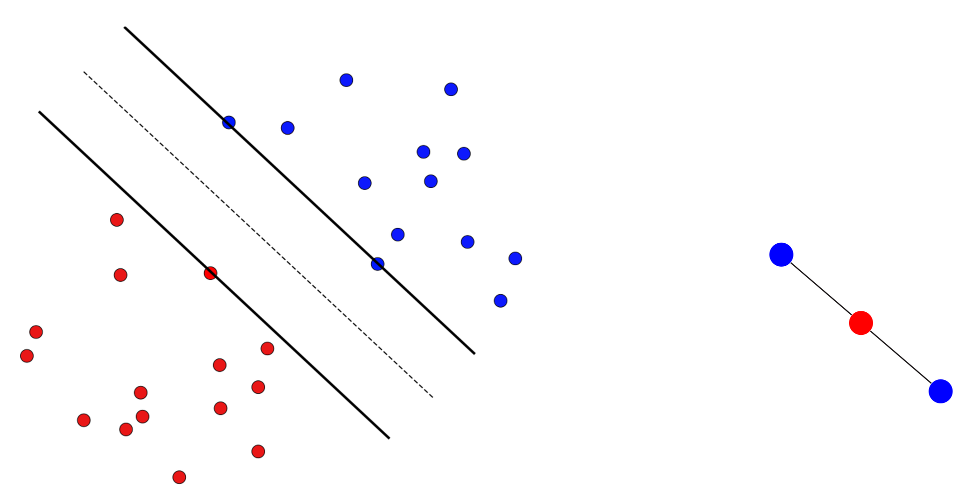
Example



Radon's Theorem and SVM Configurations

If $X \subset \mathbb{R}^n$ is a set of linearly separable labeled points, then the projections of the convex hulls of the positive and negative support vectors onto the separating hyperplane intersect.

Example



For points in "general position", we want to show there is only one Radon point.

Strong General Position

A set of points $X \subseteq \mathbb{R}^n$ is in strong general position if

(i) for $k < n$, no $k+2$ subset of X lies in a k-flat

(ii) for any $k+1$ points in X (determining a k-flat), the orthogonal projection of any other point in X to that k-flat does not hit the affine span of k of those points

(iii) for $k + l \leq n$, no disjoint k-flats and l-flats contain parallel vectors.

Radon Point Existence

If $X \subset \mathbb{R}^n$ is a set of linearly separable labeled points in strong general position, then the projections of the convex hulls of the positive and negative support vectors onto the separating hyperplane intersect at a single Radon point.

Example

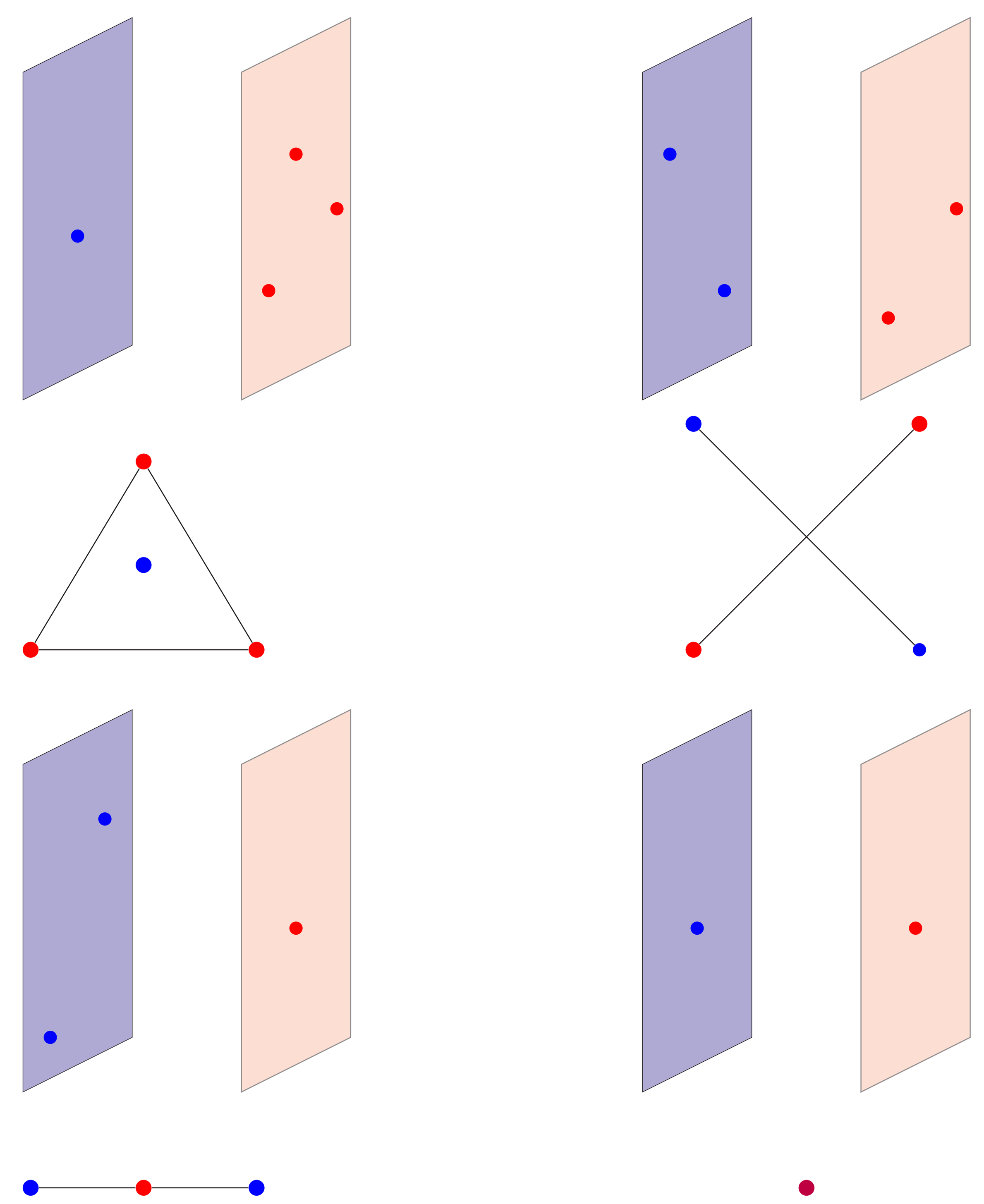


Figure: All possible support vector configurations in $\mathbb{R}^3$ and their projections onto the separating hyperplane.

Results

- Suppose $X \subseteq \mathbb{R}^n$ is in strong general position, and that X is equipped with linearly-separable labels. Then there exists at most $n+1$ supporting vectors.
- If $X \subseteq \mathbb{R}^n$ is a set of linearly separable labeled points with positive margin, then there exists an $\varepsilon > 0$ such that upon perturbing any point by at most ε, X remains linearly separable.

Conjecture

Let $X \subseteq \mathbb{R}^n$ be a set of linearly separable labeled points in strong general position. Let $\varepsilon_0 > 0$ be the minimum distance between any two distinct points in X. Then there exists an $\varepsilon > 0$ with $\varepsilon < \frac{\varepsilon_0}{2}$ such that if each point is perturbed by at most ε, then the set of supporting vectors remains unchanged.

Open Questions

- What configurations of support vectors would happen for soft margin, spherical, or ellipsoidal SVM?
- The kernel method is a type of SVM for linearly inseparable data. What could be said about the support vectors?
- What is the probability of obtaining one support vector configuration over another?