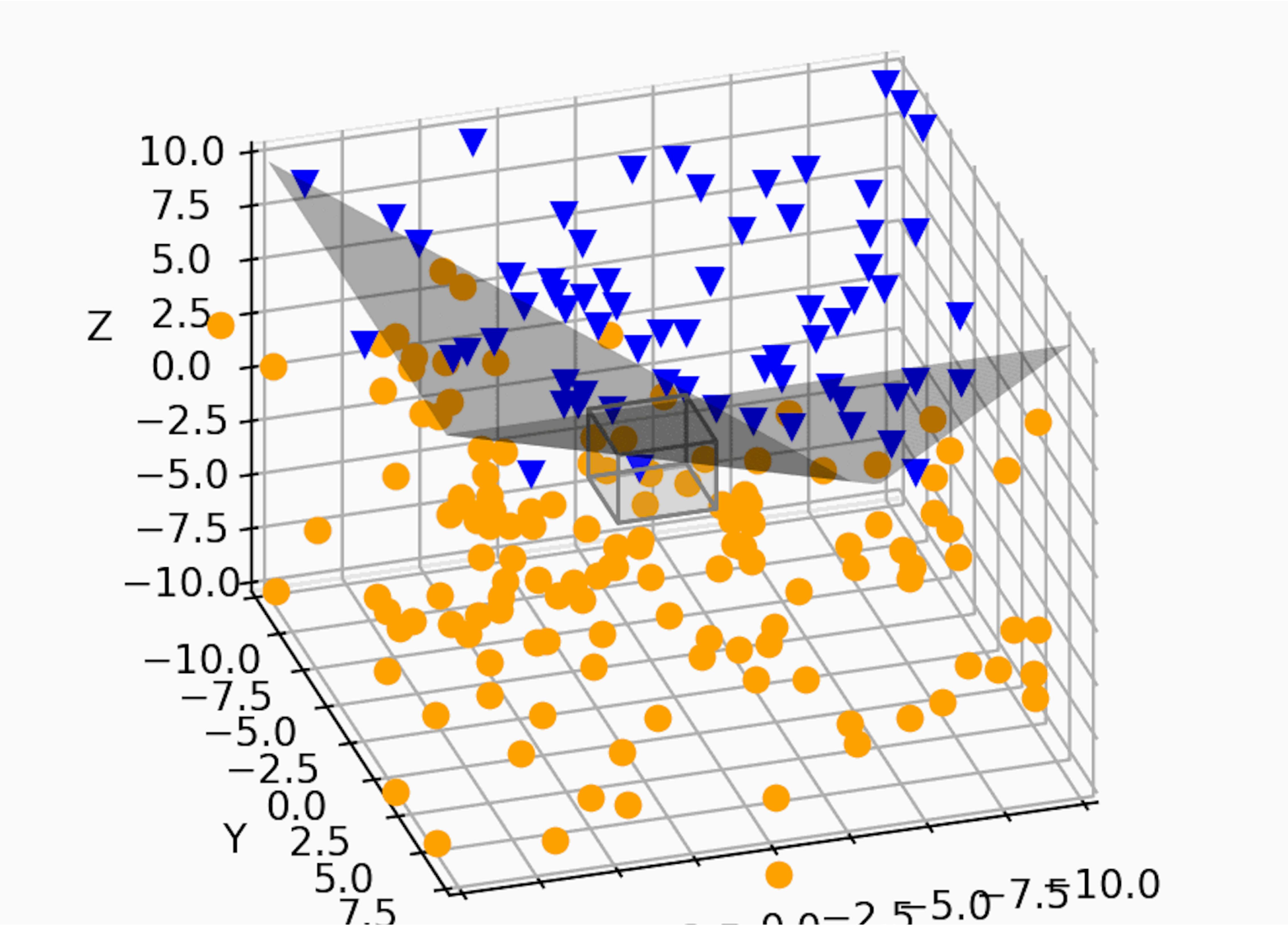
Tropical Support Vector Machines

MAP513 — Tuesday, December 19th

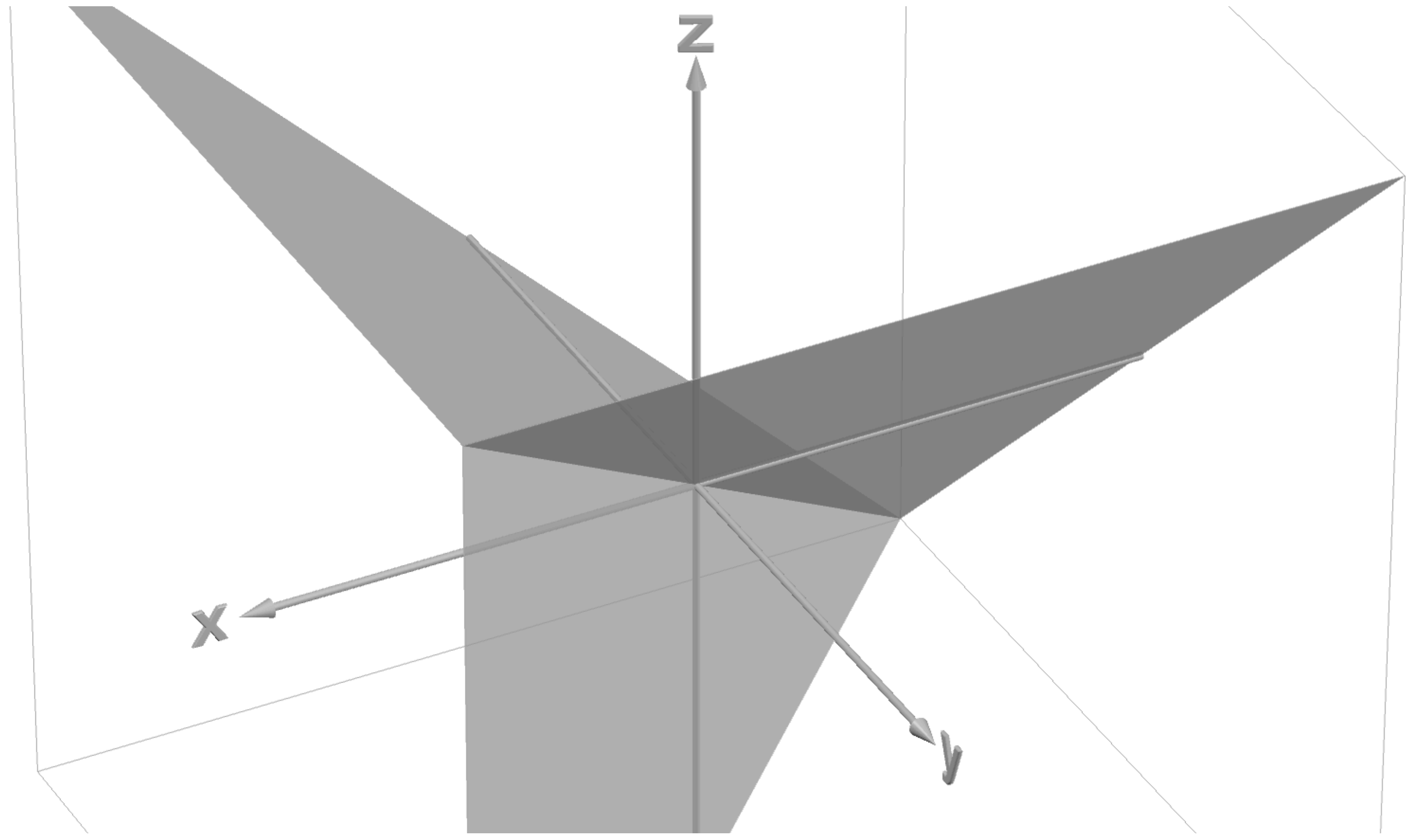


Samuel Boïté, Théo Molfessis
Supervised by Xavier Allamigeon, Stéphane Gaubert

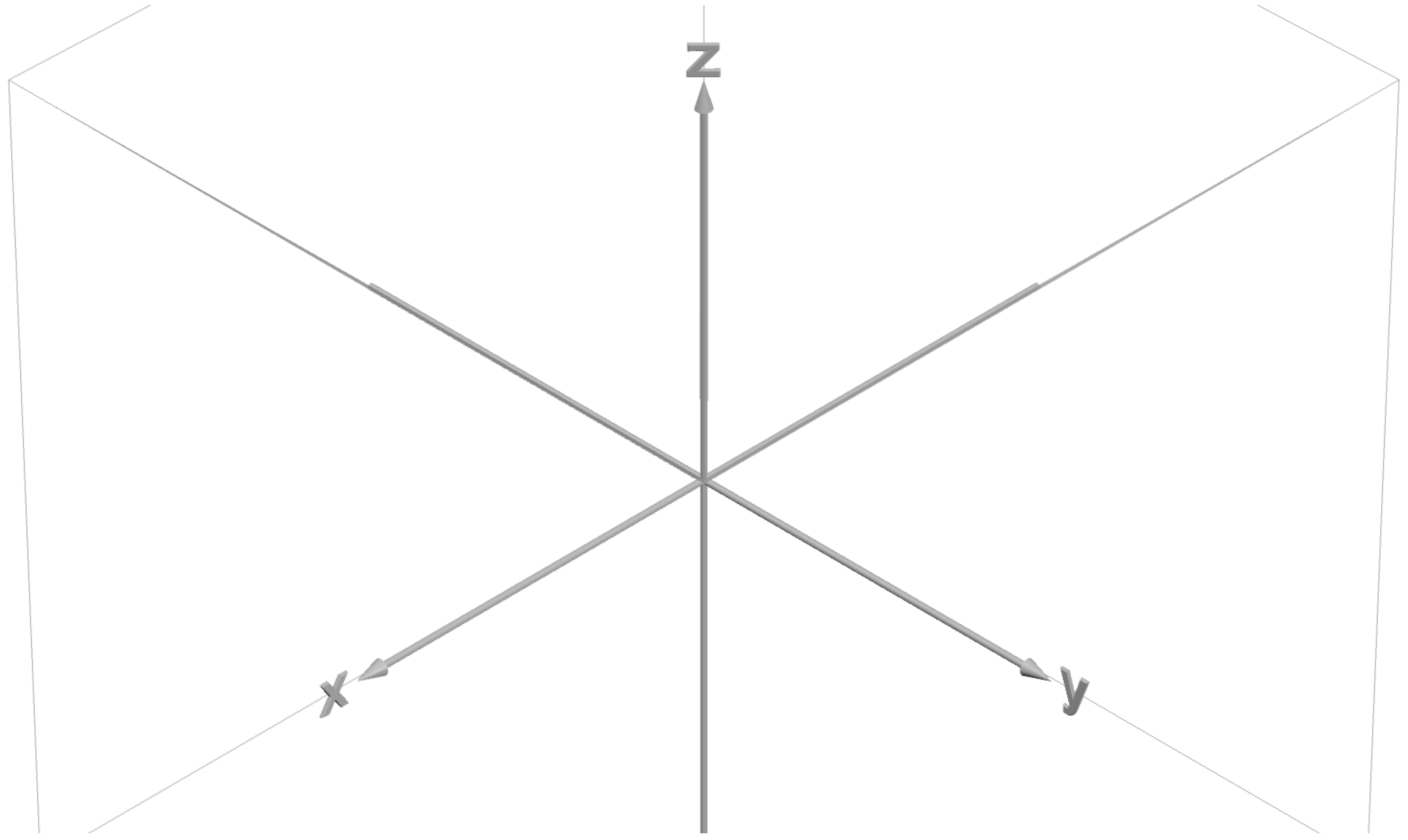
Tropical hyperplanes

- Max-plus semi-field: $\mathbb{R}_{\max} = \mathbb{R} \cup \{-\infty\}$ with addition $a \oplus b = \max(a, b)$ and multiplication $a \odot b = a + b$.

- Tropical hyperplane of apex $a \in \mathbb{R}^d_{\max}$: splits space depending on where $x - a$ reaches its maximum coordinate.



In 3D space

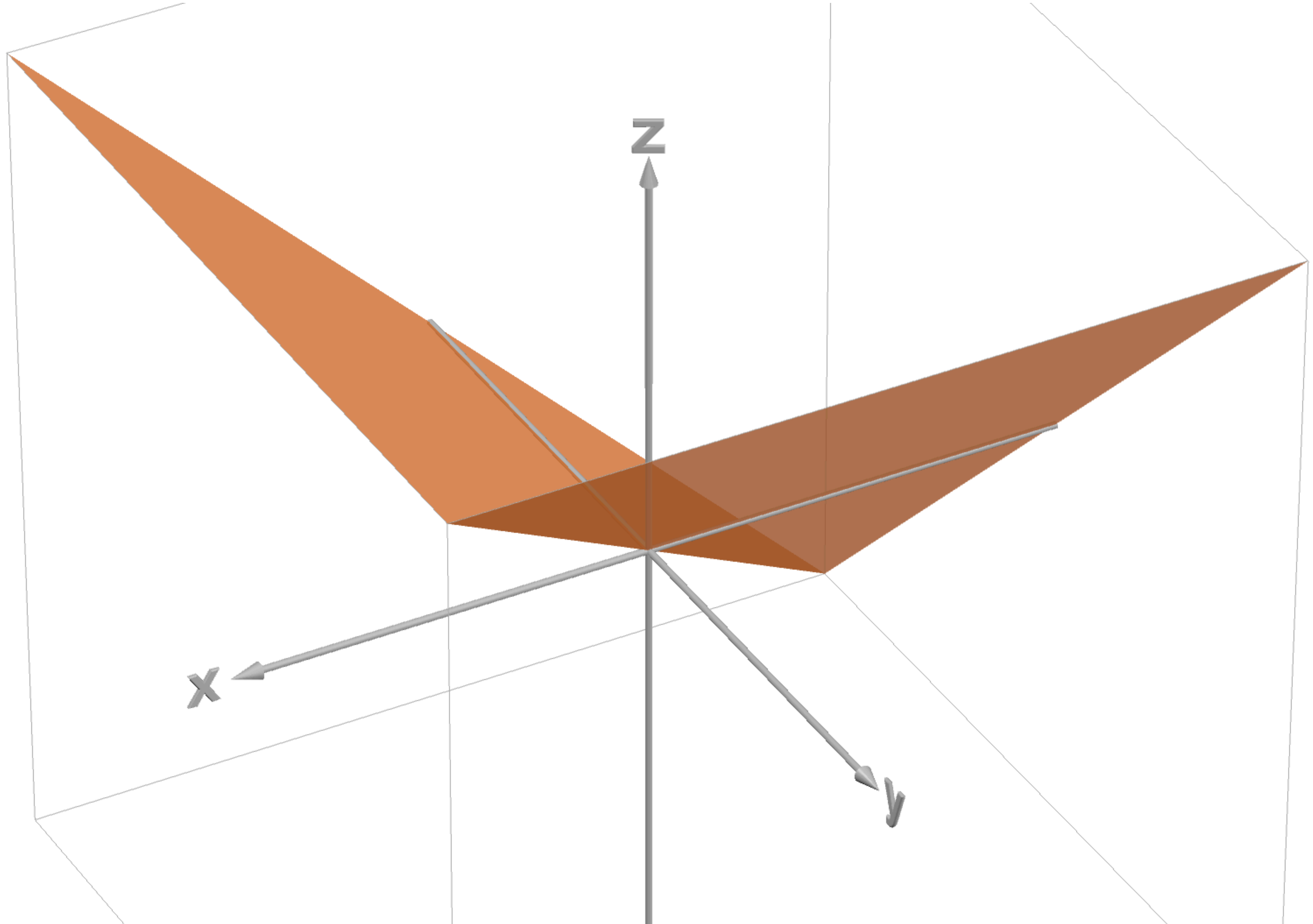


In the projective plane

- $H_a := \left\{ x \in \mathbb{R}^d_{\max}, \quad (x - a) \text{ reaches its max coordinate at least twice} \right\}$.

Tropical hyperplanes

- *Tropical parametrised hyperplane* of config. $\sigma = \{I^{\pm}\}$:
  $H_a^\sigma := \left\{x \in \mathbb{R}_{\max}^d, \quad (x-a) \text{ reaches its max coordinate in } I^+ \text{ and } I^-\right\}$



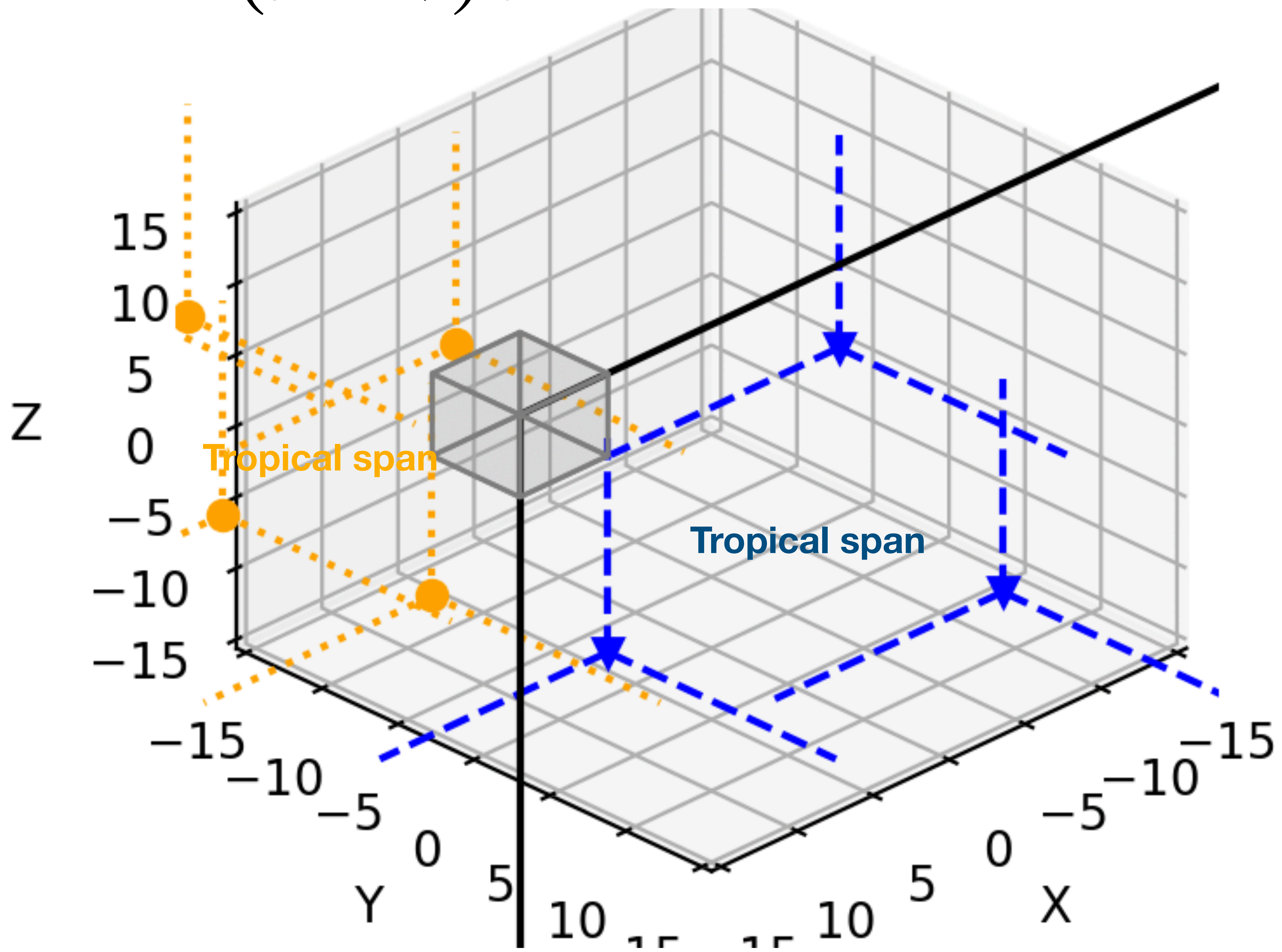
Tropical classification problem

- n classes of d-dimensional points: $X_1, \ldots, X_n \subset \mathbb{R}^d$.
- Tropical distance: $d(u, v) := \max(u - v) - \min(u - v)$.
- Hard separation with margin ν:

$$\forall x^k \in X^k, \quad \arg\max_{i \in [d]} x_i^k \in I^k,$$

$$d(H_a^\sigma, x^k) = \max(x^k - a) - \max_{\bigsqcup_{\ell \neq k} I^\ell} (x^k - a) \geq \nu \,.$$

- Amounts to separate tropical spans.



Tropical Binary Classification Example

Describing tropical spans

- **Shapley operator**: non-decreasing map over $\mathbb{R}^d_{\max}$ verifying $\forall \lambda \in \mathbb{R}_{\max}, T(\lambda + x) = \lambda + T(x)$.
- For T a Shapley map, we define $\mathcal{S}(T) = \{x \in \mathbb{R}^d, \quad x \leq T(x)\}$.
- Important example: *tropical projections:*

$$\forall i \in [p], \quad P_X(x)_i = \max_{j \in [p]} \left\{ X_{ij} + \min_{k \in [d]} (-X_{kj} + x_k) \right\} .$$

$$\boxed{\mathcal{S}(P_X) = \mathrm{Span}(X) \, .}$$

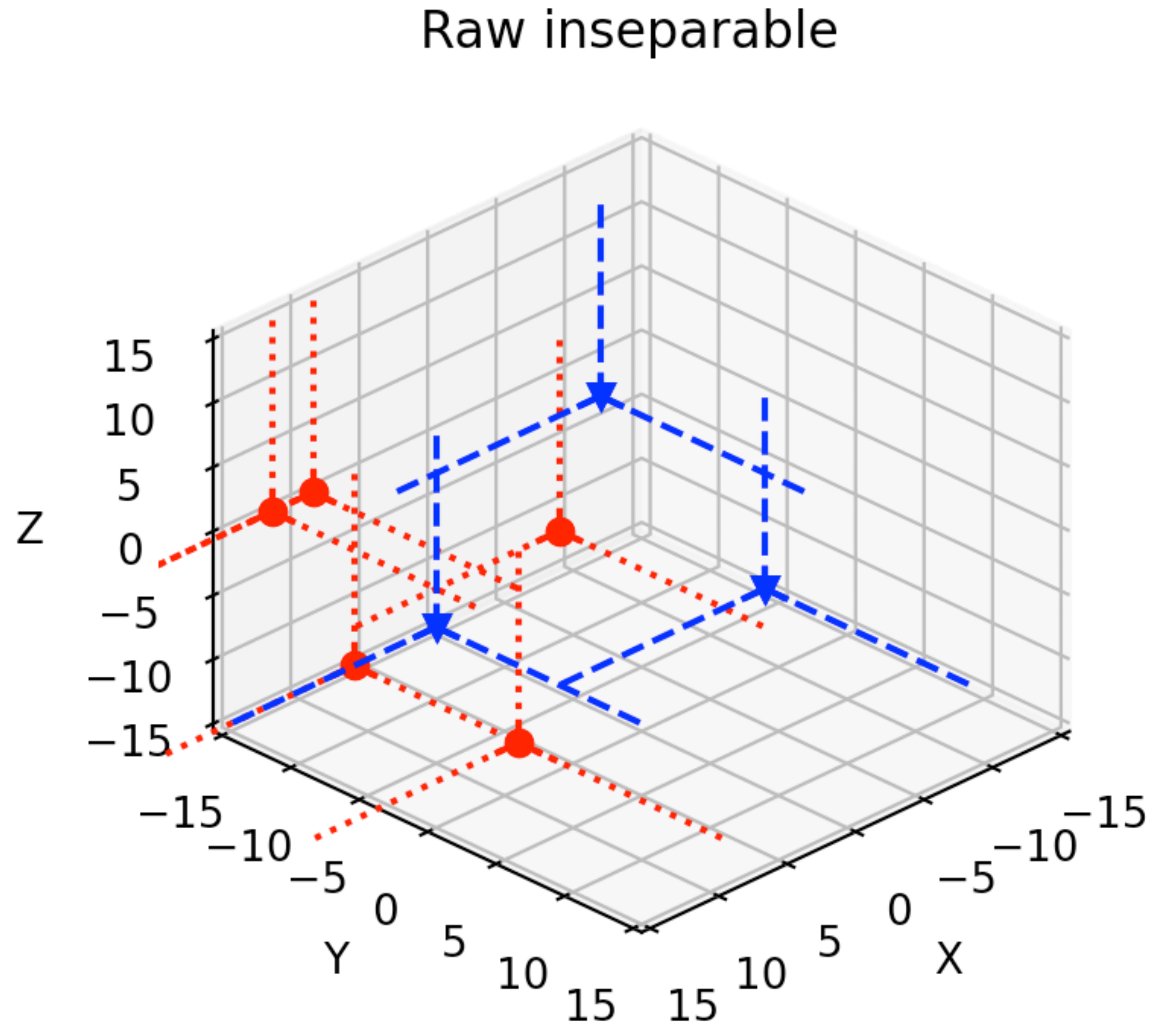
- Hence *tropical projections* describe the sets we want to separate.

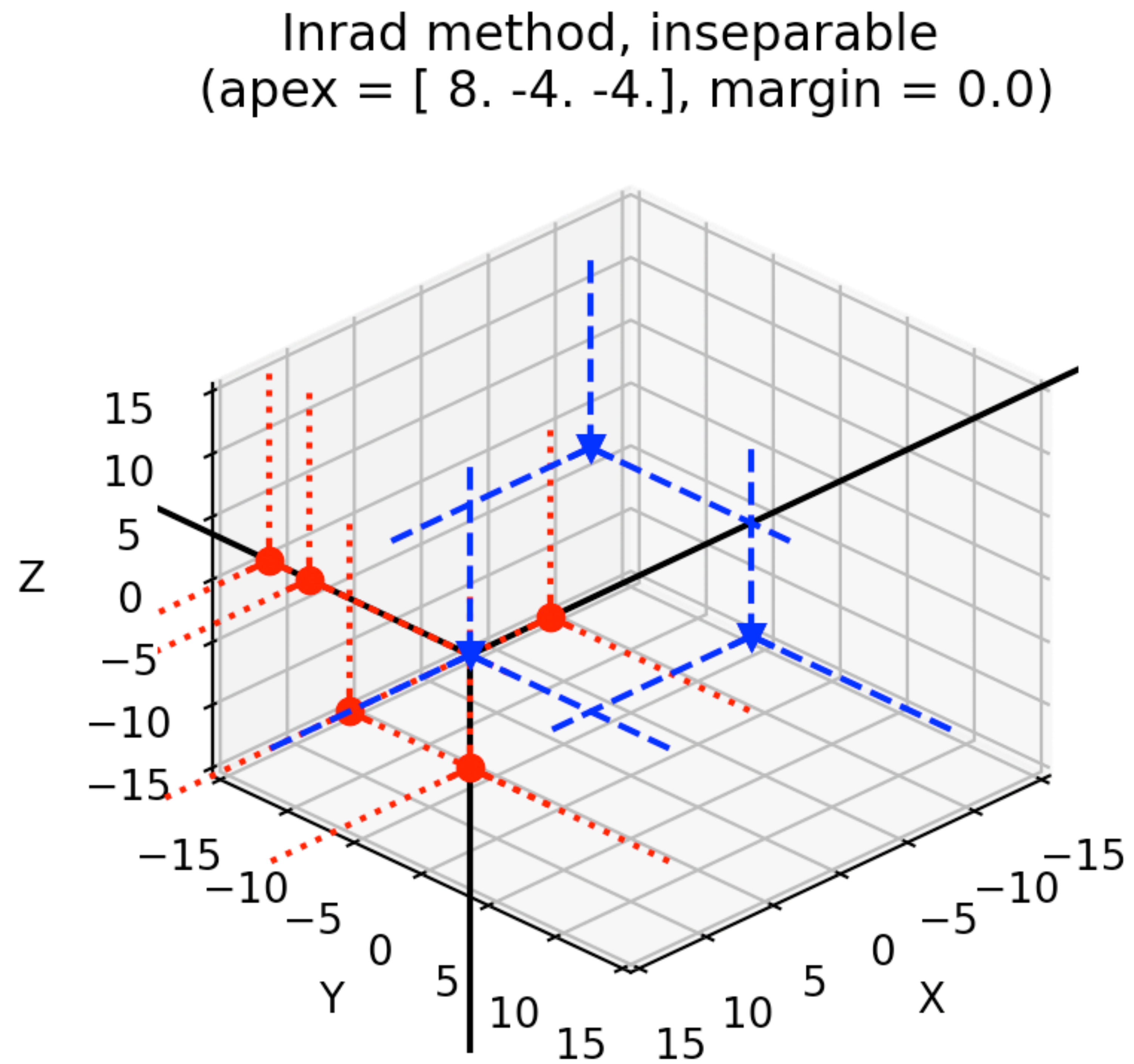
Dealing with non-separability

- **Objective:** Evaluating overlap $V^+ \cap V^-$ in terms of distance to separability.
- $\mathcal{S}(T^+ \wedge T^-) = \mathcal{S}(T^+) \cap \mathcal{S}(T^-)$: we have a Shapley for the intersection.
- We can make it *diagonal-free*.
- *[Allamigeon, Gaubert et al.]* $\rho(T) = \text{inrad}\, \mathcal{S}(T)$ when T DF.
- Eigenpairs computable in pseudo-polynomial time with Krasnoselskii-Mann.

Dealing with non-separability

- **Proposition.** Projecting points at distance less than $\rho(T)$ over H_a nullifies the interior of $V^+ \cap V^-$, with a eigenvector associated with $\rho(T)$.





Optimal margin in the separable case

- Let (a, λ) the unique eigenpair: in the separable case, $\lambda < 0$.
- Let the sectors $I^{\pm} := \{i \in [d], \quad T^{\pm}(a)_i > \lambda + a_i\}$.
- **Proposition.** H_a^{σ}, given the sectors defined above, separates $V^{\pm}$ with a margin of $-\lambda$. It is optimal in the case of finite point clouds.

Optimal margin in the separable case

- Let (a, λ) the unique eigenpair: in the separable case, $\lambda < 0$.
- Let the sectors $I^{\pm} := \{i \in [d], \quad T^{\pm}(a)_i > \lambda + a_i\}$.
- **Proposition.** H_a^σ, given the sectors defined above, separates $V^{\pm}$ with a margin of $-\lambda$. It is optimal in the case of finite point clouds.
- *Proof.* As $T^{\pm}$ are non-expansive, we have for $x^{\pm} \in V^{\pm}$:

$$x_i^{\pm} \le \max(x^{\pm} - a) + T^{\pm}(a)_i \,.$$

- For instance, let $i \in I^-$. Then $T^+(a)_i = \lambda + a_i$, so for $x^+ \in V^+$:

$$x_i^+ - a_i \le \max(x^+ - a) + \lambda \,.$$

- And:

$$d(H_a^\sigma, x^+) = \max(x^+ - a) - \max(x^+ - a)_{I^-} \geq -\lambda\,.$$

- For the optimality, when $T^\pm$ is in the form:

$$T^\pm(x) := \sup_{v \in V^\pm} \big(v_i + \min(-v + x)\big)\,.$$

- For $\varepsilon > 0$, we can find $v \in V^+$ such that:

$$T^+(a)_i - \varepsilon \leq v_i - \max(v - a) \leq T^+(a)_i\,.$$

- Hence:

$$\lambda - \varepsilon \leq v_i - a_i - \max(v - a) \leq \lambda\,.$$

Topics being explored

1. Hard-Margin Multi-Classification

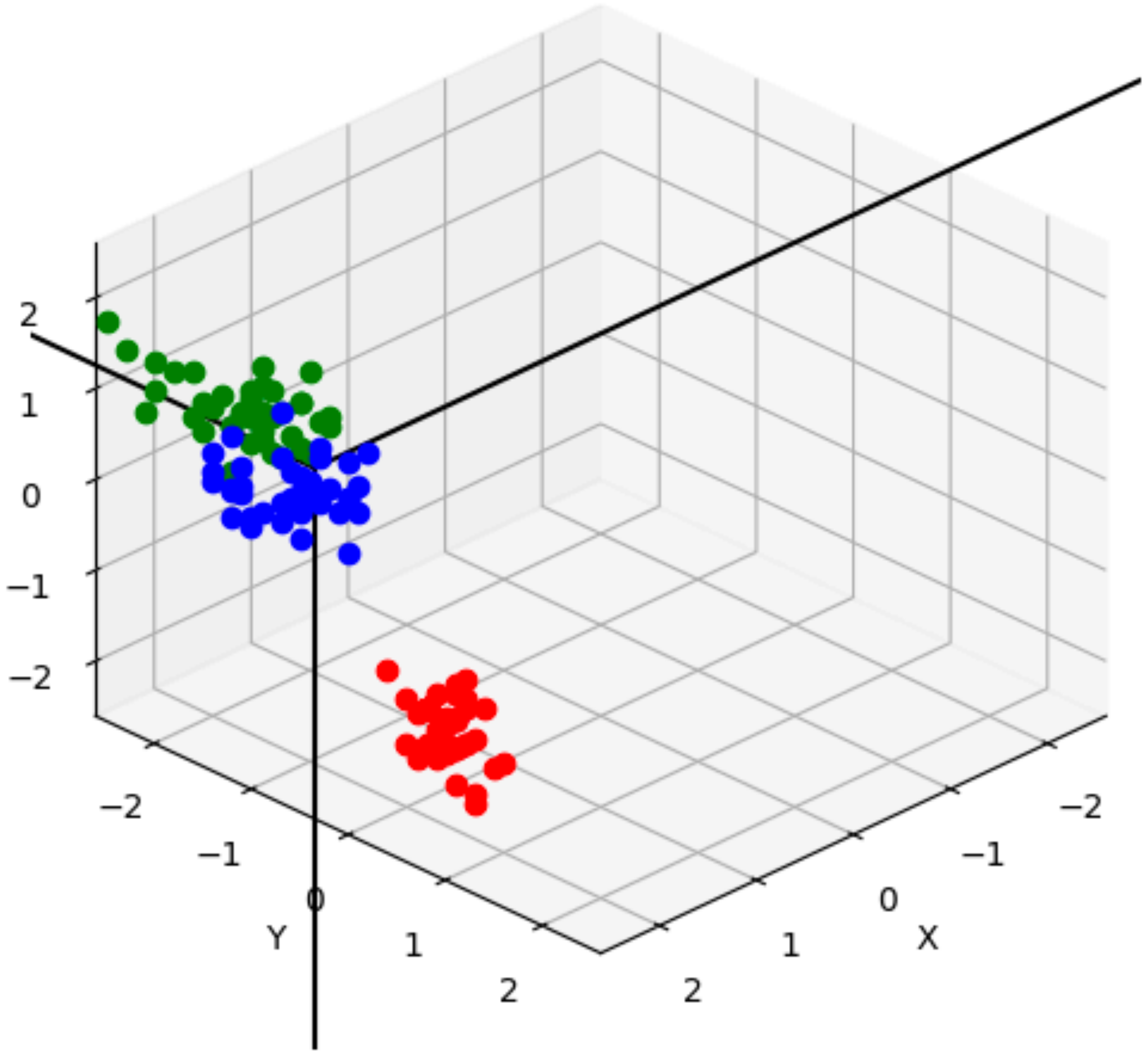
All-Vs-All

Same proofs!

$$T := \bigvee_{1 \leq k < l \leq n} T^k \wedge T^l$$

One-Vs-All?

$$T^{(1)} := T^1 \wedge \left(\bigvee_{j \neq 1} T^j \right).$$



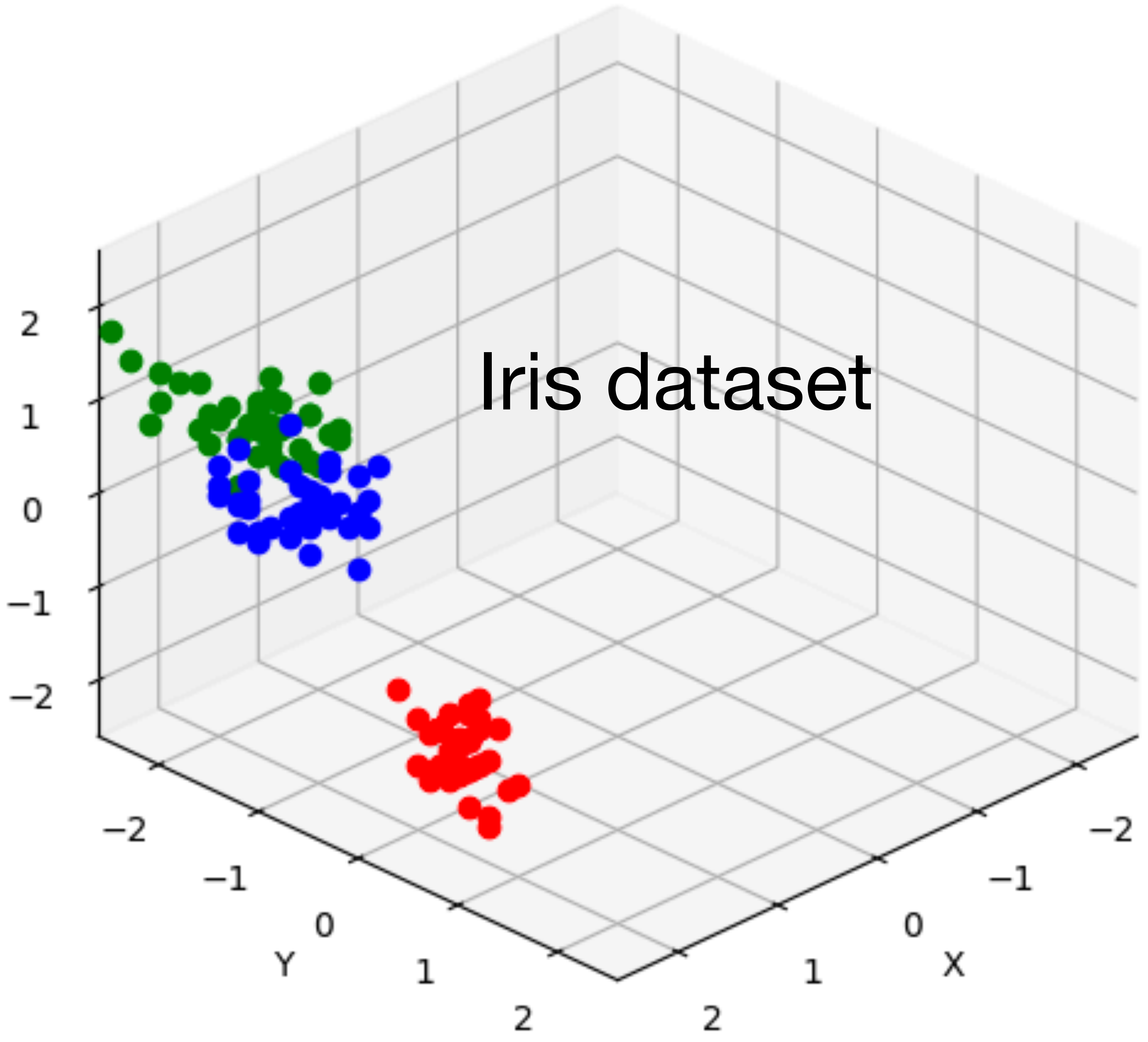
Topics being explored

2. Adding Features: Tropically Polynomial Decision Boundaries

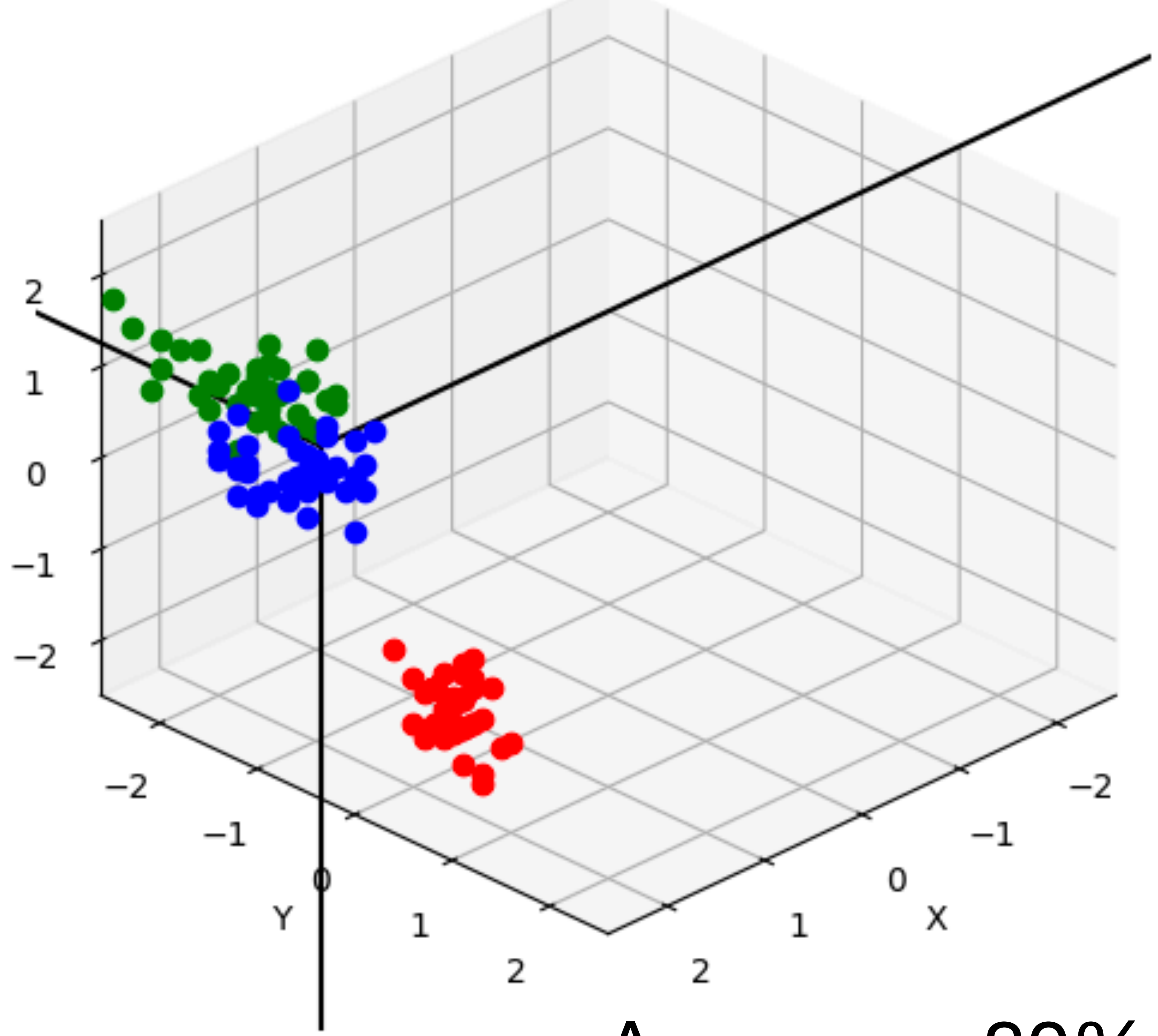
- Initial space: $\mathbb{R}^d$.
- New space: ver$(x) := \left(\langle x, \alpha \rangle\right)_{\alpha \in \mathscr{A}} \in \mathbb{R}^{\mathscr{A}}$.
- We consider integer combinations of features: $\mathscr{A}_s := (s\Delta_d) \cap \mathbb{Z}^d$.
- Ex: $\mathscr{A}_1$ corresponds to initial space, $\mathscr{A}_2$ also contains all sums of 2 features…
- *[Zhang, 18]* Feedforward neural networks + ReLU ≡ Tropical Rational Functions

Topics being explored

2. Adding Features: Tropically Polynomial Decision Boundaries



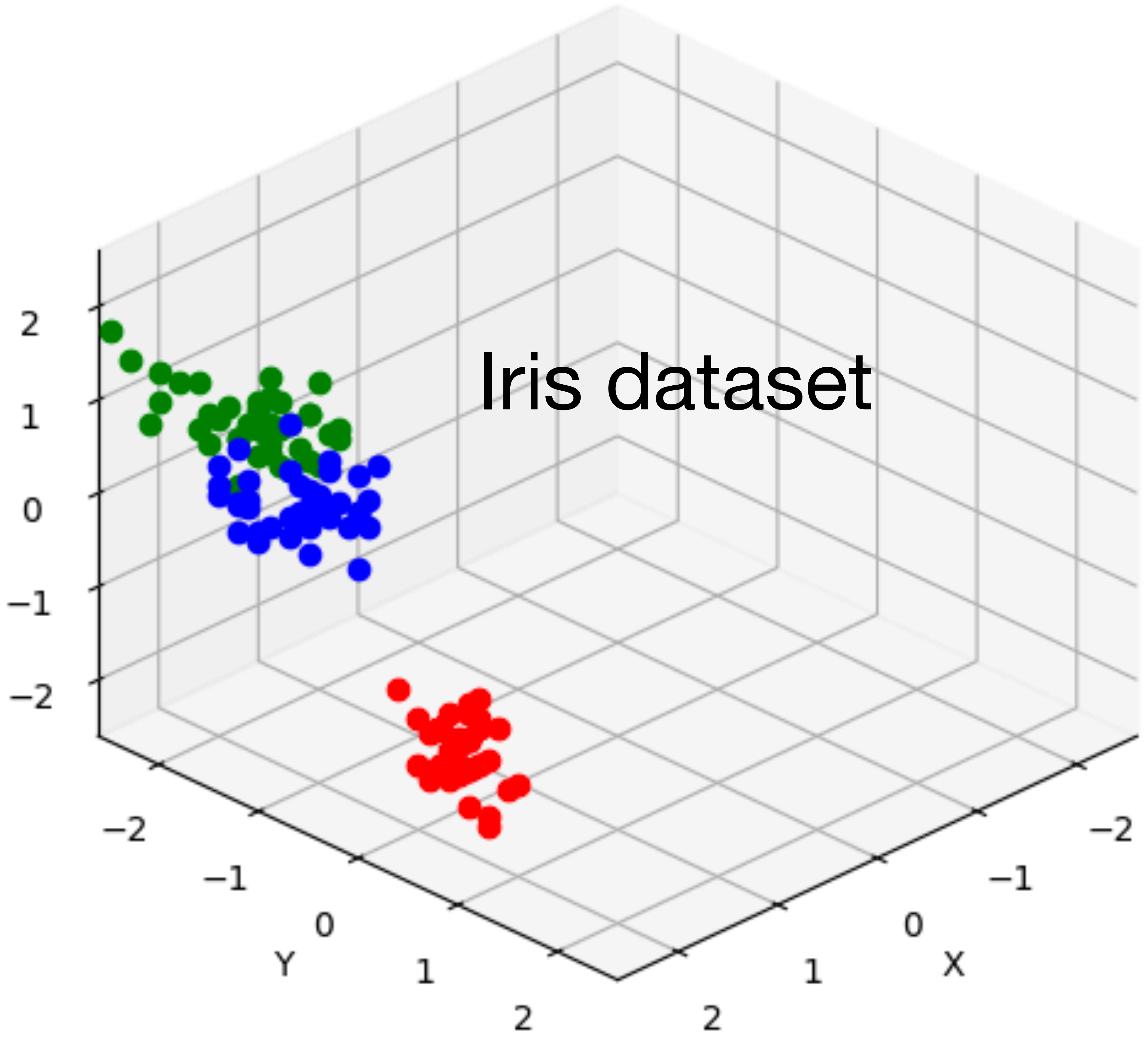
Tropical hyperplane ($\mathscr{A}_1$)



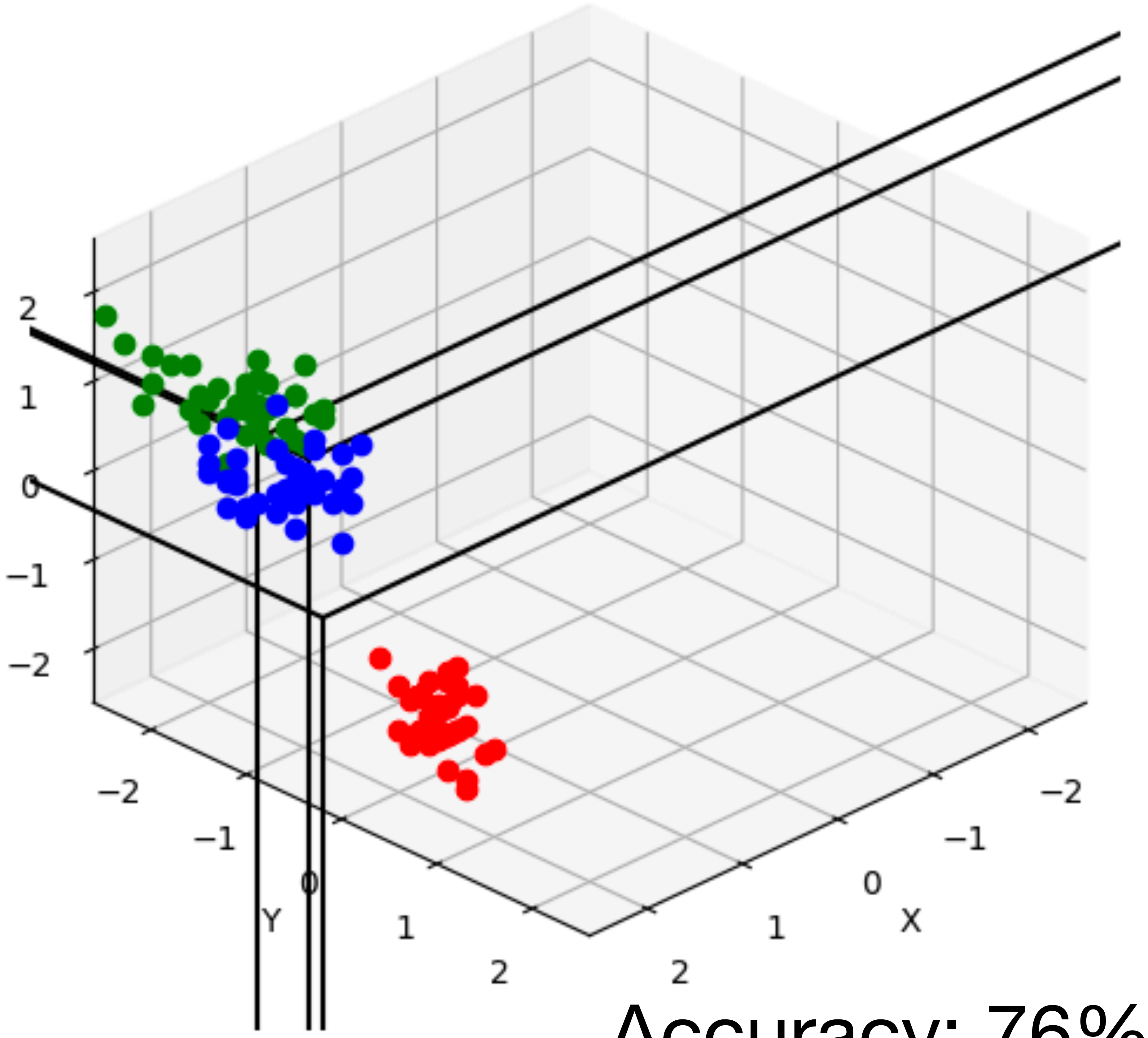
Accuracy: 80%

Topics being explored

2. Adding Features: Tropically Polynomial Decision Boundaries



One-vs-all hyperplanes ($\mathscr{A}_1$)



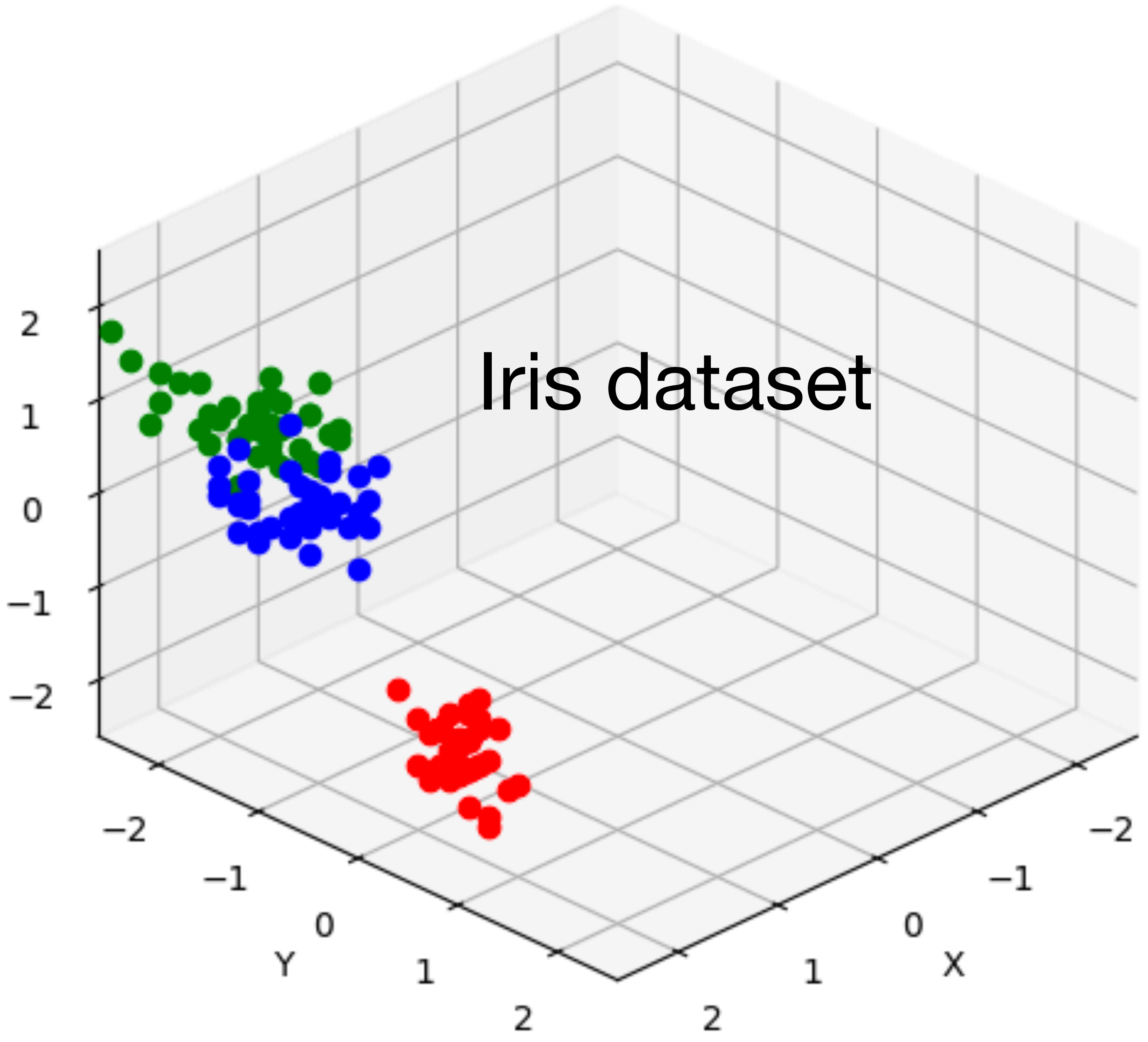
Accuracy: 76%

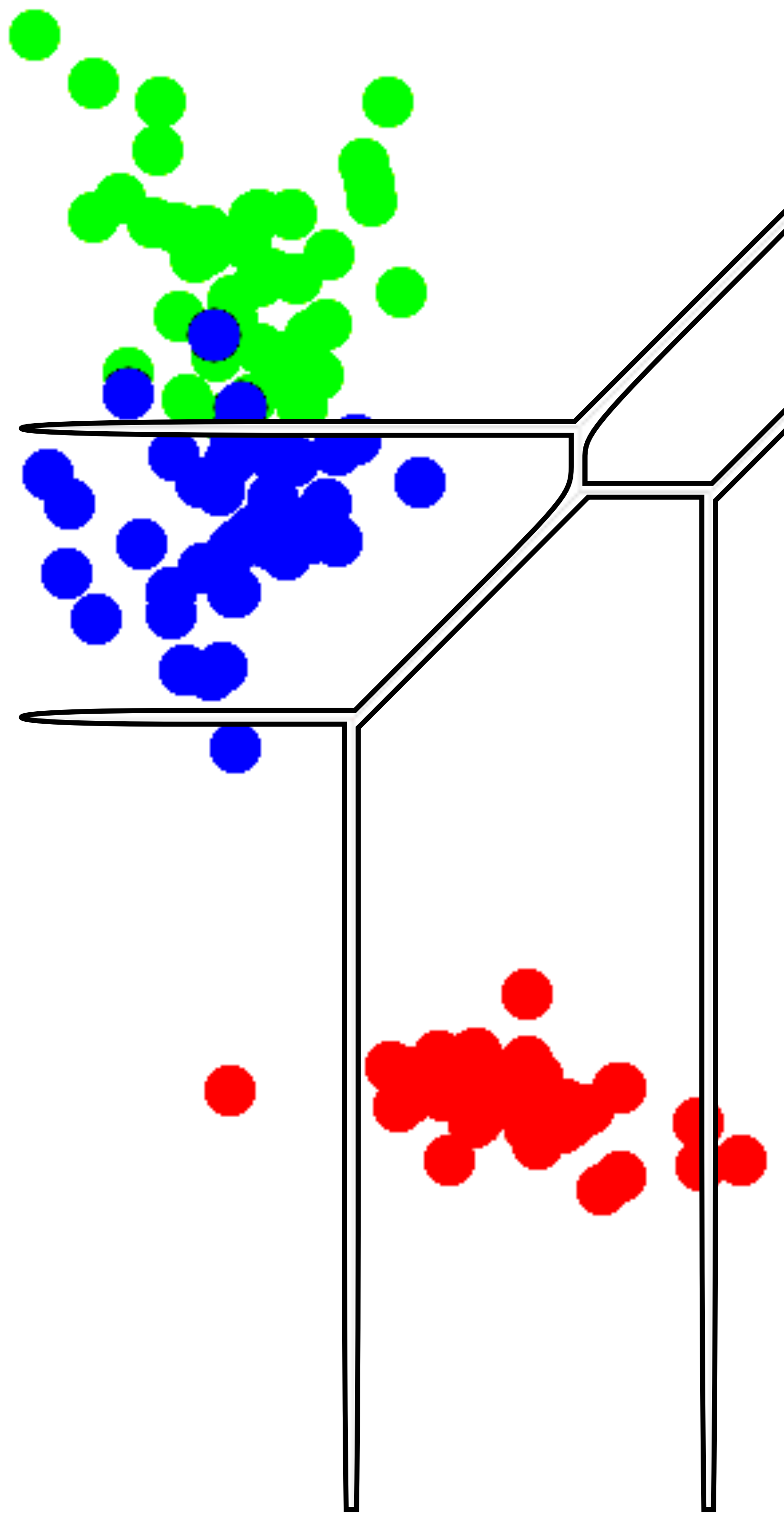
Blue class hard to separate

Topics being explored

2. Adding Features: Tropically Polynomial Decision Boundaries

Veronese trick ($\mathscr{A}_2$) : better!

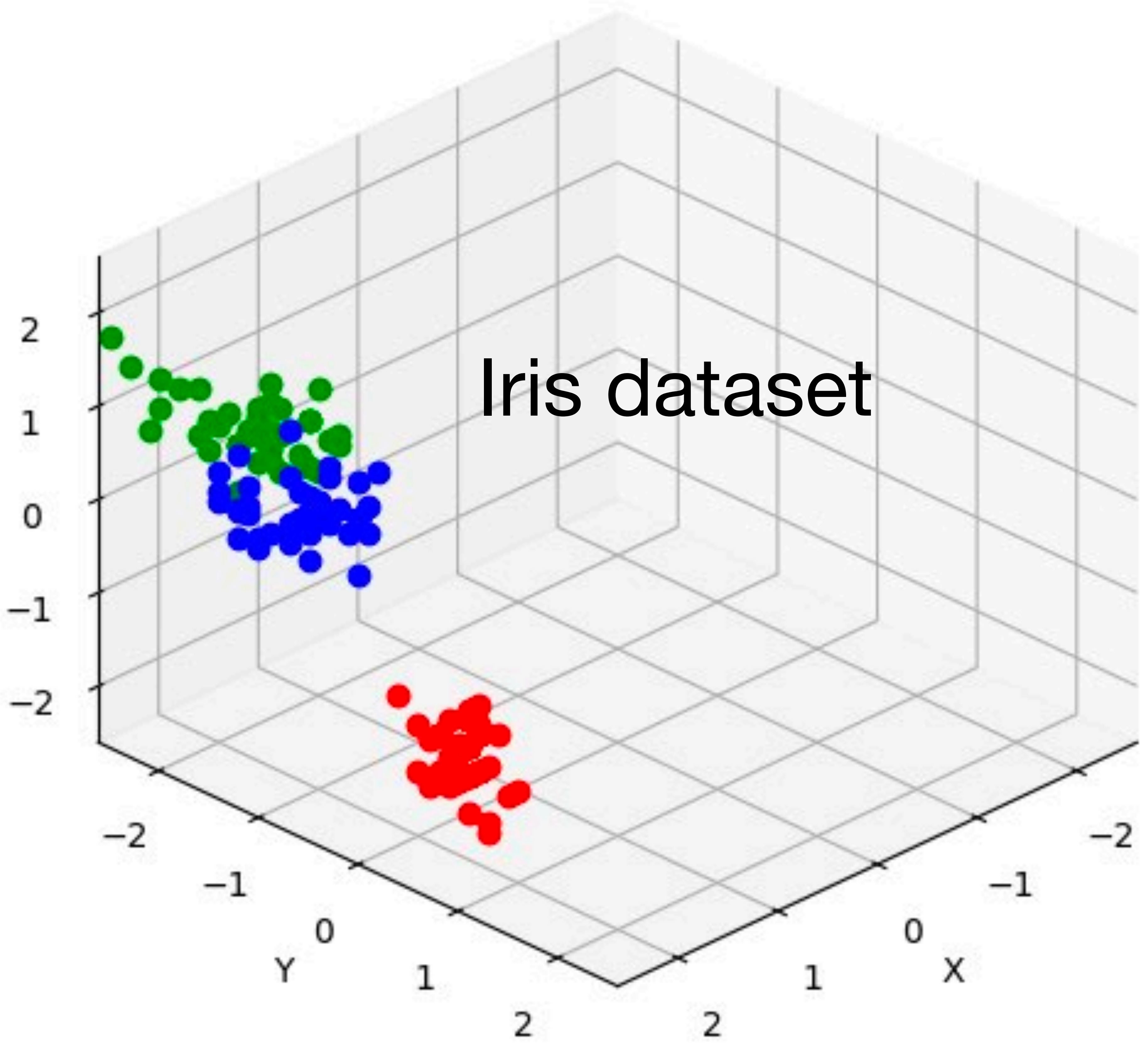




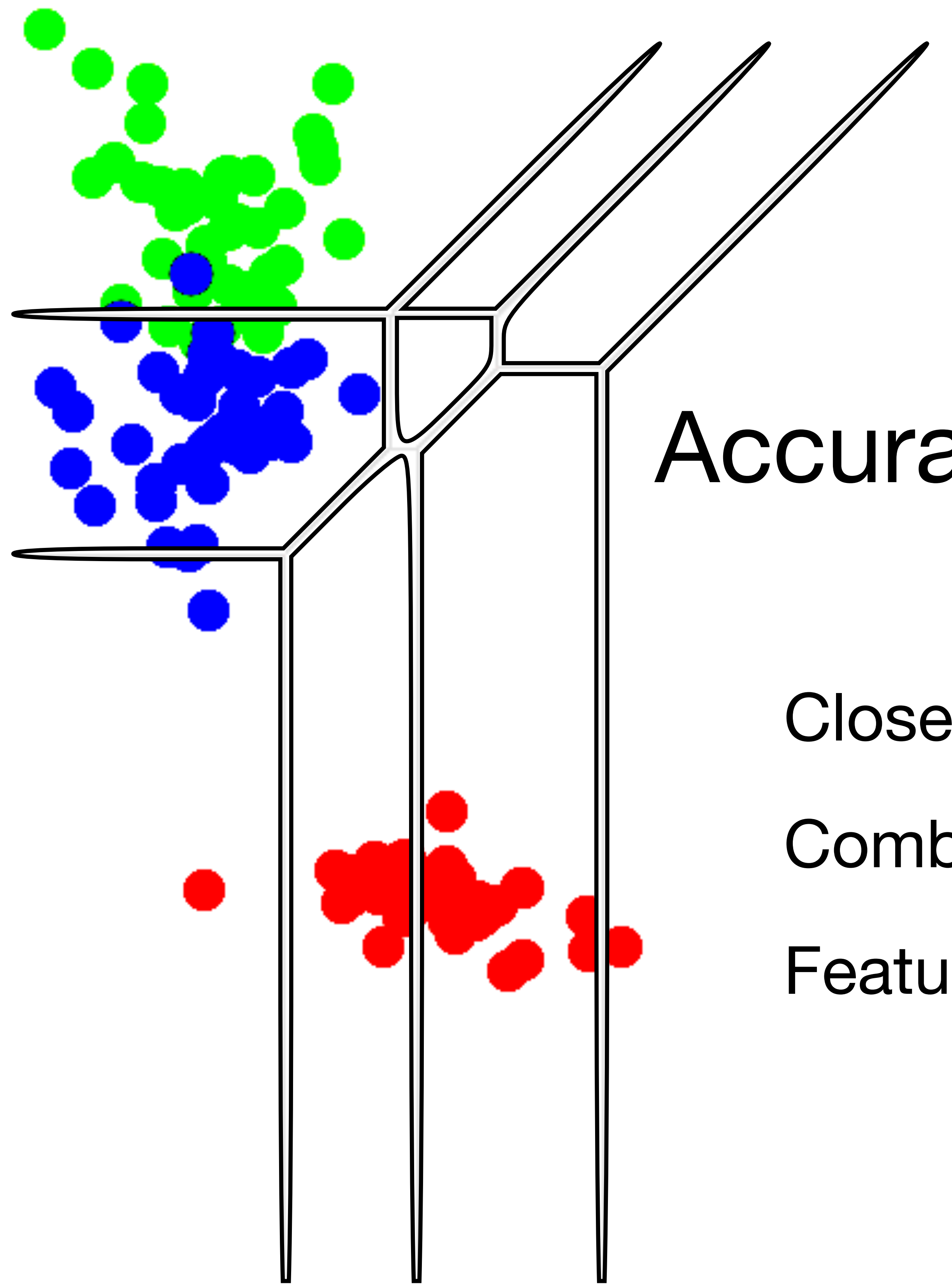
Generated with Polymake, approximate overlap

Topics being explored

2. Adding Features: Tropically Polynomial Decision Boundaries



Veronese trick ($\mathscr{A}_3$) : overfitting!



Accuracy: 98%

Closed cells

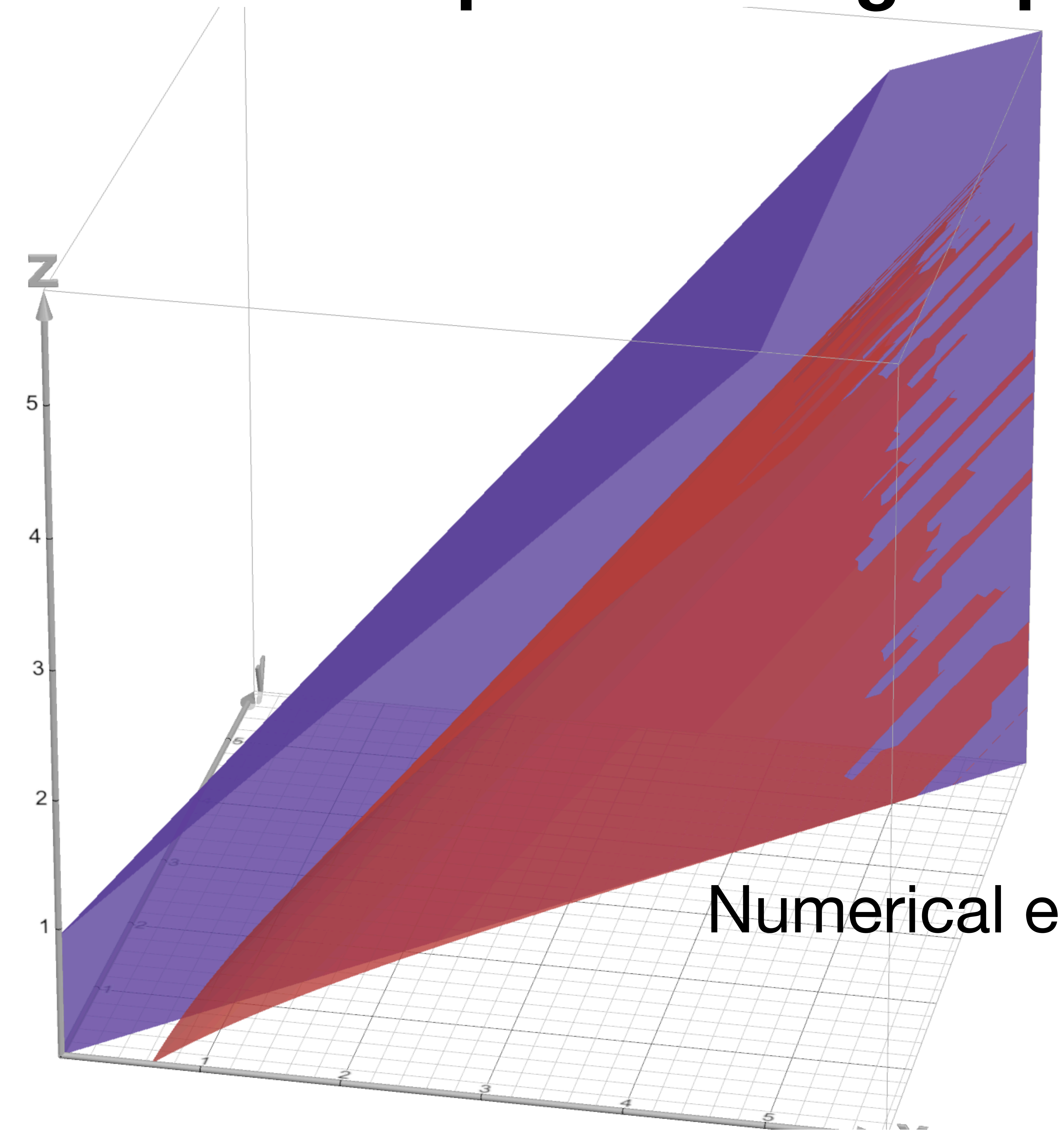
Combinatorics

Feature selection?

Generated with Polymake, approximate overlap

Topics being explored

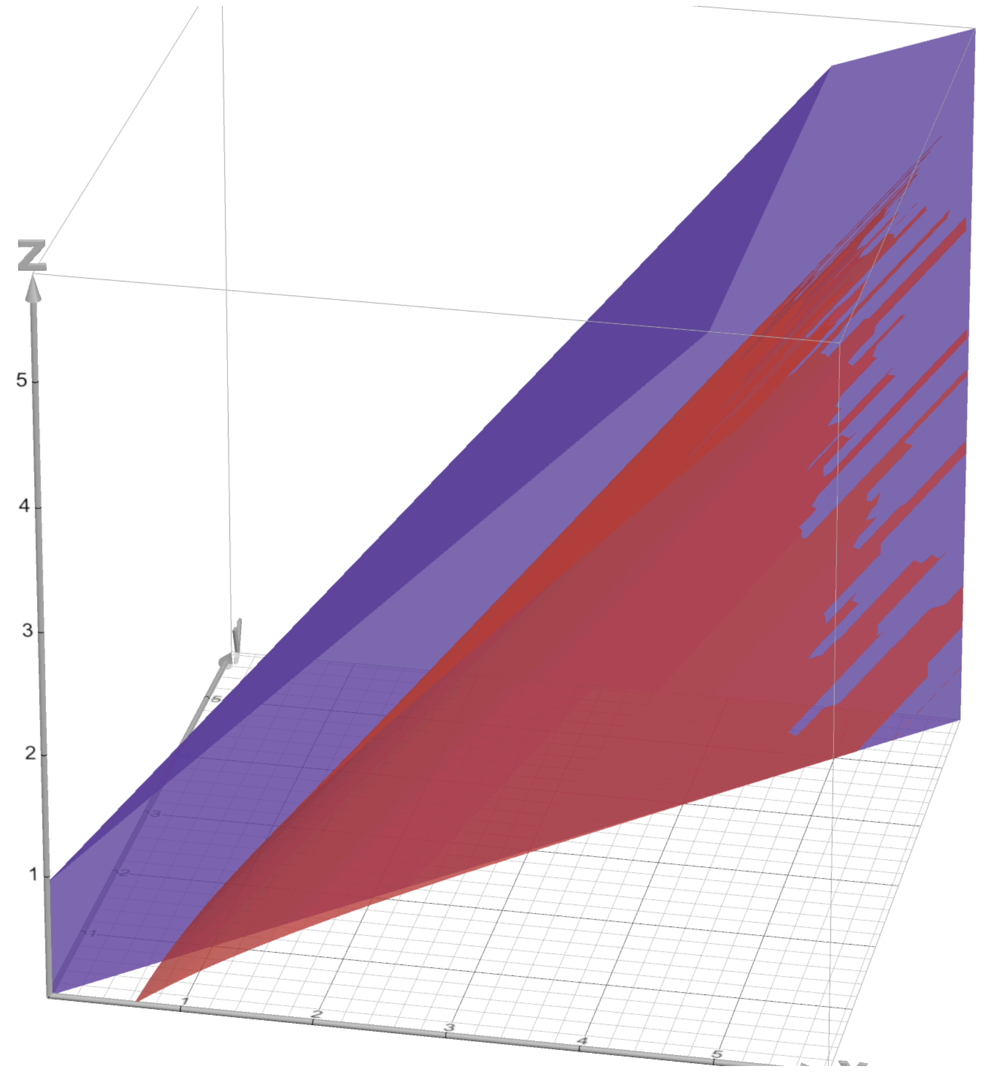
3. Linear Hyperplanes Look Tropical on Log Paper



Numerical errors!

Topics being explored

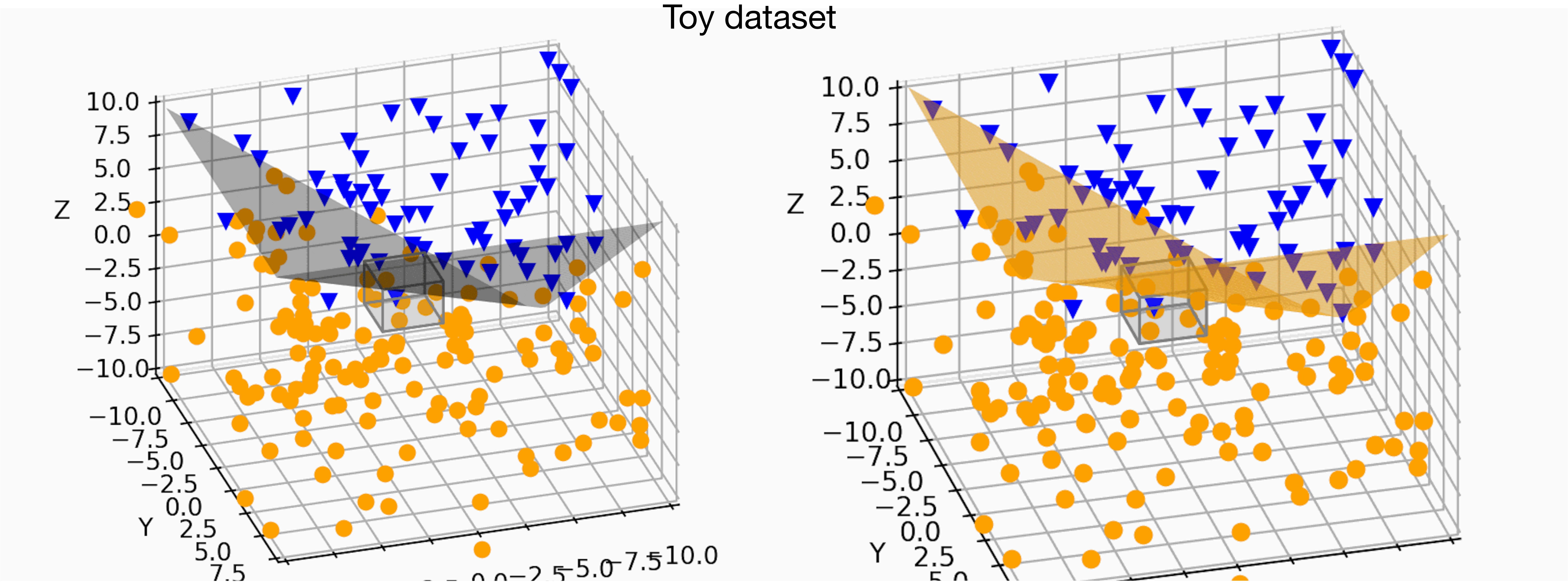
3. Linear Hyperplanes Look Tropical on Log Paper



- New data: $x^\beta = (x_{ij}^\beta := e^{\beta X_{ij}})_{ij}$.
- Hypersurface converges: $d_H\left(\beta^{-1}\log H^\beta, H^{trop}\right) \leq \beta^{-1}\log d$.
- When d high, β has to compensate, however if $\beta\bar{X}$ too high, numerical error!
- Tropical method will give higher numerical accuracy on higher dimensions.
- ***Bridging linear and tropical SVM theories?***

Topics being explored

3. Linear Hyperplanes Look Tropical on Log Paper



Tropical hyperplane

Classical hyperplane on logarithmic paper

$$d = 3, \quad \beta = 10$$