MAX PLANCK INSTITUTE
FOR MATHEMATICS IN THE SCIENCES

Separating points by piecewise-linear functions

Polyhedral and real tropical geometry of neural networks

Discrete Geometry Workshop

Oberwolfach

Marie-Charlotte Brandenburg
MPI MiS Leipzig (→ soon: KTH Stockholm)
23 January 2024

Georg Loho
FU Berlin | University of Twente

Guido Montúfar
UC LA | MPI MiS

*work in progress

Linear Classifiers and Hyperplane Arrangements

Setup

Given data points $D = \{p_1, \ldots, p_M\} \in \mathbb{R}^d$ in input space

- a linear classifier is a linear function $f : \mathbb{R}^d \to \mathbb{R}$
- f defines a hyperplane $\{x \in \mathbb{R}^d \mid f(x) = 0\}$ in input space separating $\{p_i \mid f(p_i) > 0\}$ from $\{p_i \mid f(p_i) < 0\}$
- f can be parametrized as $f(x) = \langle s, x \rangle + a$ for some fixed $s \in \mathbb{R}^d, a \in \mathbb{R}$.
- parameter space of linear classifiers is $\{(s, a) \mid s \in \mathbb{R}^d, a \in \mathbb{R}\} \cong \mathbb{R}^{d+1}$.

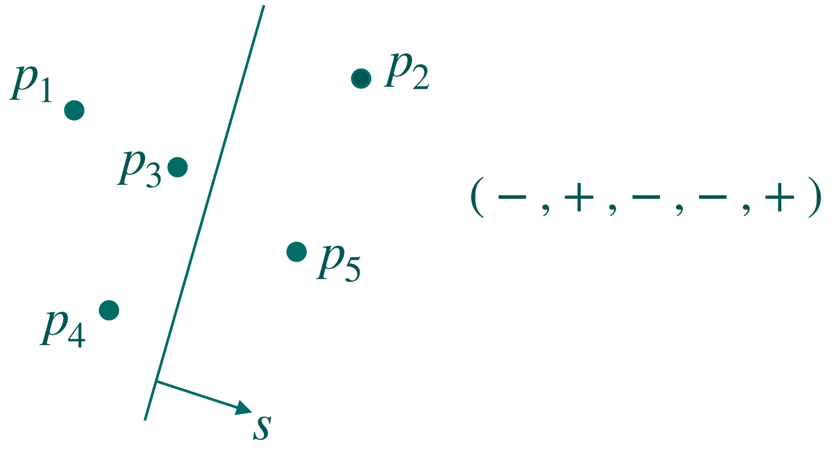
Classification by f:
(sgn($f(p_1)$), …, sgn($f(p_M)$)) $\in \{-,0,+\}^d$

Goal

Subdivide parameter space into cells, in which classifiers have the same classification

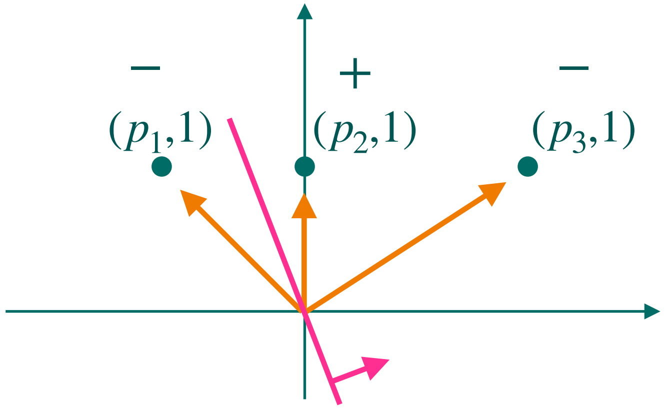
Theorem (Cover ’64, …)

These cells are chambers in the hyperplane arrangement $\bigcup_{p \in D} (p,1)^\perp \subseteq \mathbb{R}^{d+1}$ in parameter space

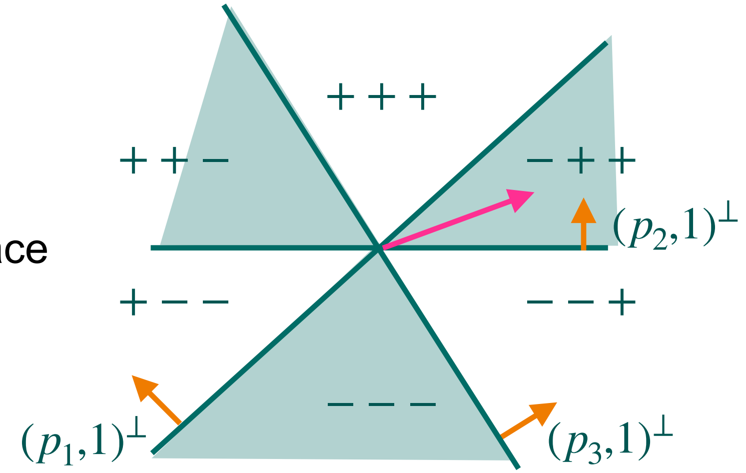


Linear Classifiers and Hyperplane Arrangements

input space $\times \mathbb{R}$



parameter space



Goal

Subdivide parameter space into cells, in which classifiers have the same classification

Theorem (Cover '64, ...)

These cells are chambers in the hyperplane arrangement $\bigcup_{p\in D}(p,1)^\perp \subseteq \mathbb{R}^{d+1}$ in parameter space

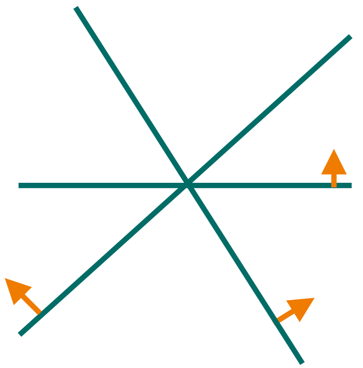
$\{+++,\ ++-,\ +--,\ ---,\ --+, -+-\}$
are the dichotomies of the data set

Fix a labelling $D = D^+ \sqcup D^-$

f makes a mistake at $p \in D^+$ if $f(p) < 0$
f makes a mistake at $p \in D^-$ if $f(p) > 0$

$0/1-$loss function counts number of mistakes of f

Linear Classifiers aka Hyperplane Arrangements



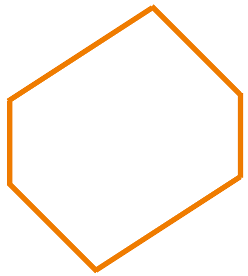
Theorem

Let $D \subset \mathbb{R}^d$ be a finite data set. Then

(i) the hyperplane arrangement $\mathscr{H}_D = \bigcup_{p \in D} (1,p)^{\perp}$ subdivides the parameter space into regions according to the represented dichotomies,

(ii) $\mathscr{H}_D$ induces the normal fan of the zonotope $P_D = \sum_{p \in D} \mathrm{conv}(\mathbf{0}, (1,p))$,

(iii) the dichotomies are the maximal covectors of the underlying realizable oriented matroid.



$\{+++,\ ++-,\ +--,\ ---,\ --+, -+-\}$

What happens for piecewise-linear functions?

From Linear to Piecewise Linear Classification

Linear function	Tropical rational functions	
Separation by hyperplane	Signed tropical hypersurface	input space
Covectors of oriented matroids	Activation patterns	parameter space $\Theta(d, N)$
Zonotope	Activation polytope	
Normal fan of zonotope	Activation fan	
Arrangement of hyperplanes	Arrangement of indecision surfaces	parameter space $\Theta(d, n, m)$
Polyhedral cone of perfect classifiers	Perfect classification fan	

TROPICAL INTERMEZZO AND RELU NEURAL NETWORKS

$a \oplus b = \max(a,b),\ a \odot b = a + b,\ a \oslash b = a - b,\ x^{\odot a} = a \cdot x$

classical rational function $\tilde{r}(x) = \left(\sum_{i=1}^{n} a_i x_1^{s_{i1}} \cdots x_d^{s_{id}} \right) / \left(\sum_{j=1}^{m} b_j x_1^{t_{j1}} \cdots x_d^{t_{jd}} \right)$

tropical rational function $r(x) = \left(\bigoplus_{i=1}^{n} a_i \odot x_1^{\odot s_{i1}} \odot \ldots \odot x_d^{\odot s_{id}} \right) \oslash \left(\bigoplus_{j=1}^{m} b_j \odot x_1^{\odot t_{j1}} \odot \ldots \odot x_d^{\odot t_{jd}} \right)$

$$= \max_{i=1,\ldots,n} \left(a_i + s_{i1}x_1 + \ldots + s_{id}x_d \right) - \max_{j=1,\ldots,m} \left(b_j + t_{j1}x_1 + \ldots + t_{jd}x_d \right)$$

$$= \max_{i=1,\ldots,n} \left(a_i + \langle s_i, x \rangle \right) - \max_{j=1,\ldots,m} \left(b_j + \langle t_j, x \rangle \right), \quad a_i, b_j \in \mathbb{R},\ s_i, t_j \in \mathbb{R}^d$$

$=$ difference of two convex piecewise linear functions

$(n,m) = (1,1)$ recovers linear classifiers

THEOREM (ARORA-BASU-MIANJY-MUKHERJEE, ZHANG-NAITZAT-LIM, '18)

A function $f\colon \mathbb{R}^d \to \mathbb{R}$ is a tropical rational function if and only if f can be represented by a ReLU neural network. The values d, n, m are related to the depth of the network.

From Linear to Piecewise Linear Classification

Linear function	Tropical rational functions	
Separation by hyperplane	Signed tropical hypersurface	} input space
Covectors of oriented matroids	Activation patterns	parameter space $\Theta(d, N)$
Zonotope	Activation polytope	
Normal fan of zonotope	Activation fan	
Arrangement of hyperplanes	Arrangement of indecision surfaces	parameter space $\Theta(d, n, m)$
Polyhedral cone of perfect classifiers	Perfect classification fan	

Decision Boundaries

$$g \oslash h = \left(\bigoplus_{i=1}^{n} a_i \odot x^{\odot s_i} \right) \oslash \left(\bigoplus_{j=1}^{m} b_j \odot x^{\odot t_j} \right)$$
$$= \max_{i=1,\dots,n} \left(a_i + \langle s_i, x \rangle \right) - \max_{j=1,\dots,m} \left(b_j + \langle t_j, x \rangle \right)$$

Decision boundary

$\mathcal{B}(g \oslash h) = \{ x \in \mathbb{R}^d \mid g(x) \oslash h(x) = 0 \}$

→ Polyhedral complex with $\leq n \cdot m$ linear pieces

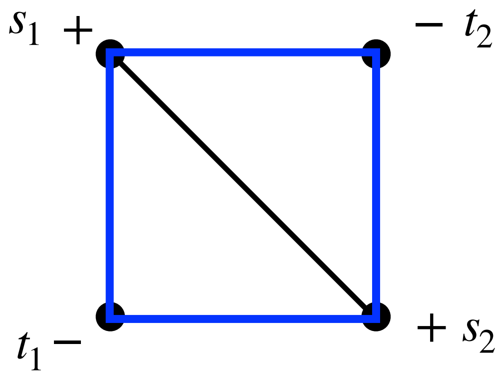
$g \oslash h : \mathbb{R}^2 \to \mathbb{R}$

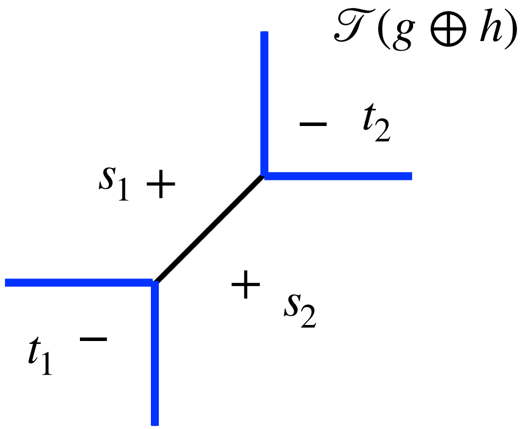
$$g(x) \oslash h(x) = \max \left(a_1 + \langle s_1, x \rangle, a_2 + \langle s_2, x \rangle \right) - \max \left(b_1 + \langle t_1, x \rangle, b_2 + \langle t_2, x \rangle \right),$$

Regular subdivision is given by $a_1, a_2, b_1, b_2 \in \mathbb{R}$

How to construct the decision boundary

- $g \oplus h = \max_{i \in [n],\, j \in [m]} \left(a_i + \langle s_i, x \rangle, b_j + \langle t_j, x \rangle \right)$
- Regular subdivision of signed Newton polytope $\mathcal{N}(g \oplus h)$
- Tropical hypersurface $\mathcal{T}(g \oplus h)$ is the codimension-1 skeleton of the dual complex
- Decision Boundary $\mathcal{B}(g \oslash h)$ is the sign-mixed subcomplex of $\mathcal{T}(g \oplus h)$





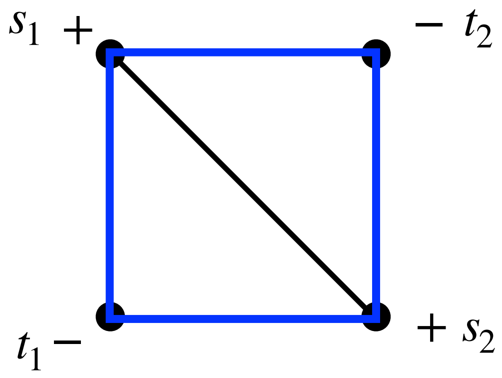
Decision Boundaries

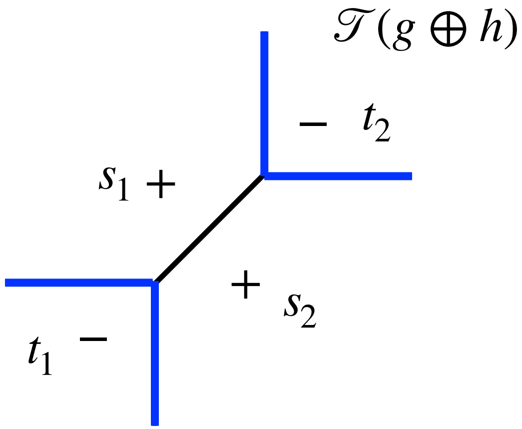
→ We can first consider tropical polynomials and later assign signs to terms

Fix number of terms N of tropical polynomial $f : \mathbb{R}^d \to \mathbb{R}, f(x) = \max_{i \in [N]} a_i + \langle s_i, x \rangle$

$f = f_\theta$ is defined through its parameters $\theta = (a_1, s_1, \ldots, a_N, s_N)$

Parameter space $\Theta(d, N) = \{(a_1, s_1, \ldots, a_N, s_N) \mid a_i \in \mathbb{R}, s_i \in \mathbb{R}^d\} \cong \mathbb{R}^{N(d+1)}$





From Linear to Piecewise Linear Classification

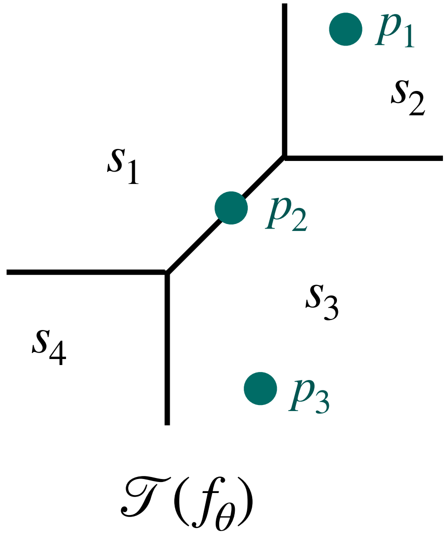
Linear function	Tropical rational functions	
Separation by hyperplane	Signed tropical hypersurface	input space
Covectors of oriented matroids	Activation patterns	parameter space $\Theta(d, N)$
Zonotope	Activation polytope	parameter space $\Theta(d, N)$
Normal fan of zonotope	Activation fan	parameter space $\Theta(d, N)$
Arrangement of hyperplanes	Arrangement of indecision surfaces	parameter space $\Theta(d, n, m)$
Polyhedral cone of perfect classifiers	Perfect classification fan	parameter space $\Theta(d, n, m)$

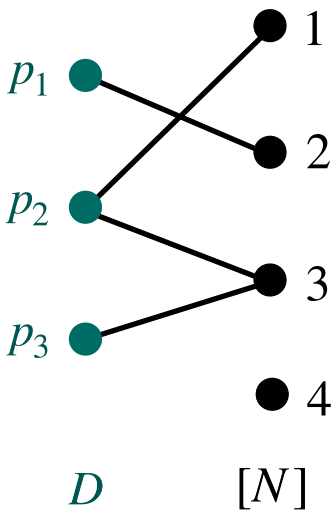
Classification By Tropical Polynomials

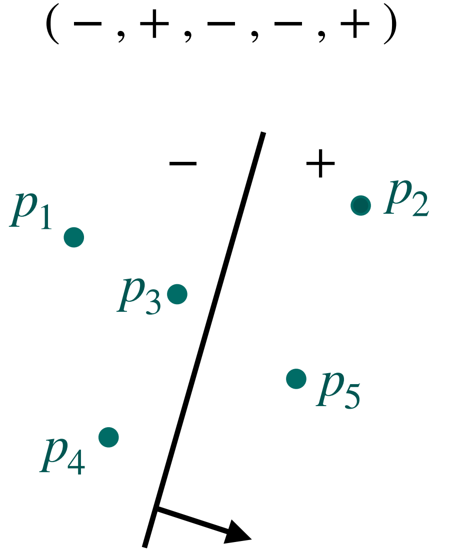
Given data points $D = \{p_1, \ldots, p_M\} \in \mathbb{R}^d$
tropical polynomial $f_\theta(x) = \max\limits_{i \in [N]} a_i + \langle s_i, x \rangle$,
parameter $\theta \in \Theta(d, N)$.

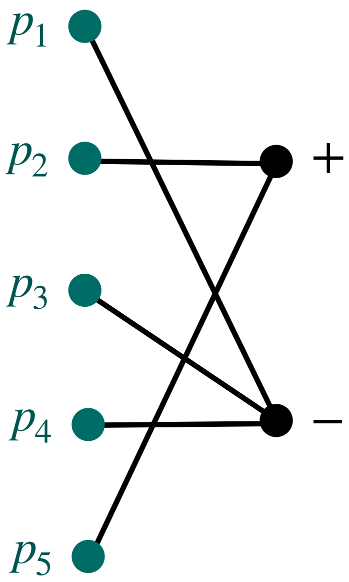
p lies in the region s_k of $\mathcal{T}(f_\theta)$
$\iff f_\theta(p) = a_k + \langle s_k, p \rangle$
$\iff p$ activates the k^{th} term

activation pattern: bipartite graph $G_\theta = (V, E)$
$V = D \sqcup [N]$
$E = \{pk \mid p \text{ activates } k^{th} \text{ term of } f_\theta\}$
$\rightarrow$ generalization of covectors of oriented matroids









Classification by Tropical Polynomials

activation cone of a bipartite graph G: $C(G) = \{\theta \mid G = G_\theta \text{ is activation pattern}\}$

activation fan = collection of all nonempty activation cones (over all bipartite graphs)
→ complete fan in $\Theta(d, N)$

A data point $p \in \mathbb{R}^d$ defines a $(N-1)$-dimensional simplex

$$\Delta(p) = \mathrm{conv}\big((1,p,0,\mathbf{0},\ldots,0,\mathbf{0}),\ (0,\mathbf{0},1,p,\ldots,0,\mathbf{0}),\ \ldots,\ (0,\mathbf{0},0,\mathbf{0},\ldots,1,p)\big)$$
$$\subseteq \Theta(d,N) \cong \mathbb{R}^{(d+1)N}$$

activation polytope $= \sum_{p\in D} \Delta(p)$ → generalization of zonotope

Theorem

The activation fan is the normal fan of the activation polytope.

What are the properties of the activation patterns appearing in an activation fan?

Activation Patterns

Theorem (Activation Patterns)

Let $\mathscr{G}$ be the set of activation patterns of the activation fan. Then $\mathscr{G}$ satisfies

- **(Zero)** $K_{N,D} \in \mathscr{G}$
- **(Symmetry)** $G \in \mathscr{G} \implies$ any graph isomorphic to G under the action of S_N is contained in $\mathscr{G}$ →
- **(Elimination)** If $G, H \in \mathscr{G}$, $p \in D$ then there exists a graph $F \in \mathscr{G}$ with $N(p;F) = N(p;G) \cup N(p;H)$ →
- **(Boundary)** For each $i \in [N]$ and G the bipartite graph with edges $E(G) = \{pi \mid p \in D\}$ holds $G \in \mathscr{G}$ →
- **(Comparability)** For any point $p \in D$ the comparability graph $\mathrm{CG}^p_{G,H}$ of any two patterns $G, H \in \mathscr{G}$ is acyclic

Definition (Tropical Oriented Matroid)

A tropical oriented matroid is a pair $([M], \mathscr{T})$, where $\mathscr{T} \subseteq \{(A_1, \ldots, A_M) \mid A_i \subseteq [N], i \in [M]\}$ are tropical covectors satisfying

- **(Elimination)** If $A, B \in T$ and $j \in [D]$ then there exists a type $C \in T$ with $C_j = A_j \cup B_j$ and $C_k \in \{A_k, B_k, A_k \cup B_k\}$ for all $k \in [D]$.
- **(Boundary)** For each $j \in [N]$ holds $(\{j\}, \ldots, \{j\}) \in T$
- **(Comparability)** The comparability graph $\mathrm{CG}_{A,B}$ of any two types A and B in T is acyclic.
- **(Surrounding)** If $A \in T$ the any refinement is also in T.

Definition (Oriented Matroid)

An oriented matroid is a pair $([M], \mathscr{C})$, where $\mathscr{C} \subseteq \{-,0,+\}^{[M]}$ are covectors satisfying

- → **(Zero)** $(0,\ldots,0) \in \mathscr{C}$
- → **(Symmetry)** $C \in \mathscr{C} \implies -C \in \mathscr{C}$
- **(Composition)** if $C, D \in \mathscr{C}$ then $(C \circ D) \in \mathscr{C}$
- → **(Elimination)** if $C, D \in \mathscr{C}$ and $i \in S(C,D)$ then there exists some $Z \in \mathscr{C}$ such that $Z_i = 0$ and $Z_j = (C \circ D)_j \forall j \in [M] \backslash S(C,D)$.

From Linear to Piecewise Linear Classification

Linear function	Tropical rational functions	
Separation by hyperplane	Signed tropical hypersurface	input space
Covectors of oriented matroids	Activation patterns	parameter space $\Theta(d, N)$
Zonotope	Activation polytope	
Normal fan of zonotope	Activation fan	
Arrangement of hyperplanes	Arrangement of indecision surfaces	parameter space $\Theta(d, n, m)$
Polyhedral cone of perfect classifiers	Perfect classification fan	

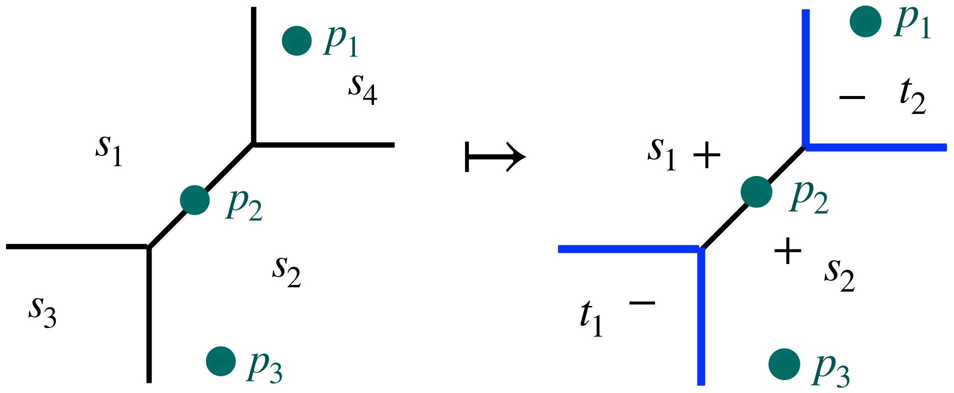
Dividing the parameter space

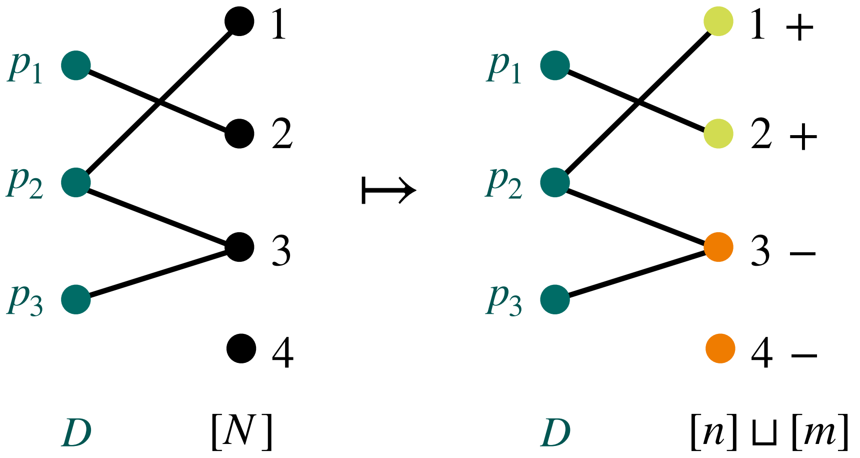
Fix a partition $n + m = N$. This induces

- a partition of the parameters $\Theta(d, N) \to \Theta(d, n, m)$

  $(a_1, s_1, \ldots, a_n, s_n, a_{n+1}, s_{n+1}, \ldots, a_N, s_N) \mapsto (a_1, s_1, \ldots, a_n, s_n, b_1, t_1, \ldots, b_m, t_m)$
- a partition into numerator and denominator

  $$f(x) = \max_{i \in [N]} a_i + \langle s_i, x \rangle \;\mapsto\; g(x) \oslash h(x) = \max_{i=1,\ldots,n} \left(a_i + \langle s_i, x\rangle\right) - \max_{j=1,\ldots,m} \left(b_j + \langle t_j, x \rangle\right)$$
- signs on regions in input space
- a coloring of nodes in activation patterns





Indecision Surfaces

What is the analogue of a hyperplane $(1,p)^{\perp} = \{(a,s) \mid a + \langle s,p \rangle = 0\}$
from the hyperplane arrangement?

The indecision surface of a data point $p \in D$ is $\mathcal{S}(p) = \{\theta \in \Theta(d,n,m) \mid (g_\theta \oslash h_\theta)(p) = 0\}$.

Theorem

The indecision surface is the sign-mixed subcomplex of the normal fan of $\Delta(p)$ with signs

$$\Delta(p) = \text{conv}\Big(\underbrace{\overset{+}{(1,p,0,\mathbf{0},\ldots,0,\mathbf{0})}, \ \ldots, \ \overset{+}{(0,\mathbf{0},\ldots,1,p,0,\mathbf{0},\ldots,0,\mathbf{0})}}_{n \text{ times}}, \underbrace{\overset{-}{(0,\mathbf{0},\ldots,0,\mathbf{0},1,p\ldots,0,\mathbf{0})}, \ \ldots, \ \overset{-}{(0,\mathbf{0},0,\mathbf{0},\ldots,1,p)}}_{m \text{ times}}\Big).$$

$\mathcal{S}(p)$ subdivides the parameter space into $\mathcal{S}^+(p) = \{\theta \mid (g_\theta \oslash h_\theta)(p) > 0\}$ and $\mathcal{S}^-(p)$.

The perfect Classification Fan

Fix a target labelling $D = D^+ \sqcup D^-$

$\text{label}(p) = +$ if $p \in D^+$,
$\text{label}(p) = -$ if $p \in D^-$

Theorem

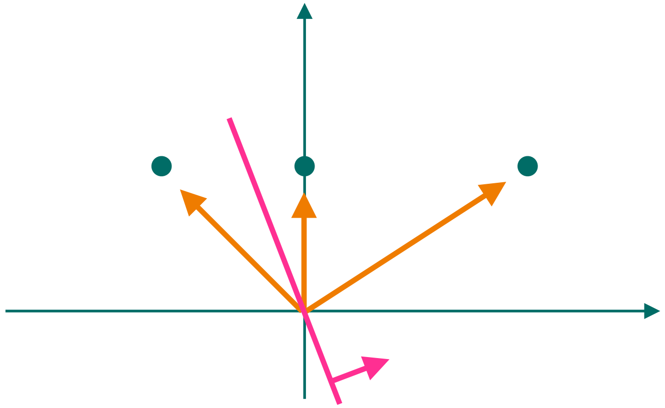
The perfect classification fan w.r.t $D^+ \sqcup D^-$ is

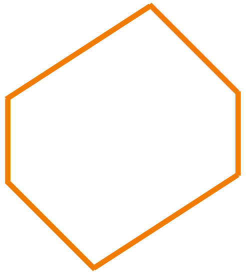
- $\bigcap_{p \in D} \mathcal{S}^{\text{label}(p)}(p)$
- the collection of all cones of the activation fan s.t. the activation pattern (bipartite graph) is compatible with the target labelling
- the set of solutions of a system of tropical polynomial inequalities (tropical semialgebraic set)

$\rightarrow$ noncomplete fan with many bad properties

What I would have told you at a different conference…

- connectivity of the perfect classification fan (through walls of codimension 1)
- local and global minima of the $0/1$-loss function (counts the number of mistakes in a classification)
- sublevel sets of the $0/1$-loss function as subfans of the activation fan
- …





THANK YOU