Optimizing Semi-Supervised Support Vector Machines using Semidefinite Programming

Joint work with Veronica Piccialli* and Antonio M. Sudoso

*Veronica Piccialli's work has been supported by PNRR MUR project PE0000013-FAIR

Future Artificial Intelligence Research

September 6, 2023

FWF Der Wissenschaftsfonds.

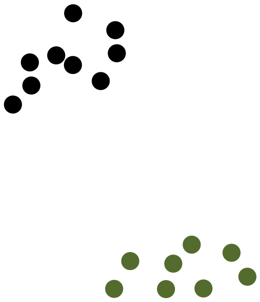
UNIVERSITÄT KLAGENFURT

Modeling - Analysis - Optimization DOC.FUNDS DOCTORAL SCHOOL

Support Vector Machines (SVMs) Vapnik & Chervonenkis (1963)

Input

- data points $\{(x_i, y_i)\}_{i=1}^n$ with $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, 1\}$



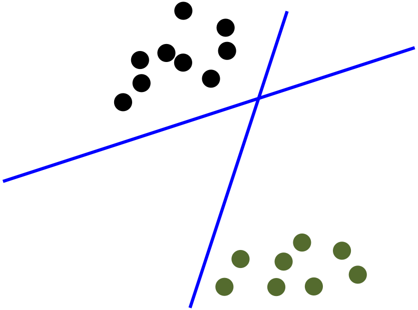
Support Vector Machines (SVMs) Vapnik & Chervonenkis (1963)

Input

- data points $\{(x_i, y_i)\}_{i=1}^n$ with $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, 1\}$

Goal/Output

- hyperplane $w^\top x + b = 0$ separating classes



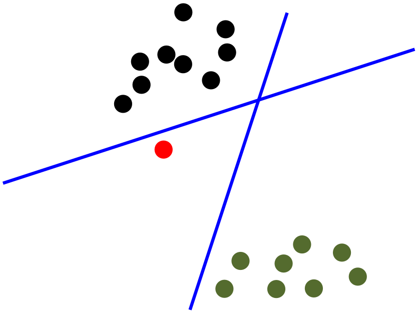
Support Vector Machines (SVMs) Vapnik & Chervonenkis (1963)

Input

- data points $\{(x_i, y_i)\}_{i=1}^n$ with $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, 1\}$

Goal/Output

- hyperplane $w^\top x + b = 0$ separating classes
- prediction model $y(x) = \text{sign}(w^\top x + b)$ for new data



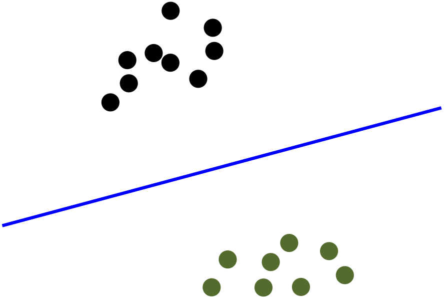
Support Vector Machines (SVMs) Vapnik & Chervonenkis (1963)

Input

- data points $\{(x_i, y_i)\}_{i=1}^n$ with $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, 1\}$

Goal/Output

- hyperplane $w^\top x + b = 0$ separating classes
- prediction model $y(x) = \text{sign}(w^\top x + b)$ for new data



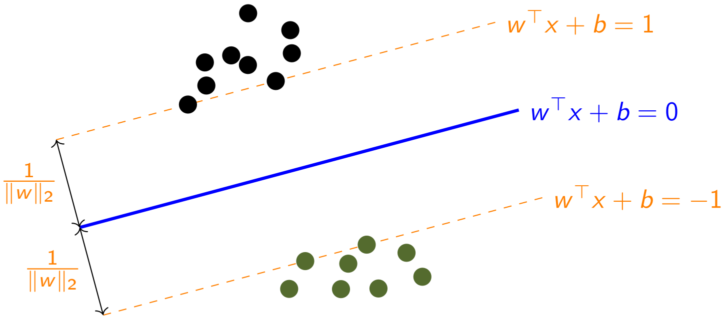
Support Vector Machines (SVMs) Vapnik & Chervonenkis (1963)

Input

- data points $\{(x_i, y_i)\}_{i=1}^n$ with $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, 1\}$

Goal/Output

- hyperplane $w^\top x + b = 0$ separating classes (maximum margin)
- prediction model $y(x) = \text{sign}(w^\top x + b)$ for new data



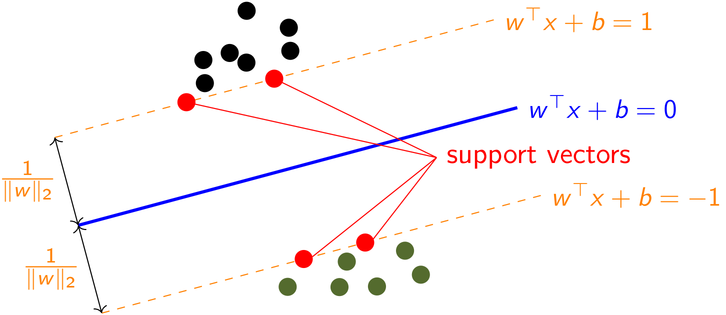
Support Vector Machines (SVMs) Vapnik & Chervonenkis (1963)

Input

- data points $\{(x_i, y_i)\}_{i=1}^n$ with $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, 1\}$

Goal/Output

- hyperplane $w^\top x + b = 0$ separating classes (maximum margin)
- prediction model $y(x) = \text{sign}(w^\top x + b)$ for new data



Hard Margin

Maximum hard margin hyperplane

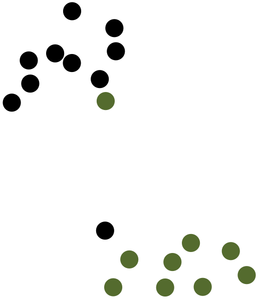
$$\begin{aligned} \min_{w,b} \quad & \tfrac{1}{2}\|w\|_2^2 \\ \text{s. t.} \quad & y_i(w^\top x_i + b) \geq 1,\ i = 1, \ldots, n \end{aligned}$$

Hard Margin

Maximum hard margin hyperplane

$$\min_{w,b} \quad \frac{1}{2}\|w\|_2^2$$
$$\text{s.t.} \quad y_i(w^\top x_i + b) \geq 1, \;\; i = 1, \ldots, n$$

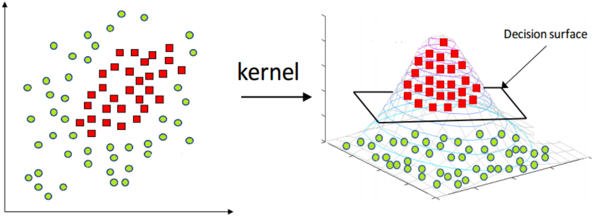
Question: What if data is **not** linearly separable?



The Kernel Trick Boser, Guyon, Vapnik (1992)

Kernel trick

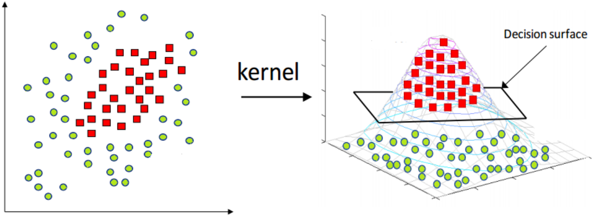
Map data into a higher-dimensional space via $\phi\colon \mathbb{R}^d \to \mathbb{R}^m$, $m \geq d$.
Then find a separating hyperplane in the new space.



The Kernel Trick Boser, Guyon, Vapnik (1992)

Kernel trick

Map data into a higher-dimensional space via $\phi\colon \mathbb{R}^d \to \mathbb{R}^m$, $m \geq d$.
Then find a separating hyperplane in the new space.



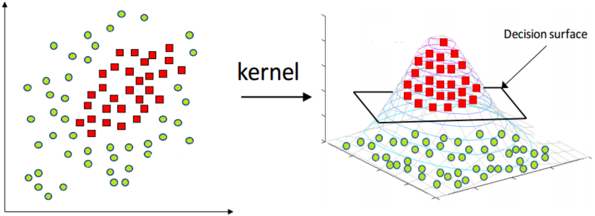
- linear or polynomial kernel, radial basis function kernel, ...
- no explicit mapping into higher dimension via kernel function

$$k(x_i, x_j) := \langle \phi(x_i), \phi(x_j) \rangle$$

The Kernel Trick Boser, Guyon, Vapnik (1992)

Kernel trick

Map data into a higher-dimensional space via $\phi\colon \mathbb{R}^d \to \mathbb{R}^m$, $m \geq d$. Then find a separating hyperplane in the new space.



- linear or polynomial kernel, radial basis function kernel, ...
- no explicit mapping into higher dimension via kernel function

$$k(x_i, x_j) := \langle \phi(x_i), \phi(x_j) \rangle$$

- separator is nonlinear in original space
- parameters must be chosen, risk of overfitting

Soft Margin Cortes & Vapnik (1995)

Maximum soft margin hyperplane w.r.t. $C > 0$

- data 'almost' linearly separable $\Rightarrow$ allow misclassifications
- introduce slack variables ξ_i and add penalty term to objective:

Soft Margin Cortes & Vapnik (1995)

Maximum soft margin hyperplane w.r.t. $C > 0$

- data 'almost' linearly separable $\Rightarrow$ allow misclassifications
- introduce slack variables ξ_i and add penalty term to objective:

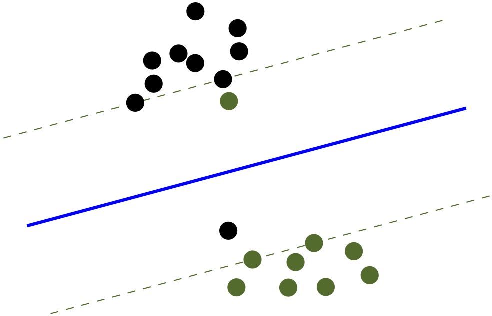
$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i(w^\top x_i + b) \geq 1 - \xi_i, \; i = 1, \dots, n \\ & \xi_i \geq 0, \; i = 1, \dots, n \end{aligned}$$

Soft Margin Cortes & Vapnik (1995)

Maximum soft margin hyperplane w.r.t. $C > 0$

- data 'almost' linearly separable ⇒ allow misclassifications
- introduce slack variables ξ_i and add penalty term to objective:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & y_i(w^\top x_i + b) \geq 1 - \xi_i, \; i = 1, \dots, n \\ & \xi_i \geq 0, \; i = 1, \dots, n \end{aligned}$$

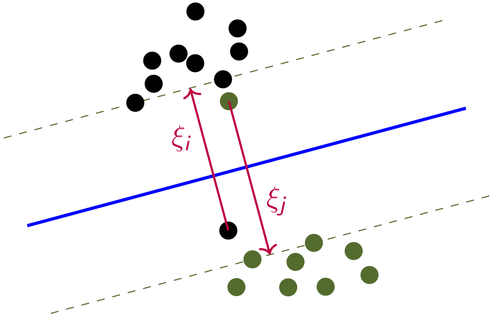


Soft Margin Cortes & Vapnik (1995)

Maximum soft margin hyperplane w.r.t. $C > 0$

- data 'almost' linearly separable ⇒ allow misclassifications
- introduce slack variables ξ_i and add penalty term to objective:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i(w^\top x_i + b) \geq 1 - \xi_i, \; i = 1, \ldots, n \\ & \xi_i \geq 0, \; i = 1, \ldots, n \end{aligned}$$

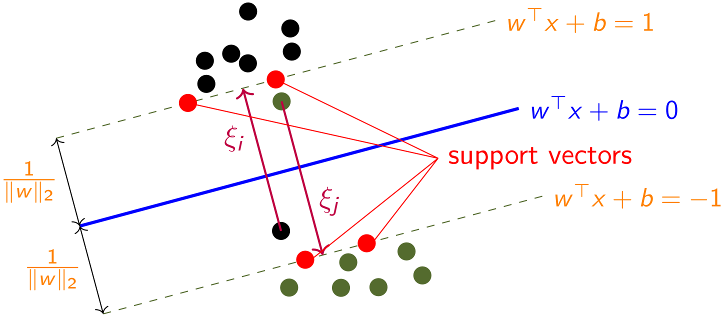


Soft Margin Cortes & Vapnik (1995)

Maximum soft margin hyperplane w.r.t. $C > 0$

- data 'almost' linearly separable $\Rightarrow$ allow misclassifications
- introduce slack variables ξ_i and add penalty term to objective:

$$\begin{aligned} \min_{w,b,\xi} \quad & \tfrac{1}{2}\|w\|_2^2 + C \textstyle\sum_{i=1}^n \xi_i \\ \text{s. t.} \quad & y_i(w^\top x_i + b) \geq 1 - \xi_i, \ \ i = 1, \dots, n \\ & \xi_i \geq 0, \ \ i = 1, \dots, n \end{aligned}$$



Summary: SVMs

Properties

- robust prediction technique
- applicable to very large data sets
- convex quadratic problem must be solved

Applications

- image processing and classification
- face detection, pattern recognition, ...
- "Support Vector Machines Applications" (Ma & Guo, 2014)

Summary: SVMs

Properties

- robust prediction technique
- applicable to very large data sets
- convex quadratic problem must be solved

Applications

- image processing and classification
- face detection, pattern recognition, . . .
- "Support Vector Machines Applications" (Ma & Guo, 2014)

Practical limitations

- supervised learning: all data must be labeled
- high costs, high time expenditure, limited resources, . . .

Semi-Supervised Support Vector Machines (S3VMs)

Bennett & Demiriz (1998)

Input

- n data points $x_i \in \mathbb{R}^d, \ i = 1, \dots, n$
- ℓ labeled points $\{(x_i, y_i)\}_{i=1}^{\ell}$ with $y_i \in \{-1, +1\}, \ i = 1, \dots, \ell$
- $n - \ell$ unlabeled points $\{x_i\}_{i=\ell+1}^{n}$

Semi-Supervised Support Vector Machines (S3VMs)

Bennett & Demiriz (1998)

Input

- n data points $x_i \in \mathbb{R}^d,\ i = 1, \ldots, n$
- ℓ labeled points $\{(x_i, y_i)\}_{i=1}^{\ell}$ with $y_i \in \{-1, +1\},\ i = 1, \ldots, \ell$
- $n - \ell$ unlabeled points $\{x_i\}_{i=\ell+1}^{n}$

Assumption

All data points are centered around the origin ($\Rightarrow b = 0$).

Semi-Supervised Support Vector Machines (S3VMs)

Bennett & Demiriz (1998)

Input

- n data points $x_i \in \mathbb{R}^d,\ i = 1, \dots, n$
- ℓ labeled points $\{(x_i, y_i)\}_{i=1}^{\ell}$ with $y_i \in \{-1, +1\},\ i = 1, \dots, \ell$
- $n - \ell$ unlabeled points $\{x_i\}_{i=\ell+1}^{n}$

Assumption

All data points are centered around the origin ($\Rightarrow b = 0$).

S3VM model with linear kernel

$$\begin{aligned} \min_{w,\xi,y^u} \quad & \tfrac{1}{2}\|w\|_2^2 + C \textstyle\sum_{i=1}^{n} \xi_i^2 \\ \text{s. t.} \quad & y_i\, w^\top x_i \geq 1 - \xi_i,\ i = 1, \dots, n \\ & y^u := (y_{\ell+1}, \dots, y_n) \in \{-1, +1\}^{n-\ell} \end{aligned}$$

Reformulation with Fewer Variables

Notation

- $K^* \succeq 0$ kernel matrix with $K^*_{ij} := k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$
- $K := K^* + \frac{1}{2C} I_n \succ 0$

Reformulation with Fewer Variables

Notation

- $K^* \succeq 0$ kernel matrix with $K^*_{ij} := k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$
- $K := K^* + \frac{1}{2C} I_n \succ 0$

Reformulation Bai & Yan (2016)

$$
\begin{aligned}
\min \quad & \tfrac{1}{2} v^\top K^{-1} v \\
\text{s. t.} \quad & y_i v_i \geq 1, \;\; i = 1, \ldots, \ell \\
& v_i^2 \geq 1, \;\; i = \ell + 1, \ldots, n \\
& v \in \mathbb{R}^n
\end{aligned}
\qquad (*)
$$

Reformulation with Fewer Variables

Notation

- $K^* \succeq 0$ kernel matrix with $K^*_{ij} := k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$
- $K := K^* + \frac{1}{2C} I_n \succ 0$

Reformulation Bai & Yan (2016)

$$\begin{aligned} \min \quad & \tfrac{1}{2} v^\top K^{-1} v \\ \text{s. t.} \quad & y_i v_i \geq 1, \; i = 1, \ldots, \ell \\ & v_i^2 \geq 1, \; i = \ell + 1, \ldots, n \\ & v \in \mathbb{R}^n \end{aligned} \tag{$*$}$$

- quadratic programming problem in continuous variables
- convex objective function
- nonconvex feasible set
- bound constraints: $y_i v_i \geq 1$ means either $v_i \leq -1$ or $v_i \geq 1$

Reformulation with Fewer Variables

Notation

- $K^* \succeq 0$ kernel matrix with $K^*_{ij} := k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$
- $K := K^* + \frac{1}{2C} I_n \succ 0$

Reformulation Bai & Yan (2016)

$$\begin{aligned} \min \quad & \tfrac{1}{2} v^\top K^{-1} v \\ \text{s. t.} \quad & y_i v_i \geq 1, \; i = 1, \ldots, \ell \\ & v_i^2 \geq 1, \; i = \ell + 1, \ldots, n \\ & v \in \mathbb{R}^n \end{aligned} \tag{$*$}$$

- quadratic programming problem in continuous variables
- convex objective function
- nonconvex feasible set
- bound constraints: $y_i v_i \geq 1$ means either $v_i \leq -1$ or $v_i \geq 1$

Goal: exact approach for $(*)$ using branch-and-cut

Global Optimization Problem

Textbook-like form

$$\begin{aligned} \min \quad & x^\top C x \\ \text{s. t.} \quad & L_i \le x_i \le U_i,\ i = 1, \ldots, n \\ & x_i^2 \ge 1,\ i = 1, \ldots, n \\ & x \in \mathbb{R}^n \end{aligned}$$

- rename variables

Global Optimization Problem

Textbook-like form

$$\begin{aligned} \min \quad & x^\top C x \\ \text{s.t.} \quad & L_i \le x_i \le U_i,\ i = 1, \dots, n \\ & x_i^2 \ge 1,\ i = 1, \dots, n \\ & x \in \mathbb{R}^n \end{aligned}$$

- rename variables
- C symmetric and positive definite

Global Optimization Problem

Textbook-like form

$$
\begin{aligned}
\min \quad & x^\top C x \\
\text{s. t.} \quad & L_i \leq x_i \leq U_i, \ i = 1, \dots, n \\
& x_i^2 \geq 1, \ i = 1, \dots, n \\
& x \in \mathbb{R}^n
\end{aligned}
$$

- rename variables
- C symmetric and positive definite
- $L_i \in \mathbb{R} \cup \{-\infty\}$ and $U_i \in \mathbb{R} \cup \{+\infty\}$

Global Optimization Problem

Textbook-like form

$$\begin{aligned} \min \quad & x^\top C x \\ \text{s. t.} \quad & L_i \leq x_i \leq U_i,\ i = 1, \ldots, n \\ & x_i^2 \geq 1,\ i = 1, \ldots, n \\ & x \in \mathbb{R}^n \end{aligned}$$

- rename variables
- C symmetric and positive definite
- $L_i \in \mathbb{R} \cup \{-\infty\}$ and $U_i \in \mathbb{R} \cup \{+\infty\}$
- some constraints redundant

Convex Relaxations

Quadratic programming (QP) relaxation

$$\begin{aligned} \min \quad & x^\top C x \\ \text{s.t.} \quad & L_i \leq x_i \leq U_i, \;\; i = 1, \ldots, n \\ & x \in \mathbb{R}^n \end{aligned} \tag{QP}$$

Convex Relaxations

Quadratic programming (QP) relaxation

$$\begin{aligned} \min \quad & x^\top C x \\ \text{s.t.} \quad & L_i \le x_i \le U_i, \; i = 1, \ldots, n \\ & x \in \mathbb{R}^n \end{aligned} \tag{QP}$$

Bai & Yan (2016) introduce $X := xx^\top$ and relax to $X - xx^\top \succeq 0$:

Semidefinite programming (SDP) relaxation

$$\begin{aligned} \min \quad & \langle C, X \rangle \\ \text{s.t.} \quad & X_{ii} \ge 1, \; i = 1, \ldots, n \\ & L_i \le x_i \le U_i, \; i = 1, \ldots, n \\ & \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq 0 \end{aligned} \tag{SDP}$$

Convex Relaxations

Quadratic programming (QP) relaxation

$$\begin{aligned} \min \quad & x^\top C x \\ \text{s. t.} \quad & L_i \leq x_i \leq U_i, \;\; i = 1, \ldots, n \\ & x \in \mathbb{R}^n \end{aligned} \tag{QP}$$

Bai & Yan (2016) introduce $X := xx^\top$ and relax to $X - xx^\top \succeq 0$:

Semidefinite programming (SDP) relaxation

$$\begin{aligned} \min \quad & \langle C, X \rangle \\ \text{s. t.} \quad & X_{ii} \geq 1, \;\; i = 1, \ldots, n \\ & L_i \leq x_i \leq U_i, \;\; i = 1, \ldots, n \\ & \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq 0 \end{aligned} \tag{SDP}$$

- Bai & Yan (2016): doubly nonnegative relaxation

Finding Good Upper Bounds

$$\begin{aligned} \min \quad & x^\top C x \\ \text{s.t.} \quad & y_i x_i \geq 1, \; i = 1, \ldots, \ell \\ & x_i^2 \geq 1, \; i = \ell + 1, \ldots, n \\ & x \in \mathbb{R}^n \end{aligned} \tag{P}$$

Let $\hat{x} \in \mathbb{R}^n$ be the optimal solution of (QP) or (SDP).

Finding Good Upper Bounds

$$\begin{aligned} \min \quad & x^\top C x \\ \text{s. t.} \quad & y_i x_i \geq 1, \; i = 1, \ldots, \ell \\ & x_i^2 \geq 1, \; i = \ell + 1, \ldots, n \\ & x \in \mathbb{R}^n \end{aligned} \tag{P}$$

Let $\hat{x} \in \mathbb{R}^n$ be the optimal solution of (QP) or (SDP).

Direct approach (fast)

Construct feasible solution $x \in \mathbb{R}^n$ for (P) via

$$x_i = \begin{cases} \hat{x}_i, & \text{if } |\hat{x}_i| \geq 1 \\ \text{sign}(\hat{x}_i), & \text{otherwise.} \end{cases}$$

Finding Good Upper Bounds

$$
\begin{aligned}
\min \quad & x^\top C x \\
\text{s. t.} \quad & y_i x_i \geq 1, \; i = 1, \dots, \ell \\
& x_i^2 \geq 1, \; i = \ell + 1, \dots, n \\
& x \in \mathbb{R}^n
\end{aligned} \tag{P}
$$

Let $\hat{x} \in \mathbb{R}^n$ be the optimal solution of (QP) or (SDP).

Direct approach (fast)

Construct feasible solution $x \in \mathbb{R}^n$ for (P) via

$$
x_i = \begin{cases} \hat{x}_i, & \text{if } |\hat{x}_i| \geq 1 \\ \operatorname{sign}(\hat{x}_i), & \text{otherwise.} \end{cases}
$$

Indirect approach (via solving convex QP)

1. Construct labeling vector $y \in \{-1, 1\}^n$ with $y_i = \operatorname{sign}(\hat{x}_i)$.
2. Solve convex QP with full labeling y.

Computing Finite Box Constraints $L_i \le x_i \le U_i$

Via QP

$$
\begin{aligned}
L_i/U_i := \min/\max \quad & x_i \\
\text{s.t.} \quad & L_i \le x_i \le U_i, \; i = 1, \ldots, n \\
& x^\top C x \le \text{UB} \\
& x \in \mathbb{R}^n
\end{aligned}
$$

Computing Finite Box Constraints $L_i \leq x_i \leq U_i$

Via QP

$$
\begin{aligned}
L_i / U_i := \min / \max \quad & x_i \\
\text{s. t.} \quad & L_i \leq x_i \leq U_i, \;\; i = 1, \ldots, n \\
& x^\top C x \leq \mathrm{UB} \\
& x \in \mathbb{R}^n
\end{aligned}
$$

Via SDP

$$
\begin{aligned}
L_i / U_i := \min / \max \quad & x_i \\
\text{s. t.} \quad & X_{ii} \geq 1, \;\; i = 1, \ldots, n \\
& L_i \leq x_i \leq U_i, \;\; i = 1, \ldots, n \\
& \langle C, X \rangle \leq \mathrm{UB} \\
& \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq 0
\end{aligned}
$$

Computing Finite Box Constraints $L_i \leq x_i \leq U_i$

Via QP

$$\begin{aligned} L_i/U_i := \min/\max \quad & x_i \\ \text{s.t.} \quad & L_i \leq x_i \leq U_i,\ i = 1, \dots, n \\ & x^\top C x \leq \mathrm{UB} \\ & x \in \mathbb{R}^n \end{aligned}$$

Via SDP

$$\begin{aligned} L_i/U_i := \min/\max \quad & x_i \\ \text{s.t.} \quad & X_{ii} \geq 1,\ i = 1, \dots, n \\ & L_i \leq x_i \leq U_i,\ i = 1, \dots, n \\ & \langle C, X \rangle \leq \mathrm{UB} \\ & \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq 0 \end{aligned}$$

Any further convex feasibility or optimality cuts can be added!

RLT Cuts Sherali & Adams (1998)

For any $x_i, x_j,\ i, j = 1, \ldots, n$, we have:

- $U_i - x_i \geq 0$
- $x_i - L_i \geq 0$
- $U_j - x_j \geq 0$
- $x_j - L_j \geq 0$

RLT Cuts Sherali & Adams (1998)

For any $x_i, x_j,\ i,j = 1,\dots,n$, we have:

- $U_i - x_i \geq 0$
- $x_i - L_i \geq 0$
- $U_j - x_j \geq 0$
- $x_j - L_j \geq 0$

- $(U_i - x_i)(x_j - L_j) \geq 0 \quad \Leftrightarrow \quad X_{ij} \leq U_i x_j + L_j x_i - U_i L_j$

RLT Cuts Sherali & Adams (1998)

For any $x_i, x_j,\ i,j = 1, \ldots, n$, we have:

- $U_i - x_i \geq 0$
- $x_i - L_i \geq 0$
- $U_j - x_j \geq 0$
- $x_j - L_j \geq 0$

- $(U_i - x_i)(x_j - L_j) \geq 0 \quad \Leftrightarrow \quad X_{ij} \leq U_i x_j + L_j x_i - U_i L_j$

RLT cuts

$$X_{ij} \geq U_i x_j + U_j x_i - U_i U_j$$
$$X_{ij} \geq L_i x_j + L_j x_i - L_i L_j$$
$$X_{ij} \leq L_i x_j + U_j x_i - L_i U_j$$
$$X_{ij} \leq U_i x_j + L_j x_i - U_i L_j$$

RLT Cuts Sherali & Adams (1998)

For any $x_i, x_j,\ i, j = 1, \dots, n$, we have:

- $U_i - x_i \geq 0$
- $x_i - L_i \geq 0$
- $U_j - x_j \geq 0$
- $x_j - L_j \geq 0$

- $(U_i - x_i)(x_j - L_j) \geq 0 \quad \Leftrightarrow \quad X_{ij} \leq U_i x_j + L_j x_i - U_i L_j$

RLT cuts

$$X_{ij} \geq U_i x_j + U_j x_i - U_i U_j$$
$$X_{ij} \geq L_i x_j + L_j x_i - L_i L_j$$
$$X_{ij} \leq L_i x_j + U_j x_i - L_i U_j$$
$$X_{ij} \leq U_i x_j + L_j x_i - U_i L_j$$

- cutting plane approach
- only between n and $3n$ cuts active at optimum
- significant stronger dualbounds

Triangle Inequalities Lambert (2023)

With three variables x_i, x_j, x_k:

$$(x_i - L_i)(x_j - L_j)(U_k - x_k) \geq 0$$

$$\Leftrightarrow$$

$$L_i L_j U_k - L_i L_j x_k - L_i U_k x_j + L_i x_j x_k - L_j U_k x_i + L_j x_i x_k + U_k x_i x_j \geq x_i x_j x_k$$

Triangle Inequalities Lambert (2023)

With three variables x_i, x_j, x_k:

$$(x_i - L_i)(x_j - L_j)(U_k - x_k) \geq 0$$

$$\Leftrightarrow$$

$$L_i L_j U_k - L_i L_j x_k - L_i U_k x_j + L_i x_j x_k - L_j U_k x_i + L_j x_i x_k + U_k x_i x_j \geq x_i x_j x_k$$

$$(U_i - x_i)(U_j - x_j)(x_k - L_k) \geq 0$$

$$\Leftrightarrow$$

$$x_i x_j x_k \geq L_k U_i U_j + L_k x_i x_j - U_i U_j x_k + U_j x_i x_k + U_i x_j x_k - L_k U_j x_i - L_k U_i x_j$$

Triangle Inequalities Lambert (2023)

With three variables x_i, x_j, x_k:

$$(x_i - L_i)(x_j - L_j)(U_k - x_k) \geq 0$$
$$\Leftrightarrow$$
$$L_iL_jU_k - L_iL_jx_k - L_iU_kx_j + L_ix_jx_k - L_jU_kx_i + L_jx_ix_k + U_kx_ix_j \geq x_ix_jx_k$$

$$(U_i - x_i)(U_j - x_j)(x_k - L_k) \geq 0$$
$$\Leftrightarrow$$
$$x_ix_jx_k \geq L_kU_iU_j + L_kx_ix_j - U_iU_jx_k + U_jx_ix_k + U_ix_jx_k - L_kU_jx_i - L_kU_ix_j$$

Triangle cut

$$(U_k - L_k)x_ix_j + (L_j - U_j)x_ix_k + (L_i - U_i)x_jx_k + L_iL_jU_k - L_kU_iU_j$$
$$+ (L_kU_j - L_jU_k)x_i + (L_kU_i - L_iU_k)x_j + (U_iU_j - L_iL_j)x_k \geq 0$$

Triangle Inequalities Lambert (2023)

With three variables x_i, x_j, x_k:

$$(x_i - L_i)(x_j - L_j)(U_k - x_k) \geq 0$$
$$\Leftrightarrow$$
$$L_i L_j U_k - L_i L_j x_k - L_i U_k x_j + L_i x_j x_k - L_j U_k x_i + L_j x_i x_k + U_k x_i x_j \geq x_i x_j x_k$$

$$(U_i - x_i)(U_j - x_j)(x_k - L_k) \geq 0$$
$$\Leftrightarrow$$
$$x_i x_j x_k \geq L_k U_i U_j + L_k x_i x_j - U_i U_j x_k + U_j x_i x_k + U_i x_j x_k - L_k U_j x_i - L_k U_i x_j$$

Triangle cut

$$(U_k - L_k) x_i x_j + (L_j - U_j) x_i x_k + (L_i - U_i) x_j x_k + L_i L_j U_k - L_k U_i U_j$$
$$+ (L_k U_j - L_j U_k) x_i + (L_k U_i - L_i U_k) x_j + (U_i U_j - L_i L_j) x_k \geq 0$$

- triangle cuts do not improve our dualbounds (using RLT cuts)

Marginals-based Bound Tightening Ryoo & Sahinidis (1995)

- UB: best known upper bound for nonconvex problem (P)
- LB: optimal value of SDP relaxation

Marginals-based Bound Tightening Ryoo & Sahinidis (1995)

- UB: best known upper bound for nonconvex problem (P)
- LB: optimal value of SDP relaxation

Bound tightening

If the constraint $L_i - x_i \leq 0$ is active at the optimal SDP solution with Lagrange multiplier $\lambda_i^L > 0$, then the inequality

$$x_i \geq U_i - \frac{\mathrm{UB} - \mathrm{LB}}{\lambda_i^L}$$

does not cutoff any optimal solution of (P) better than UB.

Marginals-based Bound Tightening Ryoo & Sahinidis (1995)

- UB: best known upper bound for nonconvex problem (P)
- LB: optimal value of SDP relaxation

Bound tightening

If the constraint $L_i - x_i \leq 0$ is active at the optimal SDP solution with Lagrange multiplier $\lambda_i^L > 0$, then the inequality

$$x_i \geq U_i - \frac{\mathrm{UB} - \mathrm{LB}}{\lambda_i^L}$$

does not cutoff any optimal solution of (P) better than UB.

- update L_i via $L_i := \max\left\{L_i, U_i - \frac{\mathrm{UB}-\mathrm{LB}}{\lambda_i^L}\right\}$ if $\lambda_i^L > 0$

Marginals-based Bound Tightening Ryoo & Sahinidis (1995)

- UB: best known upper bound for nonconvex problem (P)
- LB: optimal value of SDP relaxation

Bound tightening

If the constraint $L_i - x_i \leq 0$ is active at the optimal SDP solution with Lagrange multiplier $\lambda_i^L > 0$, then the inequality

$$x_i \geq U_i - \frac{\mathrm{UB} - \mathrm{LB}}{\lambda_i^L}$$

does not cutoff any optimal solution of (P) better than UB.

- update L_i via $L_i := \max\left\{L_i, U_i - \frac{\mathrm{UB}-\mathrm{LB}}{\lambda_i^L}\right\}$ if $\lambda_i^L > 0$
- update U_i via $U_i := \min\left\{U_i, L_i - \frac{\mathrm{UB}-\mathrm{LB}}{\lambda_i^U}\right\}$ if $\lambda_i^U > 0$

Applying to Main Diagonal

We know that $1 \leq X_{ii} \leq \max\{L_i^2, U_i^2\}$ must hold for $i = 1, \ldots, n$.

Applying to Main Diagonal

We know that $1 \leq X_{ii} \leq \max\{L_i^2, U_i^2\}$ must hold for $i = 1, \ldots, n$.

Lower bound on main diagonal

If $X_{ii} \geq \gamma \geq 1$ is active with Lagrange multiplier $\lambda > 0$, then update

$$L_i := \max\left\{L_i, -\sqrt{\gamma + \frac{\text{UB} - \text{LB}}{\lambda}}\right\}, \; U_i := \min\left\{U_i, \sqrt{\gamma + \frac{\text{UB} - \text{LB}}{\lambda}}\right\}.$$

Applying to Main Diagonal

We know that $1 \leq X_{ii} \leq \max\{L_i^2, U_i^2\}$ must hold for $i = 1, \ldots, n$.

Lower bound on main diagonal

If $X_{ii} \geq \gamma \geq 1$ is active with Lagrange multiplier $\lambda > 0$, then update

$$L_i := \max\left\{L_i, -\sqrt{\gamma + \frac{\mathrm{UB} - \mathrm{LB}}{\lambda}}\right\},\; U_i := \min\left\{U_i, \sqrt{\gamma + \frac{\mathrm{UB} - \mathrm{LB}}{\lambda}}\right\}.$$

Upper bound on main diagonal

If $X_{ii} \leq \gamma \leq \max\{L_i^2, U_i^2\}$ is active with Lagrange multiplier $\lambda > 0$:

- $\gamma - \frac{\mathrm{UB}-\mathrm{LB}}{\lambda} \geq 1$ and $L_i > -\sqrt{\gamma - \frac{\mathrm{UB}-\mathrm{LB}}{\lambda}} \;\Rightarrow\; L_i := \max\{L_i, 1\}$
- $\gamma - \frac{\mathrm{UB}-\mathrm{LB}}{\lambda} \geq 1$ and $U_i < \sqrt{\gamma - \frac{\mathrm{UB}-\mathrm{LB}}{\lambda}} \;\Rightarrow\; U_i := \min\{U_i, -1\}$

Applying to Main Diagonal

We know that $1 \leq X_{ii} \leq \max\{L_i^2, U_i^2\}$ must hold for $i = 1, \ldots, n$.

Lower bound on main diagonal

If $X_{ii} \geq \gamma \geq 1$ is active with Lagrange multiplier $\lambda > 0$, then update

$$L_i := \max\left\{L_i, -\sqrt{\gamma + \frac{\text{UB} - \text{LB}}{\lambda}}\right\}, \; U_i := \min\left\{U_i, \sqrt{\gamma + \frac{\text{UB} - \text{LB}}{\lambda}}\right\}.$$

Upper bound on main diagonal

If $X_{ii} \leq \gamma \leq \max\{L_i^2, U_i^2\}$ is active with Lagrange multiplier $\lambda > 0$:

- $\gamma - \frac{\text{UB}-\text{LB}}{\lambda} \geq 1$ and $L_i > -\sqrt{\gamma - \frac{\text{UB}-\text{LB}}{\lambda}} \;\Rightarrow\; L_i := \max\{L_i, 1\}$
- $\gamma - \frac{\text{UB}-\text{LB}}{\lambda} \geq 1$ and $U_i < \sqrt{\gamma - \frac{\text{UB}-\text{LB}}{\lambda}} \;\Rightarrow\; U_i := \min\{U_i, -1\}$

Projecting box constraints

$L_i > -1 \;\Rightarrow\; L_i := \max\{L_i, 1\}$ and $U_i < 1 \;\Rightarrow\; U_i := \min\{U_i, -1\}$

Preliminary Branch-and-Cut Results

- most fractional branching
  - choose $i \in \arg\min_i \{|x_i|: L_i < 0,\ U_i > 0\}$
  - add $U_i \leq -1$ in one subproblem and $L_i \geq 1$ in the other
- SDP + RLT + bound tightening using MOSEK

Datasets used for 5-fold cross-validation

- **ionosphere:** $n = 280,\ d = 32$, almost linearly separable
- **arrhythmia:** $n = 361,\ d = 191$, not linearly separable

Preliminary Branch-and-Cut Results

- linear kernel
- $C = 1$

$$\text{gap} = (\text{UB} - \text{LB})/\text{UB}$$

instance	30% labeled			60% labeled		
	root gap	nodes	gap	root gap	nodes	gap
ionosphere-0	6.24%	61	0.1%	2.60%	75	0.1%
ionosphere-1	9.15%	187	0.1%	0.94%	7	0.1%
ionosphere-2	4.82%	103	0.1%	0.70%	5	0.1%
ionosphere-3	15.38%	326	6.79%	1.55%	23	0.1%
ionosphere-4	7.20%	127	0.1%	1.20%	9	0.1%
arrhythmia-0	21.90%	15	20.51%	9.12%	83	6.57%
arrhythmia-1	20.65%	15	18.24%	5.64%	82	3.42%
arrhythmia-2	28.11%	15	25.45%	5.39%	82	2.46%
arrhythmia-3	24.21%	15	22.67%	3.01%	90	0.88%
arrhythmia-4	18.55%	15	16.99%	5.13%	103	2.90%

Future Work

Relaxation:

- test QP + RLT relaxation
- further strengthen relaxation (disjunctive cuts?)

Implementation:

- faster and parallelized
- use other solver than MOSEK
- Lagrangian relaxation to dualize cuts

Future Work

Relaxation:

- test QP + RLT relaxation
- further strengthen relaxation (disjunctive cuts?)

Implementation:

- faster and parallelized
- use other solver than MOSEK
- Lagrangian relaxation to dualize cuts

Thank you!