RUHR-UNIVERSITÄT BOCHUM

RUB

A Close Look to Margin Complexity and Related Parameters

Michael Kallweit, Hans Ulrich Simon

Lehrstuhl Mathematik & Informatik

Ruhr-University Bochum

COLT 2011

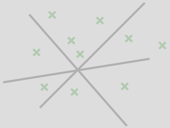
July 9, 2011

Large margin classification

- Sign matrices represent finite concept classes (columns corresponds to functions, rows to instances)

$$\begin{pmatrix} +1 & -1 & +1 & -1 \\ -1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ -1 & +1 & +1 & +1 \end{pmatrix}$$

- Linear separable concept classes admit linear arrangements (concepts as hyperplanes, instances as vectors)



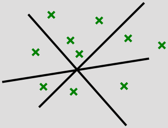
- A "large" margin is good ($\rightsquigarrow$ small generalization error)

Large margin classification

- Sign matrices represent finite concept classes (columns corresponds to functions, rows to instances)

$$\begin{pmatrix} +1 & -1 & +1 & -1 \\ -1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ -1 & +1 & +1 & +1 \end{pmatrix}$$

- Linear separable concept classes admit linear arrangements (concepts as hyperplanes, instances as vectors)



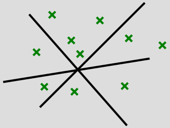
- A "large" margin is good ($\rightsquigarrow$ small generalization error)

Large margin classification

- Sign matrices represent finite concept classes (columns corresponds to functions, rows to instances)

$$\begin{pmatrix} +1 & -1 & +1 & -1 \\ -1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ -1 & +1 & +1 & +1 \end{pmatrix}$$

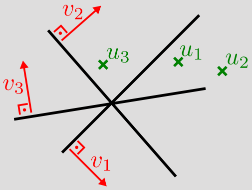
- Linear separable concept classes admit linear arrangements (concepts as hyperplanes, instances as vectors)



- A "large" margin is good ($\rightsquigarrow$ small generalization error)

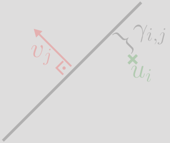
Margin parameters

$$A \in \mathbb{R}^{m \times n}$$
$$\mathscr{A} = (U; V) = (u_1, \ldots, u_n; v_1, \ldots, v_m),$$
$$u_i, v_j \in \mathbb{R}^d, \quad \|u_i\|_2, \|v_j\|_2 \leqslant 1$$
$$\mathrm{sign}(U^\top V) = A$$



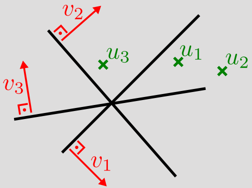
- margins of a linear arrangement:

$$\gamma_{i,j}(A \mid \mathscr{A}) := \langle u_i, v_j \rangle \cdot A_{i,j}$$

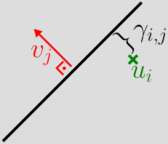


- margin complexity: $\mathrm{mc}(A) := \left(\max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A \mid \mathscr{A}) \right)^{-1}$

Margin parameters



$$A \in \mathbb{R}^{m \times n}$$
$$\mathscr{A} = (U;V) = (u_1,\dots,u_n;v_1,\dots,v_m),$$
$$u_i, v_j \in \mathbb{R}^d, \quad \|u_i\|_2, \|v_j\|_2 \leqslant 1$$
$$\operatorname{sign}(U^\top V) = A$$



- margins of a linear arrangement:
  $$\gamma_{i,j}(A \mid \mathscr{A}) := \langle u_i, v_j \rangle \cdot A_{i,j}$$
- margin complexity: $\operatorname{mc}(A) := \left(\max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A \mid \mathscr{A}) \right)^{-1}$

Margin parameters

- best total margin: $\gamma_2^*(A) = \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(A \mid \mathscr{A})$

- Y-average margin complexity for Y with $\sum_{i,j} y_{i,j} = 1$:

$$\overline{\mathrm{mc}}_Y(A) := \frac{1}{\gamma_2^*(Y \circ A)} = \left(\max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A \mid \mathscr{A}) \right)^{-1}$$

- average margin complexity: $\overline{\mathrm{mc}}(A) := \overline{\mathrm{mc}}_{\text{uniform}}(A) = \frac{m \cdot n}{\gamma_2^*(A)}$
- rank-one average margin complexity: $\overline{\mathrm{mc}}_{pq^\top}(A) = \frac{1}{\gamma_2^*(pq^\top \circ A)}$

Margin parameters

- best total margin: $\gamma_2^*(A) = \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(A \mid \mathscr{A})$

- Y-average margin complexity for Y with $\sum_{i,j} y_{i,j} = 1$:

$$\overline{\mathrm{mc}}_Y(A) := \frac{1}{\gamma_2^*(Y \circ A)} = \left(\max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A \mid \mathscr{A}) \right)^{-1}$$

- average margin complexity: $\overline{\mathrm{mc}}(A) := \overline{\mathrm{mc}}_{\mathrm{uniform}}(A) = \frac{m \cdot n}{\gamma_2^*(A)}$

- rank-one average margin complexity: $\overline{\mathrm{mc}}_{pq^\top}(A) = \frac{1}{\gamma_2^*(pq^\top \circ A)}$

Margin parameters

- best total margin: $\gamma_2^*(A) = \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(A \mid \mathscr{A})$
- Y-average margin complexity for Y with $\sum_{i,j} y_{i,j} = 1$:

$$\overline{\mathrm{mc}}_Y(A) := \frac{1}{\gamma_2^*(Y \circ A)} = \left(\max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A \mid \mathscr{A}) \right)^{-1}$$

- average margin complexity: $\overline{\mathrm{mc}}(A) := \overline{\mathrm{mc}}_{\text{uniform}}(A) = \frac{m \cdot n}{\gamma_2^*(A)}$
- rank-one average margin complexity: $\overline{\mathrm{mc}}_{pq^\top}(A) = \frac{1}{\gamma_2^*(pq^\top \circ A)}$

Our Contribution

"Min-Max" Theorems for Margin Complexity

$$\mathrm{mc}(A) \quad \xleftrightarrow{\text{duality}} \quad \max_Y \overline{\mathrm{mc}}_Y(A)$$

$$\max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A \mid \mathscr{A}) \quad = \quad \min_Y \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A \mid \mathscr{A})$$

$$\overline{\mathrm{mc}}_{p,MIN}(A) \quad \xleftrightarrow{\text{duality}} \quad \max_q \overline{\mathrm{mc}}_{pq^\top}(A)$$

$$\max_{\mathscr{A}} \min_j \sum_i p_i \gamma_{i,j}(A \mid \mathscr{A}) \quad = \quad \min_q \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(pq^\top \circ A \mid \mathscr{A})$$

Our Contribution

Some Variants of the Forster bound

already known: $$\frac{\sqrt{mn}}{\|A\|_2} \leqslant \overline{\mathrm{mc}}(A)$$

new upper bound: $$\overline{\mathrm{mc}}(A) \leqslant \frac{mn}{\|A\|_{tr}}$$

new lower bound for a variant: $$\frac{1}{\|\mathrm{diag}(p)^{1/2} A \,\mathrm{diag}(q)^{1/2}\|_2} \leqslant \overline{\mathrm{mc}}_{pq^\top}(A)$$

Our Contribution

Connections to complexity measures in statistical learning models

$$\max_{Y} \overline{\mathrm{mc}}_{Y}(A) \quad \xleftrightarrow{\text{poly}} \quad \mathrm{CSQdim}(A)$$

$$\max_{p,q} \overline{\mathrm{mc}}_{pq^\top}(A) \quad \xleftrightarrow{\text{poly}} \quad \mathrm{SQdim}(A)$$

Margin maximization and its dual

Theorem

$$\begin{aligned} & \mathrm{mc}(A) &=& \max_Y \overline{\mathrm{mc}}_Y(A) \\ \Leftrightarrow \quad & \max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A|\mathscr{A}) &=& \min_Y \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A|\mathscr{A}) \end{aligned}$$

Proof sketch

1. *Express both margin maximization terms as standard SDP problems*
2. *Compare their dual forms*

Margin maximization and its dual

Theorem

$$\begin{aligned} & \mathrm{mc}(A) &=& \max_Y \overline{\mathrm{mc}}_Y(A) \\ \Leftrightarrow \quad & \max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A|\mathscr{A}) &=& \min_Y \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A|\mathscr{A}) \end{aligned}$$

Proof sketch

1. *Express both margin maximization terms as standard SDP problems*
2. *Compare their dual forms*

Proof sketch of $\mathrm{mc}(A) = \max_Y \overline{\mathrm{mc}}_Y(A)$

Primal optimization forms:

Ingredient 1

$X = W^\top \cdot W$

$= [U\ V]^\top \cdot [U\ V]$

$= \begin{bmatrix} U^\top U & U^\top V \\ (U^\top V)^\top & V^\top V \end{bmatrix}$

SDP for $\max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A|\mathscr{A})$

$\min_{X,\mu,s} \quad -\mu$

s.t. $\quad \forall k:\ X_{k,k} = 1$

$\forall i,j:\ A_{i,j}(X_{i,m+j} + X_{m+j,i}) - s_{i,j} = 2\mu$

$X \succeq 0\ ,\ s_{i,j} \geqslant 0\ ,\ \mu \geqslant 0$

SDP for $\max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A|\mathscr{A})$

$\min_X \quad -\frac{1}{2} \cdot \sum_{i,j} (Y \circ A)_{i,j} (X_{i,m+j} + X_{m+j,i})$

s.t. $\quad \forall k:\ X_{k,k} = 1$

$X \succeq 0$

Proof sketch of $\mathrm{mc}(A) = \max_Y \overline{\mathrm{mc}}_Y(A)$

Dual optimization forms:

SDP for $\mathrm{mc}(A)^{-1} = \max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A \mid \mathscr{A})$

$$\min_{Y,y} \quad \frac{1}{2}\sum_k y_k$$

s.t.

$$\sum_{i,j} y_{i,j} = 1$$

$$y_{i,j} \geqslant 0$$

$$\begin{bmatrix} \mathrm{diag}(y_1,\ldots,y_m) & -(Y \circ A) \\ -(Y \circ A)^\top & \mathrm{diag}(y_{m+1},\ldots,y_{m+n}) \end{bmatrix} \succeq 0$$

Ingredient 2

Duality theory for SDPs

SDP for $\overline{\mathrm{mc}}_Y(A)^{-1} = \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A \mid \mathscr{A})$

$$\min_{y} \quad \frac{1}{2} \cdot \sum_k y_k$$

s.t.

$$\begin{bmatrix} \mathrm{diag}(y_1,\ldots,y_m) & -(Y \circ A) \\ -(Y \circ A)^\top & \mathrm{diag}(y_{m+1},\ldots,y_{m+n}) \end{bmatrix} \succeq 0$$

Proof sketch of $\mathrm{mc}(A) = \max_Y \overline{\mathrm{mc}}_Y(A)$

Dual optimization forms:

Ingredient 2

Duality theory for SDPs

SDP for $\mathrm{mc}(A)^{-1} = \max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A \mid \mathscr{A})$

$$\min_{Y,y} \quad \frac{1}{2}\sum_k y_k$$

s.t.

$$\sum_{i,j} y_{i,j} = 1$$

$$y_{i,j} \geqslant 0$$

$$\begin{bmatrix} \mathrm{diag}(y_1,\ldots,y_m) & -(Y \circ A) \\ -(Y \circ A)^\top & \mathrm{diag}(y_{m+1},\ldots,y_{m+n}) \end{bmatrix} \succeq 0$$

SDP for $\overline{\mathrm{mc}}_Y(A)^{-1} = \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A \mid \mathscr{A})$

$$\min_{y} \quad \frac{1}{2} \cdot \sum_k y_k$$

s.t.

$$\begin{bmatrix} \mathrm{diag}(y_1,\ldots,y_m) & -(Y \circ A) \\ -(Y \circ A)^\top & \mathrm{diag}(y_{m+1},\ldots,y_{m+n}) \end{bmatrix} \succeq 0$$

Proof sketch of $\mathrm{mc}(A) = \max_Y \overline{\mathrm{mc}}_Y(A)$

Ingredient 2

Duality theory for SDPs

Dual optimization forms:

SDP for $\mathrm{mc}(A)^{-1} = \max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A \mid \mathscr{A})$

$$\min_{Y,y} \quad \frac{1}{2}\sum_k y_k$$

s.t.

$$\sum_{i,j} y_{i,j} = 1$$

$$y_{i,j} \geqslant 0$$

$$\begin{bmatrix} \mathrm{diag}(y_1,\dots,y_m) & -(Y \circ A) \\ -(Y \circ A)^\top & \mathrm{diag}(y_{m+1},\dots,y_{m+n}) \end{bmatrix} \succeq 0$$

SDP for $\overline{\mathrm{mc}}_Y(A)^{-1} = \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A \mid \mathscr{A})$

$$\min_y \quad \frac{1}{2}\cdot\sum_k y_k$$

s.t.

$$\begin{bmatrix} \mathrm{diag}(y_1,\dots,y_m) & -(Y \circ A) \\ -(Y \circ A)^\top & \mathrm{diag}(y_{m+1},\dots,y_{m+n}) \end{bmatrix} \succeq 0$$

"Support Vectors" for Linear Arrangements

Slater Constraint Qualification $\rightsquigarrow \exists$ maximizer $\mathscr{A}^*$ for $\min_{i,j} \gamma_{i,j}(A \mid \mathscr{A})$

Min-Max theorem $\rightsquigarrow \exists$ minimizer Y^* for $\max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A \mid \mathscr{A})$.

Optimal margin: $\gamma^* := \min_{i,j} \gamma_{i,j}(A \mid \mathscr{A}^*) = \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y^* \circ A \mid \mathscr{A}^*)$

"Hard part" of matrix A:

$$K(A \mid \mathscr{A}^*) := \{(i,j) : \gamma_{i,j}(A \mid \mathscr{A}^*) = \gamma^*\}$$

$$K(A) := \bigcap_{\mathscr{A}^*} K(A \mid \mathscr{A}^*)$$

"Support Vectors" for Linear Arrangements

Slater Constraint Qualification $\rightsquigarrow \exists$ maximizer $\mathscr{A}^*$ for $\min_{i,j} \gamma_{i,j}(A \mid \mathscr{A})$

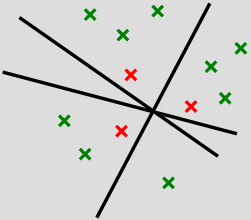
Min-Max theorem $\rightsquigarrow \exists$ minimizer Y^* for $\max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A \mid \mathscr{A})$.

Optimal margin: $\gamma^* := \min_{i,j} \gamma_{i,j}(A \mid \mathscr{A}^*) = \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y^* \circ A \mid \mathscr{A}^*)$

"Hard part" of matrix A:

$$K(A \mid \mathscr{A}^*) := \{(i,j) : \gamma_{i,j}(A \mid \mathscr{A}^*) = \gamma^*\}$$

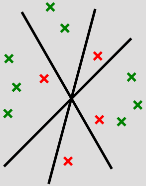
$$K(A) := \bigcap_{\mathscr{A}^*} K(A \mid \mathscr{A}^*)$$

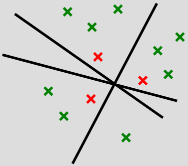


"Support Vectors" for Linear Arrangements

Corollary

1. *Every minimizer* Y^* *for* $\max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y^* \circ A \mid \mathscr{A})$ *is centered on* $K(A)$, *i.e.* $y^*_{i,j} = 0$ *for all* $(i,j) \notin K(A)$
2. $\max\limits_{\mathscr{A}} \min\limits_{i,j} \gamma_{i,j}(A \mid \mathscr{A}) = \max\limits_{\mathscr{A}} \min\limits_{(i,j) \in K(A)} \gamma_{i,j}(A \mid \mathscr{A})$.





$$A \rightsquigarrow \begin{pmatrix} 0 & 0 & +1 & 0 \\ -1 & -1 & +1 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Connection to CSQdim

Correlational Statistical Query Model

Definition

$h_1, \ldots, h_d \in [-1,1]^m$ are universally correlated with $A \in \mathbb{R}^{m \times n}$ iff

$$\forall \text{ probability vectors } p \; \forall j \; \exists k : \; \left| \langle h_k, A_j \rangle_p \right| \geqslant 1/d \; .$$

$\mathrm{CSQdim}(A) := \min\{d : \exists \, d$ vectors that are univ. corr. with $A\}$

Lemma

For every $A \in \mathbb{R}^{m \times n}$:

1. $\mathrm{mc}(A) \leqslant \mathrm{CSQdim}(A)^{1.5}$
2. $\mathrm{CSQdim}(A) \leqslant \left\lceil 32 \ln(4mn) \cdot \mathrm{mc}(A)^2 \right\rceil$

Connection to CSQdim

Correlational Statistical Query Model

Definition

$h_1, \ldots, h_d \in [-1,1]^m$ are universally correlated with $A \in \mathbb{R}^{m \times n}$ iff

$$\forall \text{ probability vectors } p \ \forall j \ \exists k : \left| \langle h_k, A_j \rangle_p \right| \geqslant 1/d \ .$$

$\mathrm{CSQdim}(A) := \min\{d : \exists\, d$ vectors that are univ. corr. with $A\}$

Lemma

For every $A \in \mathbb{R}^{m \times n}$:

1. $\mathrm{mc}(A) \leqslant \mathrm{CSQdim}(A)^{1.5}$
2. $\mathrm{CSQdim}(A) \leqslant \left\lceil 32 \ln(4mn) \cdot \mathrm{mc}(A)^2 \right\rceil$

Connection to CSQdim

Correlational Statistical Query Model

Definition

$h_1, \ldots, h_d \in [-1,1]^m$ are universally correlated with $A \in \mathbb{R}^{m \times n}$ iff

$$\forall \text{ probability vectors } p \; \forall j \; \exists k : \left| \langle h_k, A_j \rangle_p \right| \geqslant 1/d \; .$$

$\mathrm{CSQdim}(A) := \min\{d : \exists \; d$ vectors that are univ. corr. with $A\}$

Lemma

For every $A \in \mathbb{R}^{m \times n}$:

1. $\mathrm{mc}(A) \leqslant \mathrm{CSQdim}(A)^{1.5}$
2. $\mathrm{CSQdim}(A) \leqslant \left\lceil 32 \ln(4mn) \cdot \mathrm{mc}(A)^2 \right\rceil$

Connection to CSQdim

Lemma

$\mathrm{mc}(A) \leqslant \mathrm{CSQdim}(A)^{1.5}$

Proof sketch

1. *Let* $d := \mathrm{CSQdim}(A)$ *and* $h_1, \ldots, h_d$ *be universally correlated with* A.
2. *Let* $Y = (y_{i,j})$ *s.t.* $\mathrm{mc}(A) = \overline{\mathrm{mc}}_Y(A)$
3. *Then:* $\forall j\ \exists k(j) : \left| \left\langle h_{k(j)}, A_j \right\rangle_{p_j} \right| \geqslant 1/d$
4. *Arrangement:* $u_i := \frac{1}{\sqrt{d}} \cdot (h_{i,1}, \ldots, h_{i,d}), \quad v_j := \mathrm{sign}(\langle h_{k(j)}, A_j \rangle) \cdot e_{k(j)}$
5. *Calculation:* $\sum_{i,j} y_{i,j} \langle u_i, v_j \rangle = \ldots \geqslant \frac{1}{d^{1.5}}$

Connection to CSQdim

Lemma

$\mathrm{CSQdim}(A) \leqslant \left\lceil 32 \ln(4mn) \cdot \mathrm{mc}(A)^2 \right\rceil$

Proof sketch

1. *Let* $\mathscr{A}$ *be an optimal arrangement i.e.* $\gamma := \min_{i,j} \gamma_{i,j}(A \mid \mathscr{A}) = {}^1\!/_{\mathrm{mc}(A)}$
2. *Via random projections: Arrangement* $\mathscr{A}' = (u_1, \ldots, u_m; v_1, \ldots, v_n)$ *that is* d*-dimensional for* $d := \left\lceil 32 \ln(4mn) \cdot \mathrm{mc}(A)^2 \right\rceil$ *and satisfies* $\min_{i,j} \gamma_{i,j}(A \mid \mathscr{A}') \geqslant \gamma/2$
3. *Collection of* d *universally correlated vectors:* $h_k := (u_{1,k}, \ldots, u_{m,k})$

Connection to SQdim

Statistical Query Model

Definition

$$\mathrm{SQdim}_p(A) := \max\{d \mid \exists A_{i_1},\ldots,A_{i_d} \forall s,t : |\langle A_{i_s}, A_{i_t}\rangle_p| \leqslant 1/d\}$$

$$\mathrm{SQdim}(A) := \max_p \mathrm{SQdim}_p(A)$$

Lemma

For every $A \in \{-1,1\}^{m\times n}$ and every probability vector p:

1. $\max_q \overline{\mathrm{mc}}_{pq^\top}(A) < \sqrt{\mathrm{SQdim}_p(A)} \cdot (\mathrm{SQdim}_p(A)+1)$
2. $\mathrm{SQdim}_p(A) < 2 \cdot \max_q \overline{\mathrm{mc}}_{pq^\top}(A)$

Connection to SQdim

Statistical Query Model

Definition

$$\mathrm{SQdim}_p(A) := \max\{d \mid \exists A_{i_1}, \ldots, A_{i_d} \forall s,t : \left|\langle A_{i_s}, A_{i_t}\rangle_p\right| \leqslant {}^1/_d\}$$

$$\mathrm{SQdim}(A) := \max_p \mathrm{SQdim}_p(A)$$

Lemma

For every $A \in \{-1,1\}^{m\times n}$ and every probability vector p:

1. $\max_q \overline{\mathrm{mc}}_{pq^\top}(A) < \sqrt{\mathrm{SQdim}_p(A) \cdot (\mathrm{SQdim}_p(A)+1)}$
2. $\mathrm{SQdim}_p(A) < 2 \cdot \max_q \overline{\mathrm{mc}}_{pq^\top}(A)$

Connection to SQdim

Statistical Query Model

Definition

$$\mathrm{SQdim}_p(A) := \max\{d \mid \exists A_{i_1},\ldots,A_{i_d} \forall s,t : \left|\langle A_{i_s}, A_{i_t}\rangle_p\right| \leqslant 1/d\}$$

$$\mathrm{SQdim}(A) := \max_p \mathrm{SQdim}_p(A)$$

Lemma

For every $A \in \{-1,1\}^{m\times n}$ and every probability vector p:

1. $\max_q \overline{\mathrm{mc}}_{pq^\top}(A) < \sqrt{\mathrm{SQdim}_p(A) \cdot (\mathrm{SQdim}_p(A)+1)}$
2. $\mathrm{SQdim}_p(A) < 2 \cdot \max_q \overline{\mathrm{mc}}_{pq^\top}(A)$

Connection to SQdim

Conclusion

$$\mathrm{SQdim}(A) < 2 \cdot \max_{p,q} \overline{\mathrm{mc}}_{pq^\top}(A)^2 < 2 \cdot \mathrm{SQdim}(A) \cdot (\mathrm{SQdim}(A)+1)^2$$

Previous result from Sherstov[2008]:

$$\mathrm{SQdim}(A^\top) < 32 \cdot \mathrm{SQdim}(A)^4$$

Corollary

$$\mathrm{SQdim}(A^\top) < 2 \cdot \mathrm{SQdim}(A) \cdot (\mathrm{SQdim}(A)+1)^2$$

Connection to SQdim

Conclusion

$$\mathrm{SQdim}(A) < 2 \cdot \max_{p,q} \overline{\mathrm{mc}}_{pq^\top}(A)^2 < 2 \cdot \mathrm{SQdim}(A) \cdot (\mathrm{SQdim}(A) + 1)^2$$

Previous result from Sherstov[2008]:

$$\mathrm{SQdim}(A^\top) < 32 \cdot \mathrm{SQdim}(A)^4$$

Corollary

$$\mathrm{SQdim}(A^\top) < 2 \cdot \mathrm{SQdim}(A) \cdot (\mathrm{SQdim}(A) + 1)^2$$

Connection to SQdim

Conclusion

$\mathrm{SQdim}(A) < 2 \cdot \max_{p,q} \overline{\mathrm{mc}}_{pq^\top}(A)^2 < 2 \cdot \mathrm{SQdim}(A) \cdot (\mathrm{SQdim}(A)+1)^2$

Previous result from Sherstov[2008]:

$$\mathrm{SQdim}(A^\top) < 32 \cdot \mathrm{SQdim}(A)^4$$

Corollary

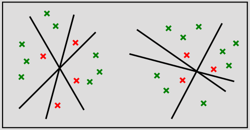
$\mathrm{SQdim}(A^\top) < 2 \cdot \mathrm{SQdim}(A) \cdot (\mathrm{SQdim}(A)+1)^2$

Summing up

- We have min-max theorems for margin complexity measures

$$\max_{\mathscr{A}} \min_{i,j} \gamma_{i,j}(A|\mathscr{A}) = \min_{Y} \max_{\mathscr{A}} \sum_{i,j} \gamma_{i,j}(Y \circ A|\mathscr{A})$$

- We have "support vectors" for linear arrangements



- We have connections to statistical learning models

Summing up
The nature of Y is crucial

Dual setting: Maximize the total margin of matrix $Y \circ A$.

$Y = pq^\top$ q under control of adversary	$\leftrightarrow$	$\max_q \overline{\mathrm{mc}}_{pq^\top}(A)$	$\leftrightarrow$	SQ model weak learning fixed distribution
$Y = pq^\top$ p,q under control of adversary	$\leftrightarrow$	$\max_{p,q} \overline{\mathrm{mc}}_{pq^\top}(A)$	$\leftrightarrow$	SQ model weak learning hardest fixed distribution
Y arbitrary	$\leftrightarrow$	$\max_Y \overline{\mathrm{mc}}_Y(A)$	$\leftrightarrow$	CSQ model weak learning distribution independet

Summing up

The nature of Y is crucial

Dual setting: Maximize the total margin of matrix $Y \circ A$.

$Y = pq^\top$ q under control of adversary	$\leftrightarrow$	$\max_q \overline{\mathrm{mc}}_{pq^\top}(A)$	$\leftrightarrow$	SQ model weak learning fixed distribution
$Y = pq^\top$ p, q under control of adversary	$\leftrightarrow$	$\max_{p,q} \overline{\mathrm{mc}}_{pq^\top}(A)$	$\leftrightarrow$	SQ model weak learning hardest fixed distribution
Y arbitrary	$\leftrightarrow$	$\max_Y \overline{\mathrm{mc}}_Y(A)$	$\leftrightarrow$	CSQ model weak learning distribution independet

Open problems

- Is there a parameter characterizing distribution independent SQ-learning?
- Can distribution independent SQ-learning be separated from SQ-learning w.r.t. the hardest fixed distribution?
- What about other soft-margin optimization problems?

Many thanks for your most gracious attention.