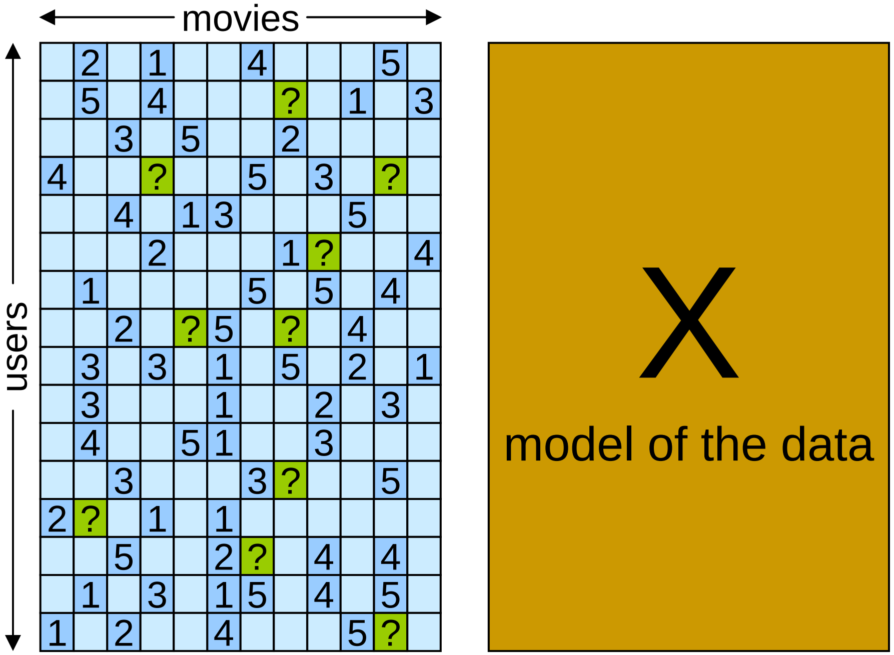
Rank, Trace-Norm & Max-Norm
as measures of matrix complexity

Nati Srebro
University of Toronto

Adi Shraibman
Hebrew University

Matrix Learning



- Reconstructing latent signal
  - gene expression (biological processes)
- Capturing structure in a corpus
  - documents, images, etc (topics, etc)
- Prediction: collaborative filtering
  - movie ratings

Fit (partially) observed **Y** with **X** from **hypothesis class of matrices**

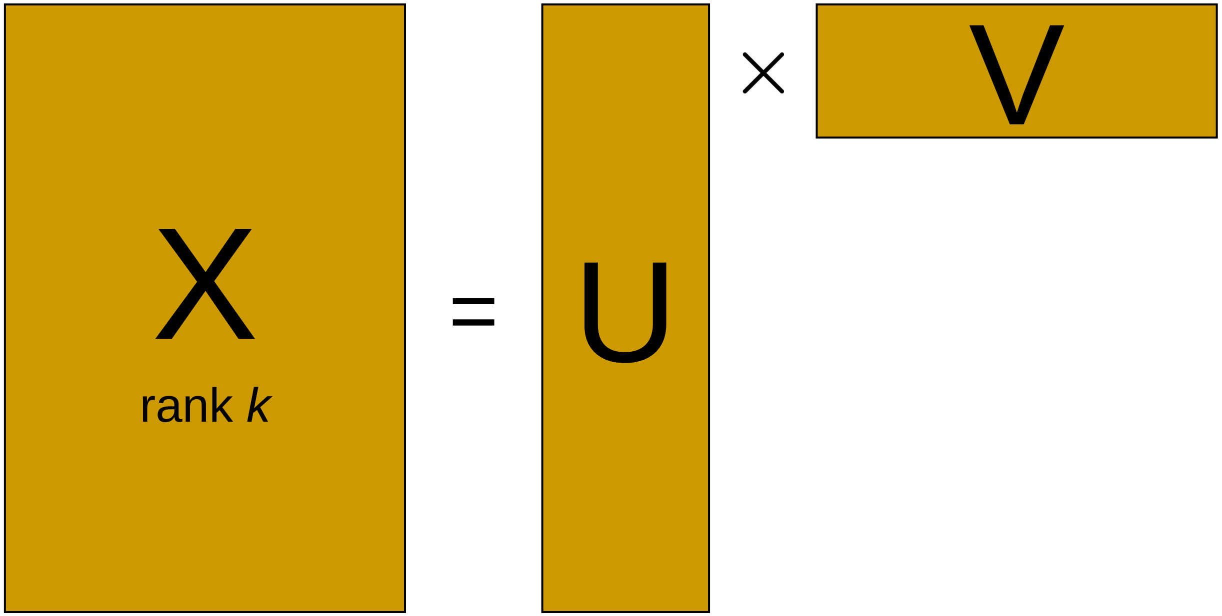
Low Rank:	$\{ X \mid \text{rank}(X) \leq k \}$	*Low dimensional factorization*
Low Trace-Norm:	$\{ X \mid \|X\|_{tr} \leq B \}$	*Low norm factorization* — **MMMF [NIPS 04]**
Low Max-Norm:	$\{ X \mid \|X\|_{max} \leq B \}$	

In this talk:

- The three hypothesis classes (measures of matrix complexity)
- Generalization error bounds (for predicting unobserved entries)
- Relationships between the three measures / hypothesis classes

Learning with Low Rank Matrices

	-1		-1			+1				+1	
	+1		+1						-1		-1
		-1		+1			+1				
+1						+1		-1			
		+1		-1	-1				+1		
			-1				-1				+1
	-1					+1		+1		+1	
		-1			+1				+1		
	+1		+1		-1		+1		-1		-1
	+1				-1			-1		+1	
	+1			+1	-1			+1			
		-1				-1				+1	
-1			-1		-1						
		+1			-1			+1		+1	
	-1		-1		-1	+1		+1		+1	
-1		-1			+1				+1		



Low Rank: $\{ X \mid \text{rank}(X) \leq k \}$

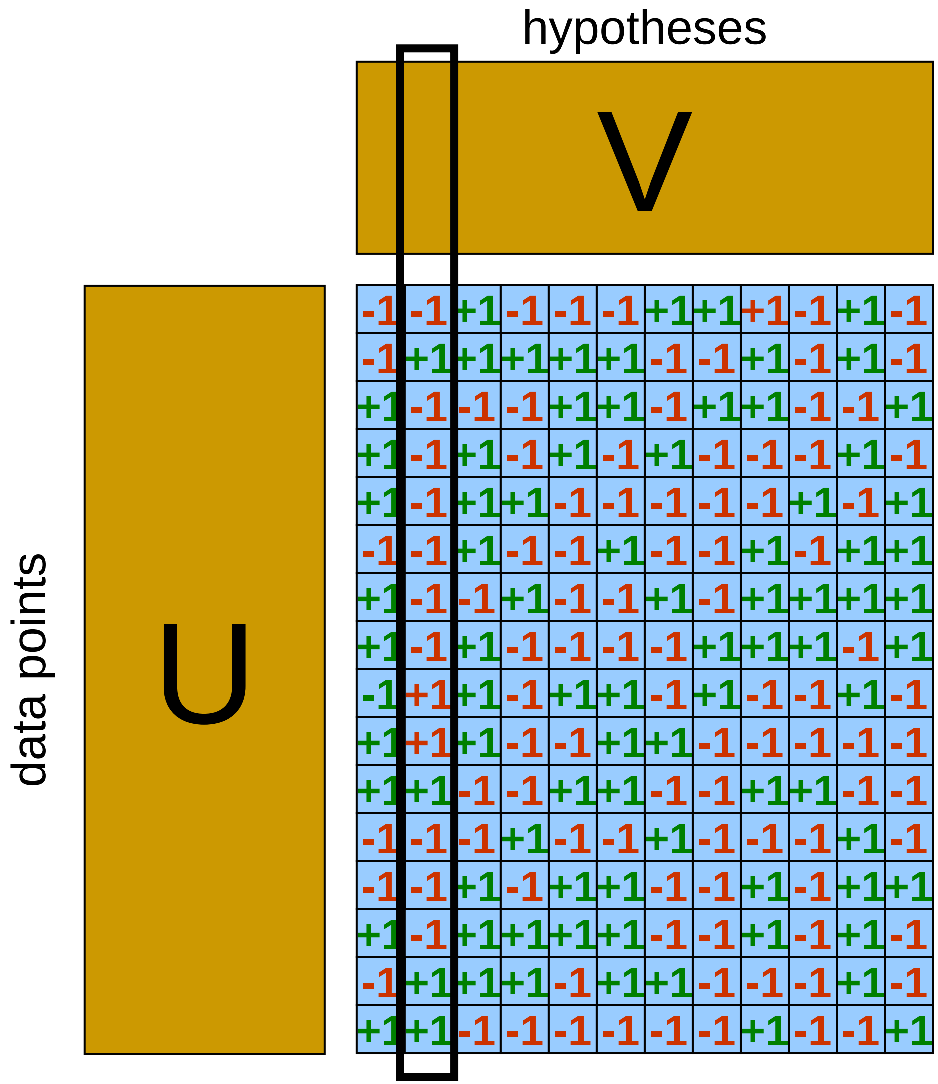
$= \{ UV' \mid U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{m \times k} \}$

For binary target matrices, only care about A=sign(X)

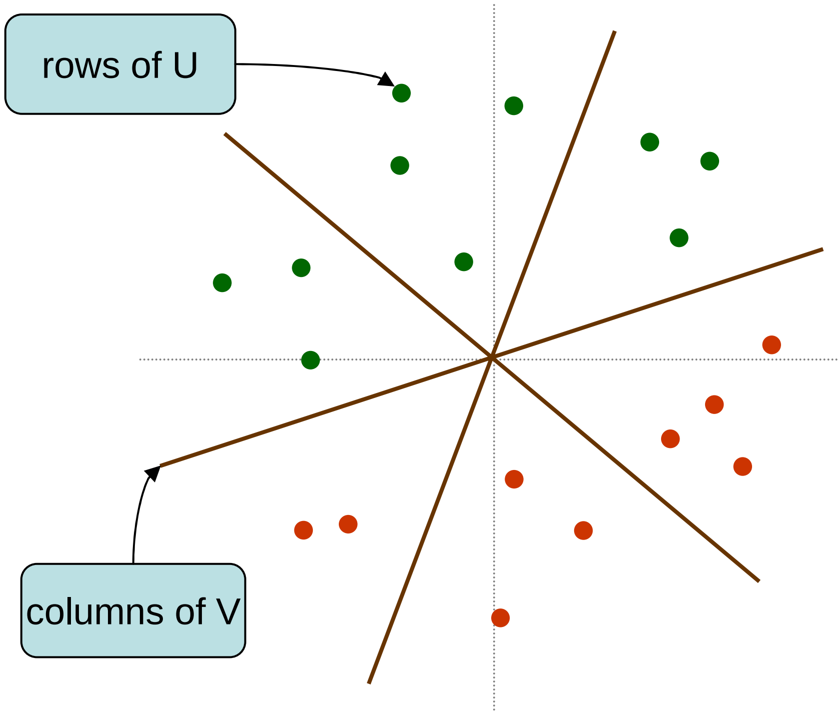
Dimensional Complexity of a binary matrix A:

$$dc(A) = \min_{\text{sign}(X)=A} \text{rank}(X)$$

Learning with Low Rank Matrices: Geometric Interpretation



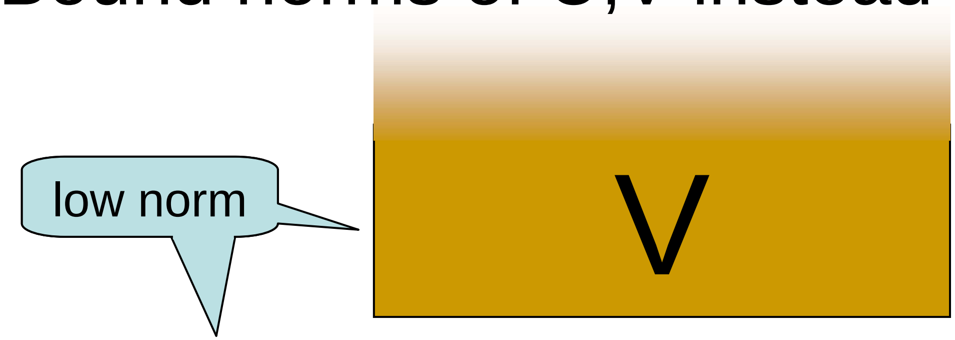
$$dc(A) = \min_{\text{rank}(X)=A} \text{rank}(X)$$

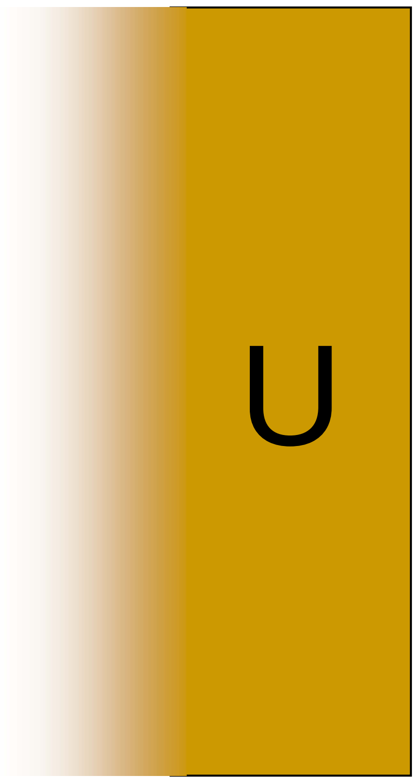


What is the minimum dimension into which a concept class can be embedded as linear separators?

Max-Margin Matrix Factorization:

Bound norms of U,V instead of their dimensionality

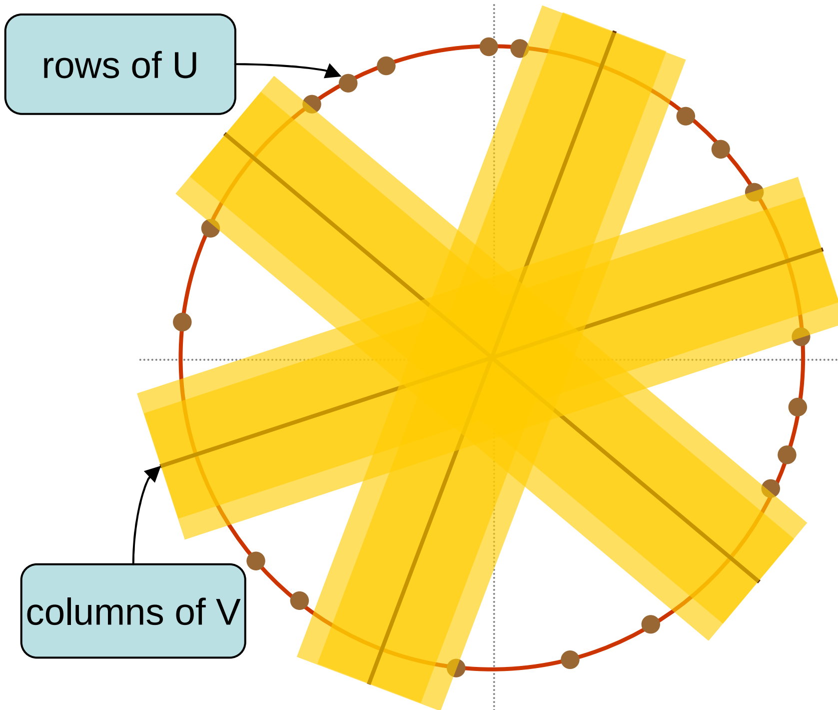




-1	-1	+1	-1	-1	-1	+1	+1	+1	-1	+1	-1
-1	+1	+1	+1	+1	+1	-1	-1	+1	-1	+1	-1
+1	-1	-1	-1	+1	+1	-1	+1	+1	-1	-1	+1
+1	-1	+1	-1	+1	-1	+1	-1	-1	-1	+1	-1
+1	-1	+1	+1	-1	-1	-1	-1	-1	+1	-1	+1
-1	-1	+1	-1	-1	+1	-1	-1	+1	-1	+1	+1
+1	-1	-1	+1	-1	-1	+1	-1	+1	+1	+1	+1
+1	-1	+1	-1	-1	-1	-1	+1	+1	+1	-1	+1
-1	+1	+1	-1	+1	+1	-1	+1	-1	-1	+1	-1
+1	+1	+1	-1	-1	+1	+1	-1	-1	-1	-1	-1
+1	+1	-1	-1	+1	+1	-1	-1	+1	+1	-1	-1
-1	-1	-1	+1	-1	-1	+1	-1	-1	-1	+1	-1
-1	-1	+1	-1	+1	+1	-1	-1	+1	-1	+1	+1
+1	-1	+1	+1	+1	+1	-1	-1	+1	-1	+1	-1
-1	+1	+1	+1	-1	+1	+1	-1	-1	-1	+1	-1
+1	+1	-1	-1	-1	-1	-1	-1	+1	-1	-1	+1

bound norms uniformly:

$(\max_i |U_i|^2)\ (\max_j |V_j|^2) \leq R^2$



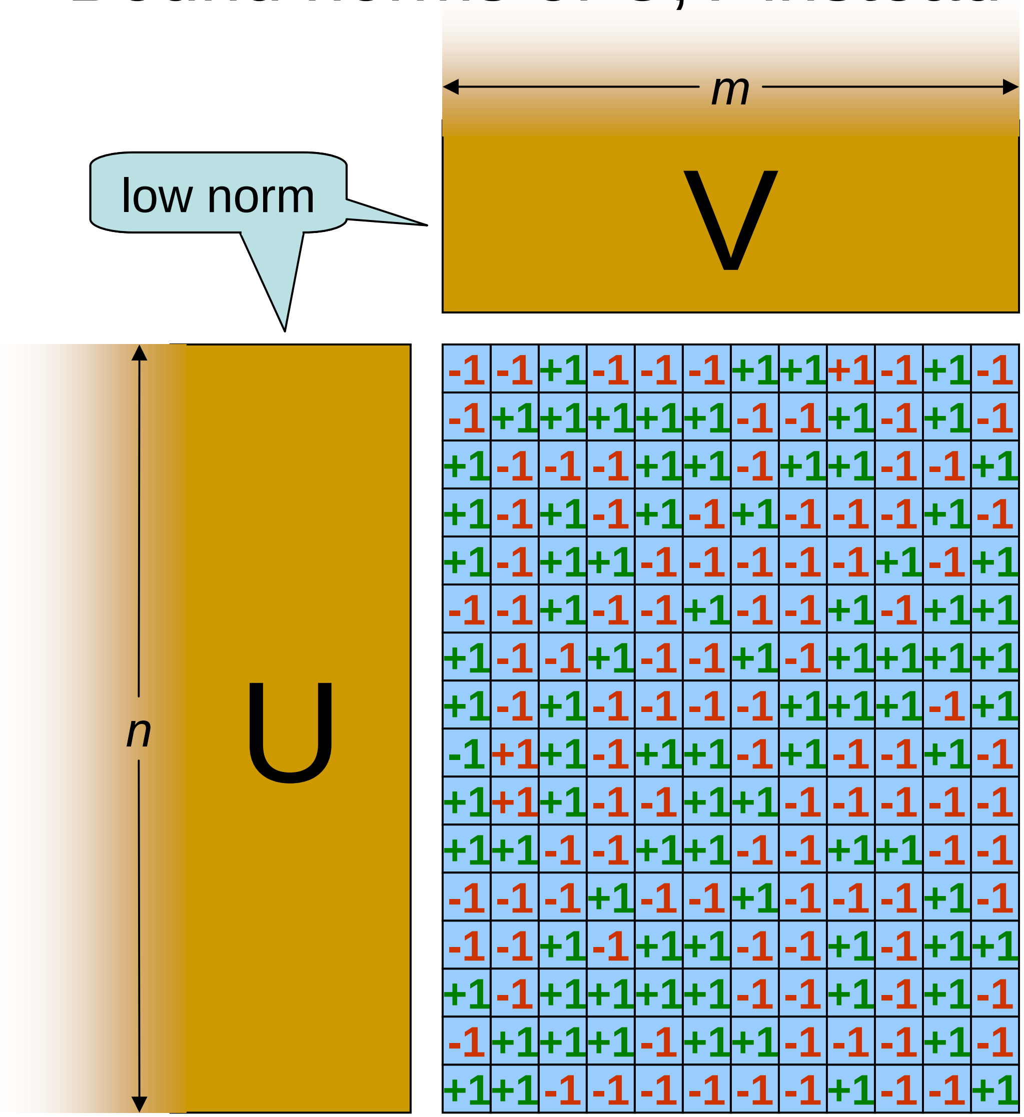
For each $Y_{ij} \in \pm 1$:

$$Y_{ij}\ X_{ij} \geq \mathit{Margin}$$

$\langle U_i, V_j \rangle$

Max-Margin Matrix Factorization:

Bound norms of U,V instead of their dimensionality



bound norms uniformly:

$$\underbrace{(\max_i |U_i|^2)\,(\max_j |V_j|^2)}_{|X|_{max} = \min_{X=UV} (\max_i |U_i|)(\max_j |V_j|)} \leq R^2$$

bound norms on avarage:

$$\underbrace{(\Sigma_i\, |U_i|^2)\,(\Sigma_j\, |V_j|^2)}_{|X|_{\Sigma} = \min_{X=UV} |U|_{Fro} |V|_{Fro}} \leq nmR^2$$

Optimize V given U:
each column is a SVM

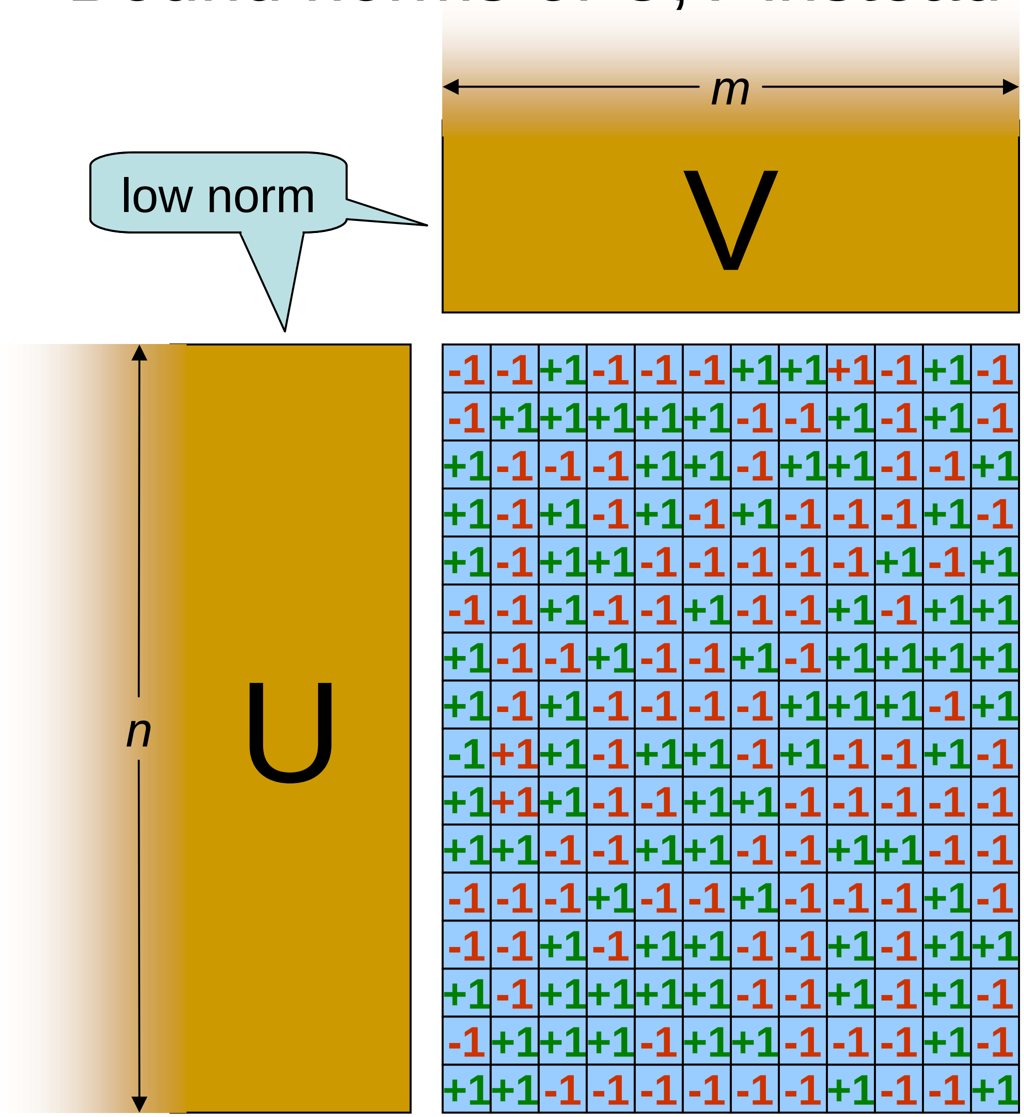
For each $\mathbf{Y_{ij}} \in \pm 1$:

$$\mathbf{Y_{ij}}\, \mathbf{X_{ij}} \geq 1$$

$\langle U_i, V_j \rangle$

Max-Margin Matrix Factorization:

Bound norms of U,V instead of their dimensionality



-1	-1	+1	-1	-1	-1	+1	+1	+1	-1	+1	-1
-1	+1	+1	+1	+1	+1	-1	-1	+1	-1	+1	-1
+1	-1	-1	-1	+1	+1	-1	+1	+1	-1	-1	+1
+1	-1	+1	-1	+1	-1	+1	-1	-1	-1	+1	-1
+1	-1	+1	+1	-1	-1	-1	-1	-1	+1	-1	+1
-1	-1	+1	-1	-1	+1	-1	-1	+1	-1	+1	+1
+1	-1	-1	+1	-1	-1	+1	-1	+1	+1	+1	+1
+1	-1	+1	-1	-1	-1	-1	+1	+1	+1	-1	+1
-1	+1	+1	-1	+1	+1	-1	+1	-1	-1	+1	-1
+1	+1	+1	-1	-1	+1	+1	-1	-1	-1	-1	-1
+1	+1	-1	-1	+1	+1	-1	-1	+1	+1	-1	-1
-1	-1	-1	+1	-1	-1	+1	-1	-1	-1	+1	-1
-1	-1	+1	-1	+1	+1	-1	-1	+1	-1	+1	+1
+1	-1	+1	+1	+1	+1	-1	-1	+1	-1	+1	-1
-1	+1	+1	+1	-1	+1	+1	-1	-1	-1	+1	-1
+1	+1	-1	-1	-1	-1	-1	-1	+1	-1	-1	+1

bound norms uniformly:

$$\underbrace{(\max_i |U_i|^2)\,(\max_j |V_j|^2)}\leq R^2$$

$$|X|_{max} = \min_{X=UV} (\max_i |U_i|)(\max_j |V_j|)$$

$$\mathbf{mc}(A) = \min_{A_{ij}X_{ij}\geq 1} |X|_{max}$$

bound norms on avarage:

$$\underbrace{(\textstyle\sum_i |U_i|^2)\,(\sum_j |V_j|^2)}\leq nmR^2$$

$$|X|_\Sigma = \min_{X=UV} |U|_{Fro}|V|_{Fro}$$

$$\mathbf{ac}(A) = \min_{A_{ij}X_{ij}\geq 1} |X|_\Sigma / \sqrt{nm}$$

For each $Y_{ij} \in \pm 1$:

$$Y_{ij}\,X_{ij} \geq 1$$

$\langle U_i, V_j \rangle$

Three Measures of Matrix Complexity

Used for fitting observed data matrices

For matrices representing concept classes:

bound norms uniformly:

$\underbrace{(\max_i |U_i|^2)\ (\max_j |V_j|^2)} \le R^2$

$|X|_{max} = \min_{X=UV} (\max_i |U_i|)(\max_j |V_j|)$

1/margin required for embedding as linear classifiers

$\mathbf{mc}(A) = \min_{A_{ij}X_{ij} \ge 1} |X|_{max}$

bound norms on avarage:

$\underbrace{(\Sigma_i |U_i|^2)\ (\Sigma_j |V_j|^2)} \le nmR^2$

$|X|_\Sigma = \min_{X=UV} |U|_{Fro} |V|_{Fro}$

$\mathbf{ac}(A) = \min_{A_{ij}X_{ij} \ge 1} |X|_\Sigma / \sqrt{nm}$

Convex!

(semi-definite programming)

[NIPS 04]

Not convex →

bound dimensionality of U,V:

dimension required for embedding as linear classifiers

$\mathbf{dc}(A) = \min_{A_{ij}X_{ij} > 0} \operatorname{rank}(X)$

Outline

- Three measures of matrix complexity
- Generalization error bounds
- Relationships between the three measures

Generalization Error Bounds

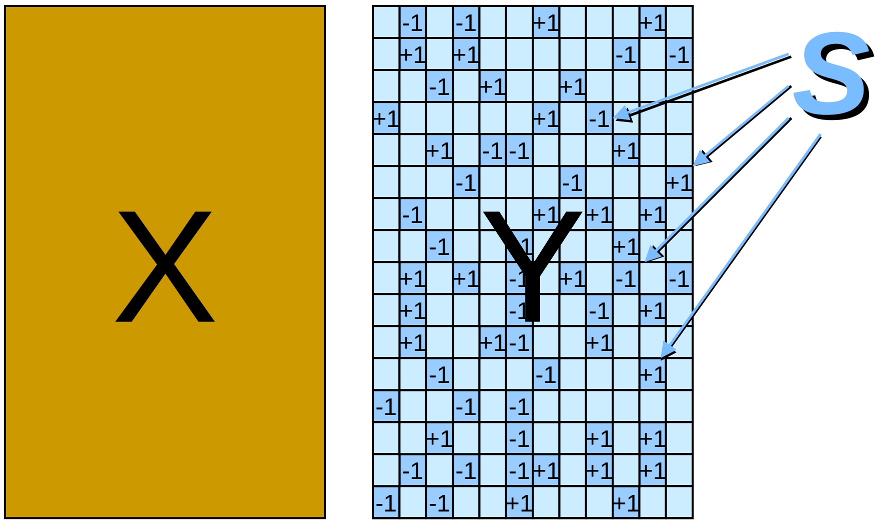
D(**X**;**Y**) = |{ ij | $X_{ij}Y_{ij} \leq 0$ }| / nm

generalization error

Assuming a low-rank structure (eigengap):

Asymptotic behavior **[Azar+01]**

Sample complexity, query strategy **[Drineas+02]**



Generalization Error Bounds

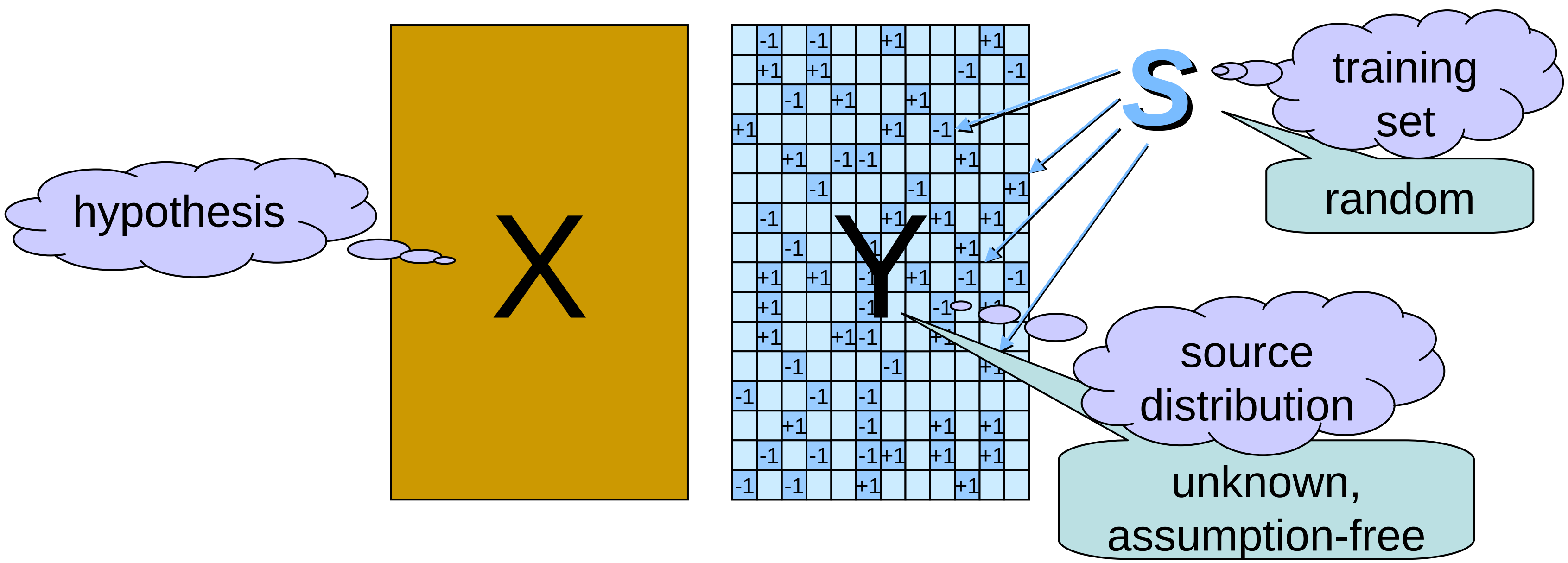
$D(\mathbf{X};\mathbf{Y}) = |\{ ij \mid X_{ij}Y_{ij} \leq 0 \}| / nm$

generalization error

$D_S(\mathbf{X};\mathbf{Y}) = |\{ ij \in S \mid X_{ij}Y_{ij} \leq 0 \}| / |S|$

empirical error

$$\forall_{\mathbf{Y}} \Pr_{S} \left(\forall_{\mathbf{X} \in \mathcal{H}}\ D(\mathbf{X};\mathbf{Y}) < D_S(\mathbf{X};\mathbf{Y}) + \varepsilon \right) > 1-\delta$$



Generalization Error Bounds: Learning with Low Rank Matrices

$D(\mathbf{X};\mathbf{Y}) = |\{ ij \mid X_{ij}Y_{ij} \leq 0 \}| / nm$

generalization error

$D_S(\mathbf{X};\mathbf{Y}) = |\{ ij \in S \mid X_{ij}Y_{ij} < 1 \}| / |S|$

empirical error

$$\forall_{\mathbf{Y}} \Pr_S \left(\forall_{\mathbf{X} \in \mathcal{H}} D(\mathbf{X};\mathbf{Y}) < D_S(\mathbf{X};\mathbf{Y}) + \varepsilon \right) > 1-\delta$$

$\mathcal{H} = \{ X \mid \text{rank}(X) \leq k \}$

- Only signs matter, equivalent to using the class:

$$\{ A \in \pm 1^{n \times m} \mid A = \text{sign}(X), \text{rank}(X) \leq k \} = \{ A \in \pm 1^{n \times m} \mid dc(A) \leq k \}$$

- Finite class of size:

$$|\{ A \in \pm 1^{n \times m} \mid dc(A) \leq k \}| \leq (8em/k)^{k(n+m)}$$

[BenDavid,Eiron,Simon01]: most concept classes have high dc(A), i.e. cannot be embedded as low-dim linear classifiers

- Union bound yields generalization bound with:

$$\varepsilon = \sqrt{\frac{k(n+m)\log\frac{8em}{k} + \log 1/\delta}{2|S|}}$$

[S,Alon,Jaakkola 04]

Generalization Error Bounds:
Learning with Low Trace-Norm Matrices

$D(\mathbf{X};\mathbf{Y}) = |\{ ij \mid X_{ij}Y_{ij} \leq 0 \}| / nm$

generalization error

$D_S(\mathbf{X};\mathbf{Y}) = |\{ ij \in S \mid X_{ij}Y_{ij} < 1 \}| / |S|$

empirical error

$$\forall_{\mathbf{Y}} \Pr_S \left(\forall_{\mathbf{X} \in \mathcal{H}} \; D(\mathbf{X};\mathbf{Y}) < D_S(\mathbf{X};\mathbf{Y}) + \varepsilon \right) > 1-\delta$$

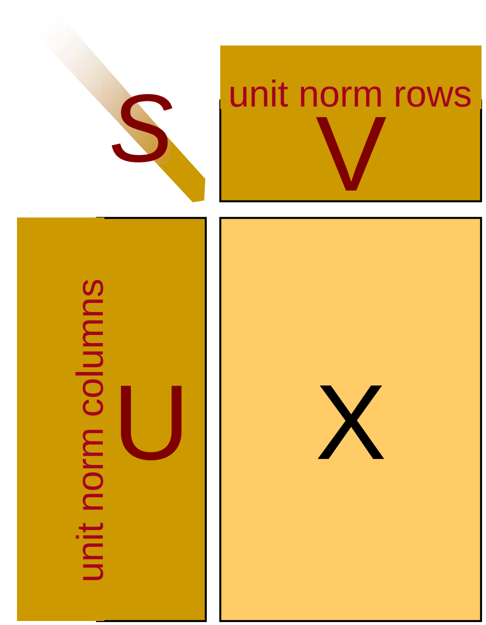
$$\mathcal{H} = \{ X \mid |X|^2_\Sigma < nmR^2 \} = \sqrt{nm}\,R \cdot \text{convex-hull}(\{ uv' \mid u \in \mathbb{R}^n, v \in \mathbb{R}^m \; |u|=|v|=1 \})$$

$$|\mathbf{X}|_\Sigma = \min_{X=UV} |U|_{Fro}|V|_{Fro}$$
$$= \sum(\text{singular values of } \mathbf{X})$$

$$\sum s_i \, u_i v'_i$$

$\sum s_i = |X|_\Sigma$

outer product of norm-1 vectors: rank-1 norm-1 matrix



Generalization Error Bounds:

Learning with Low Trace-Norm Matrices

$D(\mathbf{X};\mathbf{Y}) = |\{ ij \mid X_{ij}Y_{ij} \leq 0 \}| / nm$

generalization error

$D_{\mathbf{S}}(\mathbf{X};\mathbf{Y}) = |\{ ij \in S \mid X_{ij}Y_{ij} < 1 \}| / |S|$

empirical error

$$\forall_{\mathbf{Y}} \Pr_{\mathbf{S}} \left(\forall_{\mathbf{X}\in\mathcal{H}}\ D(\mathbf{X};\mathbf{Y}) < D_{\mathbf{S}}(\mathbf{X};\mathbf{Y}) + \varepsilon \right) > 1-\delta$$

$$\mathcal{H} = \{ X \mid |X|^2_{\Sigma} < nmR^2 \} = \sqrt{nm}\, R \cdot \text{convex-hull}(\{ uv' \mid u \in \mathbb{R}^n, v \in \mathbb{R}^m\ |u|=|v|=1 \})$$

Rademacher complexity of $\{ uv' \mid u \in R^n, v \in R^m\ |u|=|v|=1\}$

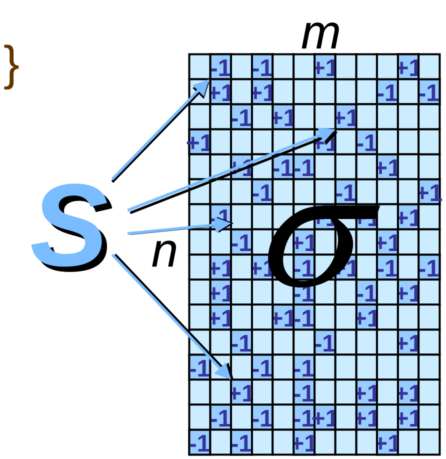
$\Downarrow$

Rademacher complexity of $\{ X \mid |X|^2_{\Sigma} < nmR^2 \}$

$\Downarrow$

Generalization error bounds for $\{ X \mid |X|^2_{\Sigma} < nmR^2 \}$

$$E_{\mathbf{S}}\, E_{\text{rand signs } \sigma} \left[\sup_{X=uv'} \left| \sum_{s \in \mathbf{S}} \sigma_s X(s) \right| \right] / |S|$$



Generalization Error Bounds:
Learning with Low Trace-Norm Matrices

$D(\mathbf{X};\mathbf{Y}) = |\{ ij \mid X_{ij}Y_{ij} \leq 0 \}| / nm$

generalization error

$D_S(\mathbf{X};\mathbf{Y}) = |\{ ij \in S \mid X_{ij}Y_{ij} < 1 \}| / |S|$

empirical error

$$\forall_{\mathbf{Y}} \Pr_S \left(\forall_{\mathbf{X} \in \mathcal{H}} \; D(\mathbf{X};\mathbf{Y}) < D_S(\mathbf{X};\mathbf{Y}) + \varepsilon \right) > 1 - \delta$$

$\mathcal{H} = \{ X \mid |X|^2_\Sigma < nmR^2 \} = \sqrt{nm}\,R \cdot \text{convex-hull}(\{ uv' \mid u \in \mathbb{R}^n, v \in \mathbb{R}^m \; |u|=|v|=1 \})$

Rademacher complexity of $\{ uv' \mid u \in R^n, v \in R^m \; |u|=|v|=1\}$

⇓

Rademacher complexity of $\{ X \mid |X|^2_\Sigma < nmR^2 \}$

⇓

Generalization error bounds for $\{ X \mid |X|^2_\Sigma < nmR^2 \}$

$$E_S \, E_{\text{rand signs } \sigma} \left[\sup_{X=uv'} \left| \textstyle\sum_{ij \in S} \sigma_{ij} X_{ij} \right| \right] / |S|$$

$$= \mathbf{E}[||\sigma||_2] / |S| \leq K \sqrt{(n+m)\log^{3/2} n \,/\, |S|} \quad \Rightarrow \quad \varepsilon = K \sqrt{\frac{R^2 (n+m) \log^{3/2} n + \log 1/\delta}{|S|}}$$

use **[Seginer00]** bound on singular values of random matrix

Generalization Error Bounds:

Learning with Low Max-Norm Matrices

$D(\mathbf{X};\mathbf{Y}) = |\{ ij \mid X_{ij}Y_{ij} \leq 0 \}| / nm$

generalization error

$D_S(\mathbf{X};\mathbf{Y}) = |\{ ij \in S \mid X_{ij}Y_{ij} < 1 \}| / |S|$

empirical error

$$\forall_{\mathbf{Y}} \Pr_S \left(\forall_{\mathbf{X} \in \mathcal{H}} D(\mathbf{X};\mathbf{Y}) < D_S(\mathbf{X};\mathbf{Y}) + \varepsilon \right) > 1-\delta$$

$\mathcal{H} = \{ X \mid |X|_{max} < R \} = R \cdot \text{convex-hull}($??? $)$

$$\text{conv}(\{ uv' \mid u \in \pm 1^n, v \in \pm 1^m \}) \subseteq$$

$$\{ X \mid |X|_{max} \leq 1 \}$$

$$\subseteq 1.79 \; \text{conv}(\{ uv' \mid u \in \pm 1^n, v \in \pm 1^m \})$$

Grothendiek's Inequality

$$\Rightarrow \quad \varepsilon = 12 \sqrt{\frac{R^2 (n+m) + \log 1/\delta}{|S|}}$$

Generalization Error Bounds:

Low Trace-Norm, Max-Norm or Rank

$D(\mathbf{X};\mathbf{Y}) = |\{ ij \mid X_{ij}Y_{ij} \leq 0 \}| / nm$

generalization error

$D_S(\mathbf{X};\mathbf{Y}) = |\{ ij \in S \mid X_{ij}Y_{ij} < 1 \}| / |S|$

empirical error

$$\forall_{\mathbf{Y}} \Pr_S \left(\forall_{\mathbf{X} \in \mathcal{H}} \; D(\mathbf{X};\mathbf{Y}) < D_S(\mathbf{X};\mathbf{Y}) + \varepsilon \right) > 1-\delta$$

$\mathcal{H} = \{ X \in R^{n\times m} \mid \mathrm{rank}(X) \leq k \}$

$\mathbf{dc}(A) = \min_{A_{ij}X_{ij}>0} \mathrm{rank}(X)$

$$\varepsilon = \sqrt{\frac{k(n+m)\log\frac{8en}{k} + \log \frac{1}{\delta}}{2|S|}}$$

$\mathcal{H} = \{ X \in R^{n\times m} \mid |X|^2_\Sigma/nm \leq R^2 \}$

$\mathbf{ac^2}(A) = \min_{A_{ij}X_{ij}\geq 1} |X|^2_\Sigma/nm$

$$\varepsilon = K\sqrt{\frac{R^2(n+m)\log^{3/2} n + \log \frac{1}{\delta}}{|S|}}$$

$\mathcal{H} = \{ X \in R^{n\times m} \mid |X|^2_{max} \leq R^2 \}$

$\mathbf{mc^2}(A) = \min_{A_{ij}X_{ij}\geq 1} |X|^2_{max}$

$$\varepsilon = 12\sqrt{\frac{R^2(n+m) + \log \frac{1}{\delta}}{|S|}}$$

Relationship between average margin complexity, max margin complexity and dimensional complexity

average $\leq$ maximum

$$ac^2(A) \leq mc^2(A)$$

$$dc(A) \leq 10\, mc^2(A) \log(3nm)$$

Randomly project high-dimensional large margin arrangement to obtain low dimensional arrangement **[Arriaga,Vempala99]**

Reverse inequalities?

Gaps?

Between the Rank and the Max-Norm

- Low margin complexity $\Rightarrow$ Low dimensional complexity:

$$\mathbf{dc(A)} \le 10\ \mathbf{mc^2(A)}\ \log(3nm)$$

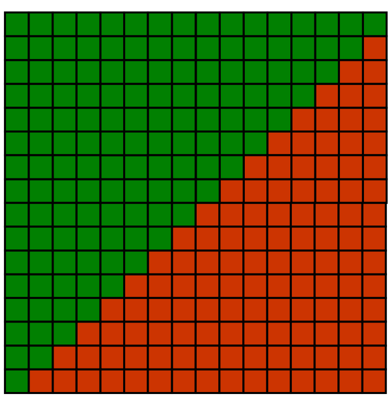
$$\tfrac{1}{2}\ \mathbf{(k-1)n}\ \log(m) \le \log |\{\ A \in \pm 1^{n\times m}\ |\ \mathbf{dc(A) \le k}\ \}| \le \mathbf{k(n+m)}\ \log(8en/k)$$

$$\mathbf{R^2 n}\ \log(n/R) \le \log |\{\ A \in \pm 1^{n\times m}\ |\ \mathbf{mc(A) \le R}\ \}| \le 10\ \mathbf{R^2(n+m)}\ \log(3nm)\ \log(n/R)$$

- Low dimensional complexity $\overset{?}{\not\Rightarrow}$ Low margin complexity:

$$\exists A, \qquad (\ \mathbf{dc(A)}\ \log(n)\ )^p \le \mathbf{mc^2(A)}$$

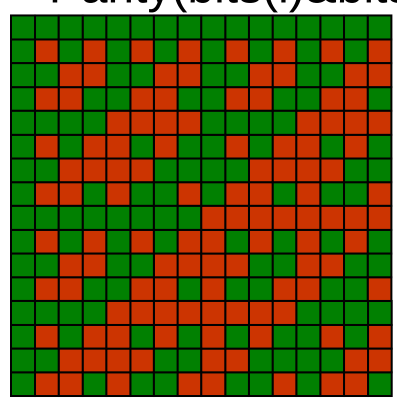
A_{ij} = Parity(bits(i)&bits(j))



$mc^2(T_n)=\Theta(\log n)$

$dc(T_n)=2$

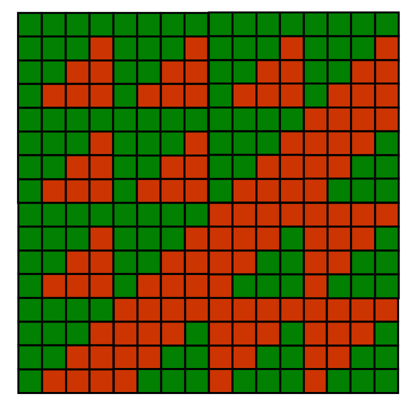
[BenDavid Eiron Simon 01]



$mc^2(H_n)=n$

$n^{0.5} \le dc(H_n) < n^{0.8}$

[Forster et al 02,03]



Kronecker exponent of triangular matrices

Between the Rank and the Max-Norm

- Low margin complexity $\Rightarrow$ Low dimensional complexity:

$$\mathbf{dc(A)} \leq 10\ \mathbf{mc^2(A)} \log(3nm)$$

$$\tfrac{1}{2}\ \mathbf{(k-1)n} \log(m) \leq \log |\{ A\in\pm1^{n\times m} \mid \mathbf{dc(A)}\leq\mathbf{k} \}| \leq \mathbf{k(n+m)} \log(8en/k)$$

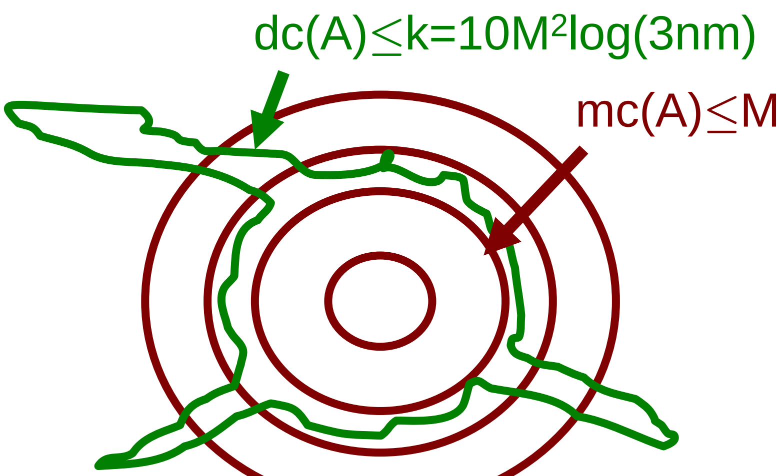
$$\mathbf{R^2n} \log(n/R) \leq \log |\{ A\in\pm1^{n\times m} \mid \mathbf{mc(A)}\leq\mathbf{R} \}| \leq 10\ \mathbf{R^2(n+m)} \log(3nm) \log(n/R)$$

- Low dimensional complexity $\not\Rightarrow$ Low margin complexity:

$$\exists\, A, \qquad (\ \mathbf{dc(A)} \log(n)\ )^p \leq \mathbf{mc^2(A)}$$

- Low max-norm $\Rightarrow$ similar matrix with low rank:

$$\mathbf{rank(X')} \leq 9/\varepsilon^2\ \mathbf{|X|^2_{max}} \log(3nm) \text{ for some } \mathbf{|X'-X|_\infty} < \varepsilon$$



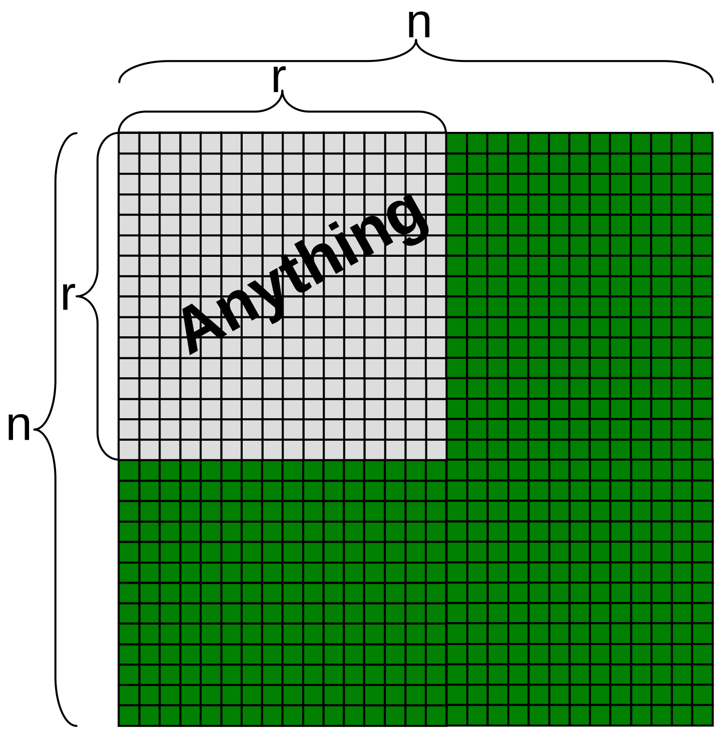
Open:

low $\mathbf{dc(A)} \overset{?}{\Rightarrow}$ low $\mathbf{mc^2(A')}$ for some $\mathbf{A'}\approx\mathbf{A}$

low $\mathbf{rank(X)} \overset{?}{\Rightarrow}$ low $\mathbf{|X'|^2_{max}}$ for some low $\mathbf{|X'-X|_1}$

Between the Trace-Norm and Max-Norm

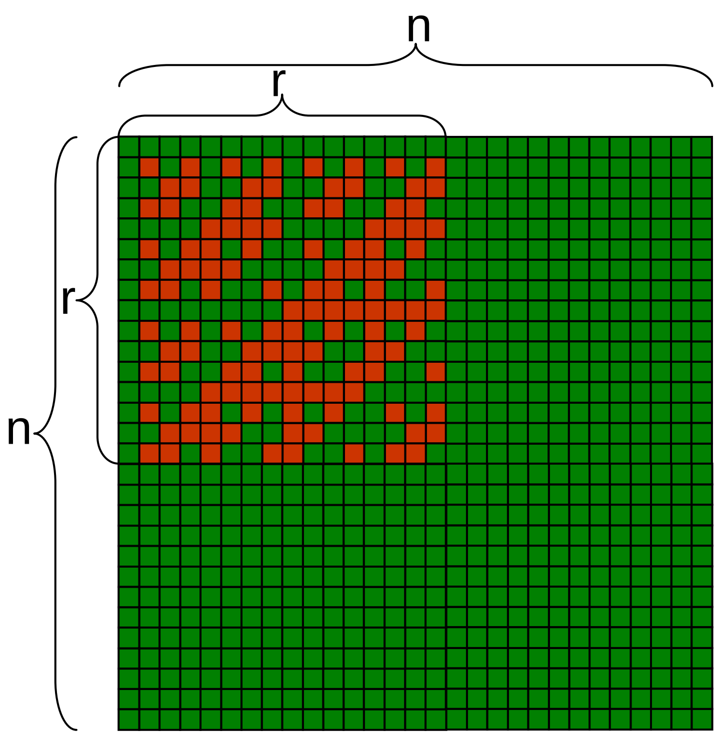
$$\mathbf{ac^2(A) \leq mc^2(A)}$$



$ac^2(A) < 2r^3/n^2 + 2$

Between the Trace-Norm and Max-Norm

$\mathbf{ac^2(A)} \leq \mathbf{mc^2(A)}$



$ac^2(A) < 2r^3/n^2 + 2$

For $r=n^{2/3}$, we can get:

$\mathbf{ac^2(A)} < 4$

while: $\mathbf{mc^2(A)} > n^{2/3}$

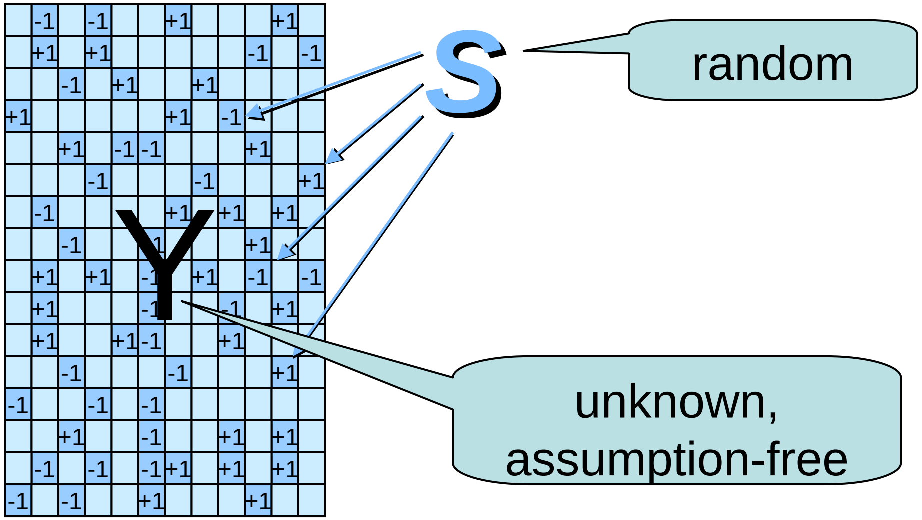
and: $dc^2(A) > n^{1/3}$

Random Sampling Assumption for Generalization Error Bounds

$D(\mathbf{X};\mathbf{Y}) = P_{ij}(X_{ij}Y_{ij} \leq 0)$

$\forall_{\mathbf{Y}} \Pr_{S} \left(\forall_{\mathbf{X}\in\mathcal{H}} D(\mathbf{X};\mathbf{Y}) < D_S(\mathbf{X};\mathbf{Y}) + \varepsilon \right) > 1-\delta$

$D_S(\mathbf{X};\mathbf{Y}) = |\{ ij\in S \mid X_{ij}Y_{ij} < 1 \}| \,/\, |S|$



$\mathcal{H} = \{ X\in R^{n\times m} \mid |X|^2_{\Sigma}/nm \leq R^2 \}$ $\quad \varepsilon = K\sqrt{\frac{R^2(n+m)\log^{3/2} n + \log 1/\delta}{|S|}}$

Requires **uniform sampling** of entries

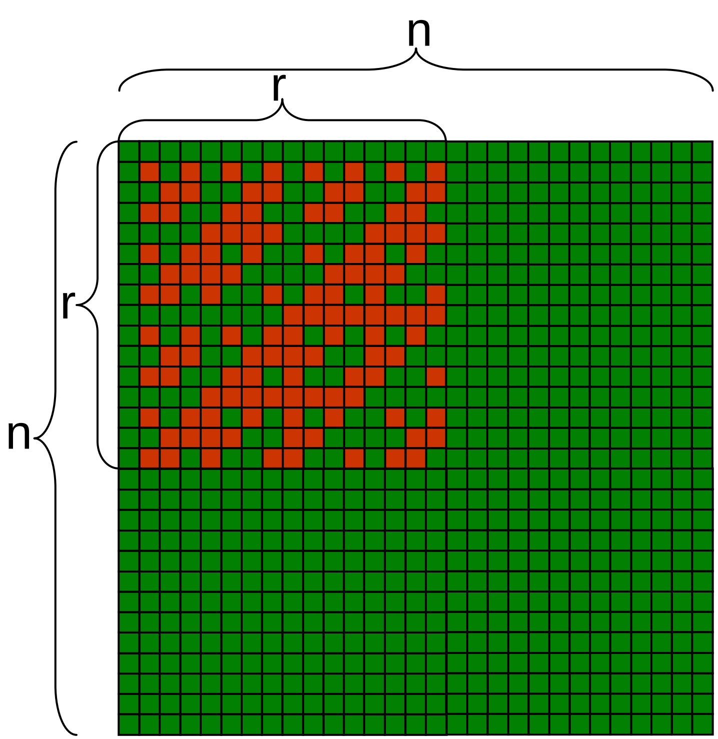
$\mathcal{H} = \{ X\in R^{n\times m} \mid |X|^2_{max} \leq R^2 \}$ $\quad \varepsilon = 12\sqrt{\frac{R^2(n+m) + \log 1/\delta}{|S|}}$

$\mathcal{H} = \{ X\in R^{n\times m} \mid \mathrm{rank}(X) \leq k \}$ $\quad \varepsilon = \sqrt{\frac{k(n+m)\log \frac{8en}{k} + \log 1/\delta}{2|S|}}$

Applies to **any distribution** over index pairs ij (which entries are observed)

Between the Trace-Norm and Max-Norm

$$\mathbf{ac^2(A)} \leq \mathbf{mc^2(A)}$$



$$ac^2(A) < 2r^3/n^2 + 2$$

For $r=n^{2/3}$, we can get:

$$\mathbf{ac^2(A)} < 4$$

while: $\mathbf{mc^2(A)} > n^{2/3}$

and: $dc^2(A) > n^{1/3}$

This is the largest gap possible: $\mathbf{mc^2(A)} \leq 9(\mathbf{ac^2(A)} \cdot n \cdot n)^{1/3}$

Using similar techniques:

$$(\mathbf{R}^{4/3}-2)\mathbf{n}^{4/3} \leq \log \left|\{ A \in \pm 1^{n \times m} \mid \mathbf{ac(A)} \leq \mathbf{R} \}\right| \leq 7\, \mathbf{R}^{2/3}\mathbf{n}^{5/3} \log(3nm) \log(n/M^2)$$

Rank, Max-Norm and Trace-Norm as complexity measures for fitting Data Matrix

$\text{rank}(X) = \min_{X=UV} \dim(U,V)$ $\quad |X|_{max} = \min_{X=UV} (\max_i |U_i|)(\max_j |V_j|)$ $\quad |X|_\Sigma = \min_{X=UV} |U|_{Fro}|V|_{Fro}$

$\mathbf{dc(A)} = \min_{A_{ij}X_{ij}>0} \text{rank}(X)$ $\quad \mathbf{mc(A)} = \min_{A_{ij}X_{ij}\geq 1} |X|_{max}$ $\quad \mathbf{ac(A)} = \min_{A_{ij}X_{ij}\geq 1} |X|_\Sigma / \sqrt{nm}$

(**dc(A)**, mc(A) previously studied as embedability of a concept class)

- Generalization error bounds scale with:

  $\mathbf{k=rank(X)}$ $\quad \mathbf{R^2=|X|^2_{max}}$ $\quad \mathbf{R^2=|X|^2_\Sigma/nm}$

- Relationships:
  - $\mathbf{dc(A)} \leq 10\, \mathbf{mc^2(A)} \log(3nm)$, but $\exists A$, $(\, \mathbf{dc(A)} \log(n)\,)^p < \mathbf{ac^2(A)} \leq \mathbf{mc^2(A)}$
  - $\mathbf{ac^2(A)} \leq \mathbf{mc^2(A)} \leq 9(\mathbf{ac^2(A)} \cdot n \cdot n)^{1/3}$, and this is tight
- Open issues:
  - low **dc(A)** $\Rightarrow$ low $\mathbf{mc^2(A')}$ for some $\mathbf{A'}\approx\mathbf{A}$?

    low **rank(X)** $\Rightarrow$ low $\mathbf{|X'|^2_{max}}$ for some bounded $\mathbf{|X'-X|_1}$?
  - **dc(A)** = O($\mathbf{mc^2(A)}$) ?

    also tighten estimate of $|\{\, A\in\pm 1^{n\times m} \mid \mathbf{mc(A)\leq R} \,\}|$; Median **mc(A)** ?
  - Tighten polynomial gap in bounds on $\log(|\{\, A\in\pm 1^{n\times m} \mid \mathbf{ac(A)\leq R} \,\}|)$