MATH 7251: High-Dimensional Probability **Spring 2022**

Lecture 1

Name: Tyler LaBonte **Professor:** Konstantin Tikhomorov

1 Overview of the Course

1. Bernstein-type inequalities for linear combinations of scalar random variables
2. Matrix Bernstein-type inequalities
3. Concentration for martingales
4. Isoperimetric inequalities
5. Concentration in product spaces
6. ϵ-net arguments
7. Applications to sample covariance matrices, dimensionality reduction, and random matrices

2 Linear Algebra Review

2.1 Orthogonality

Suppose $\mathbf{A}$ is an $n \times m$ matrix. The span of the columns is called the column space; it is a linear subspace of $\mathbb{R}^n$. Similarly, the span of the rows is called the row space, and it is a linear subspace of $\mathbb{R}^m$. The null space is $\{\boldsymbol{x} \in \mathbb{R}^m : \mathbf{A}\boldsymbol{x} = 0\}$ and it is the orthogonal complement of the row space. The dimensionalities of the row space and the column space are equal and are called the rank.

Theorem 2.1: Matrix Orthogonality

The following are equivalent for a square matrix $\mathbf{A}$:

1. $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$.
2. $\mathbf{A}\mathbf{A}^\top = \mathbf{I}$.
3. The columns of $\mathbf{A}$ form an orthonormal basis of $\mathbb{R}^n$.
4. The rows of $\mathbf{A}$ form an orthonormal basis of $\mathbb{R}^n$.
5. $\mathbf{A}$ is distance-preserving: $\|\mathbf{A}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

The eigenvalues of a square matrix $\mathbf{A}$ are numbers λ such that $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. A nonzero vector $\boldsymbol{x} \in \mathbb{C}^n$ is an eigenvector of $\mathbf{A}$ if there exists $\lambda \in \mathbb{C}$ such that $\mathbf{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. Another way to write this is $(\mathbf{A} - \lambda\mathbf{I})\boldsymbol{x} = 0$, and this means that $\mathbf{A} - \lambda\mathbf{I}$ is non-invertible (it has determinant 0). Thus, λ must be an eigenvalue of $\mathbf{A}$. A square matrix $\mathbf{A}$ is diagonalizable if there exists matrices $\mathbf{M}, \mathbf{D}$ such that $\mathbf{M}$ is invertible, $\mathbf{D}$ is diagonal, and $\mathbf{A} = \mathbf{M}^{-1}\mathbf{D}\mathbf{M}$.

Theorem 2.2: Eigendecomposition of Diagonalizable Matrices

Suppose $\mathbf{A}$ is an $n \times n$ diagonalizable matrix. Then, $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ where $\mathbf{P}$ is an $n \times n$ matrix whose columns are eigenvectors of $\mathbf{A}$ and $\mathbf{D}$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{A}$.

2.2 Symmetric Matrices

An $n \times n$ matrix $\mathbf{A}$ is symmetric if $\mathbf{A}^\top = \mathbf{A}$. Let $\boldsymbol{x}, \boldsymbol{y}$ be eigenvectors of a symmetric matrix $\mathbf{A}$ with corresponding eigenvalues $\lambda_{\boldsymbol{x}}$ and $\lambda_{\boldsymbol{y}}$. Then,

$$\begin{aligned} \langle \mathbf{A}\boldsymbol{x}, \boldsymbol{y} \rangle &= \lambda_{\boldsymbol{x}} \langle \boldsymbol{x}, \boldsymbol{y} \rangle \qquad (1) \\ &= \left\langle \boldsymbol{x}, \mathbf{A}^\top \boldsymbol{y} \right\rangle = \langle \boldsymbol{x}, \mathbf{A}\boldsymbol{y} \rangle = \langle \boldsymbol{x}, \lambda_{\boldsymbol{y}}\boldsymbol{y} \rangle = \lambda_{\boldsymbol{y}}^\star \langle \boldsymbol{x}, \boldsymbol{y} \rangle . \qquad (2) \end{aligned}$$

Applied with $\boldsymbol{y} = \boldsymbol{x}$ we obtain $\lambda_{\boldsymbol{x}} \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \lambda_{\boldsymbol{x}}^\star \langle \boldsymbol{x}, \boldsymbol{x} \rangle$, so $\lambda_{\boldsymbol{x}}$ is real. Thus all eigenvalues of a symmetric matrix are real. When the eigenvalues are distinct, we have $\lambda_{\boldsymbol{x}} \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \lambda_{\boldsymbol{y}} \langle \boldsymbol{x}, \boldsymbol{y} \rangle$, which is only possible when $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal. Thus, eigenvectors corresponding to distinct eigenvalues of a symmetric matrix must be orthogonal.

Theorem 2.3: Eigendecomposition for a Symmetric Matrix

Suppose $\mathbf{A}$ is an $n \times n$ symmetric matrix. Then, $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$ where $\mathbf{U}$ is an orthogonal matrix whose columns are eigenvectors of $\mathbf{A}$ and $\mathbf{D}$ is a diagonal matrix of eigenvalues of $\mathbf{A}$. By orthogonality, $\mathbf{A}$ is representable as the sum of rank-one matrices: $\mathbf{A} = \sum_{i=1}^n \lambda_i(\mathbf{A})\boldsymbol{x}_i\boldsymbol{x}_i^\top$ where $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$ and $\boldsymbol{x}_i$ is the unit eigenvector corresponding to $\lambda_i(\mathbf{A})$. Note that all the $\boldsymbol{x}_i$ form an orthonormal basis of $\mathbb{R}^n$.

An important class of symmetric matrices are positive semi-definite matrices. Suppose $\mathbf{A}$ is a symmetric real $n \times n$ matrix. Then, $\mathbf{A}$ is positive semi-definite if $\langle \mathbf{A}\boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and positive definite if $\langle \mathbf{A}\boldsymbol{x}, \boldsymbol{x} \rangle > 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$. For symmetric matrices, this is equivalent to all eigenvalues of $\mathbf{A}$ being non-negative or positive, respectively.

For example, let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be a random vector in $\mathbb{R}^n$. Assume that variances of all components of $\boldsymbol{x}$ are well-defined and finite. That is,

$$\operatorname{Var}(x_i) = \mathbb{E}\left[(x_i - \mathbb{E}[x_i])^2\right] < \infty \text{ for all } x_i. \qquad (3)$$

Then, the covariance matrix of $\boldsymbol{x}$ is defined as

$$\boldsymbol{\Sigma} = \mathbb{E}\left[(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^\top\right]. \tag{4}$$

Claim 2.1: Covariance Matrices are PSD

$\boldsymbol{\Sigma}$ is positive semi-definite.

Proof: We have $\Sigma_{ij} = \mathbb{E}[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])]$. It is clear that $\Sigma_{ij} = \Sigma_{ji}$. So, $\boldsymbol{\Sigma}$ is symmetric. Pick any fixed vector $\boldsymbol{y} \in \mathbb{R}^n$. We want to show that $\langle y, \boldsymbol{\Sigma}\boldsymbol{y}\rangle \geq 0$. We can alternatively write this as

$$\begin{aligned}
\boldsymbol{y}^\top \boldsymbol{\Sigma} \boldsymbol{y} &= \boldsymbol{y}^\top \mathbb{E}\left[(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^\top\right] \boldsymbol{y} & (5)\\
&= \mathbb{E}\left[\boldsymbol{y}^\top (\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^\top \boldsymbol{y}\right] & (6)\\
&= \mathbb{E}\left[(\boldsymbol{y}^\top (\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]))^2\right]. & (7)
\end{aligned}$$

Since squares are always non-negative, we have that $\boldsymbol{\Sigma}$ is positive semi-definite. ■

2.3 Singular Value Decomposition

Let $\mathbf{A}$ be an $n \times m$ matrix with $m \leq n$. Then, $\mathbf{A}^\top \mathbf{A}$ is positive semi-definite (exercise). The square roots of eigenvalues of $\mathbf{A}^\top \mathbf{A}$ are called singular values of $\mathbf{A}$. Note that all these square roots are well-defined since all eigenvalues are non-negative by positive semi-definiteness. Formally, $s_i(\mathbf{A}) = \sqrt{\lambda_i(\mathbf{A}^\top \mathbf{A})}$ with $s_1(\mathbf{A}) \geq s_2(\mathbf{A}) \geq \cdots \geq s_m(\mathbf{A}) \geq 0$.

A variational formula for extreme singular values is

$$s_{\max}(\mathbf{A}) = \max_{\mathbf{x} \in \mathbb{R}^m, \|\boldsymbol{x}\|=1} \|\mathbf{A}\boldsymbol{x}\|_2 \quad s_{\min}(\mathbf{A}) = \min_{\boldsymbol{x} \in \mathbb{R}^m, \|\boldsymbol{x}\|=1} \|\mathbf{A}\boldsymbol{x}\|_2. \tag{8}$$

The largest singular value is also called the spectral norm of $\mathbf{A}$.

Recall B_2^m is the unit Euclidean ball in $\mathbb{R}^m$, $B_2^m = \{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x}\|_2 \leq 1\}$. Then, applying a linear transformation to B_2^m creates an ellipsoid:

$$A(B_2^m) = \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y} = \mathbf{A}\boldsymbol{x} \text{ for some } \boldsymbol{x} \in B_2^m\}. \tag{9}$$

The geometric interpretation of the singular values of $\mathbf{A}$ is that they are the semiaxes of the ellipsoid $A(B_2^m)$.

Theorem 2.4: Singular Value Decomposition

For every $n \times m$ matrix $\mathbf{A}$ with $m \leq n$, we have $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ where $\mathbf{U}$ is $n \times n$ orthogonal, $\mathbf{V}$ is $m \times m$ orthogonal, and $\mathbf{D}$ is $n \times m$ rectangular diagonal whose elements are the singular values of $\mathbf{A}$. Furthermore, the columns of $\mathbf{V}$ are right singular vectors of $\mathbf{A}$, and the columns of $\mathbf{U}$ are left singular vectors of $\mathbf{A}$.