Essential Math for Machine Learning

Deep Learning 2021

E. Fatemizadeh

Matrix and Vectors

› General Definition:

- $A=\left[a_{ij}\right]_{m\times n} \Rightarrow \underbrace{A^{T}=\left[a_{ji}\right]_{n\times m}}_{\textbf{Transpose}} \quad \underbrace{A^{H}=\left[a_{ji}^{*}\right]_{n\times m}}_{\textbf{Hermitian}} \quad A^{H}=\left(A^{*}\right)^{T}=\left(A^{T}\right)^{*}$
- $Trace\left(A\right)=\sum_{i=1}^{n}a_{ii}, \quad m=n$

Matrix and Vectors

› Matrix Norm:

- $\|A\|_1 = \underset{1\le j\le n}{Max}\sum_{i=1}^{m}\left|a_{ij}\right|$: **maximum absolute column sum of the matrix**
- $\|A\|_\infty = \underset{1\le i\le m}{Max}\sum_{i=1}^{m}\left|a_{ij}\right|$: **maximum absolute row sum of the matrix**
- $\|A\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}\right|^2 = Trace\left(A^H A\right),$ **Frobenius Norm**

Matrix and Vectors

› Vector Norm:

- $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$
- $p = 2:\quad L_2$ **Norm**
- $p = 1:\quad L_1$ **Norm**
- $p = 0:\quad$ **# of non-zero entry** L_0 **Norm**

Matrix and Vectors

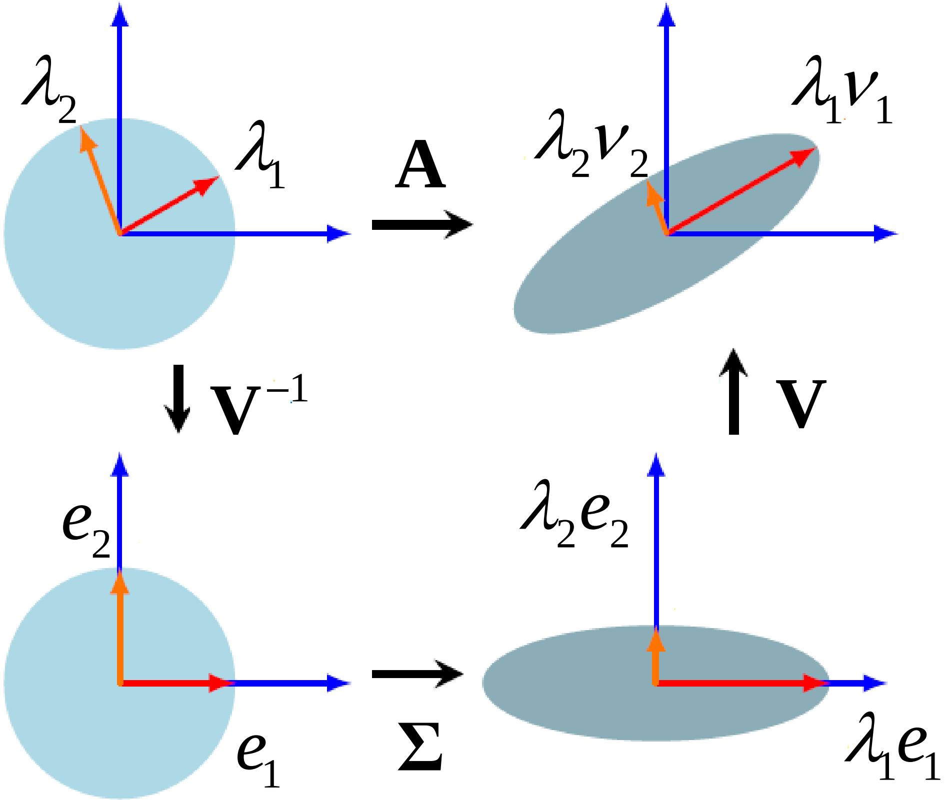
› Eigenvalues and Eigenvectors (Decomposition):

- $Av = \lambda v \Rightarrow (A - \lambda I)v = 0 \Rightarrow \{\lambda_i\}_{i=1}^{n}, \{v_i\}_{i=1}^{n}$
- $V = [v_1 | v_2 | \cdots | v_n |], \quad \Sigma = diag(\lambda_1, \lambda_2, \cdots, \lambda_n) \Rightarrow AV = V\Sigma$
- $A = V \Sigma V^{-1} \Rightarrow \Sigma = V^{-1}AV$ **Diagonalization**
- $trace(A) = \sum_{i=1}^{n} \lambda_i$
- $\det(A) = \prod_{i=1}^{n} \lambda_i$

Matrix and Vectors

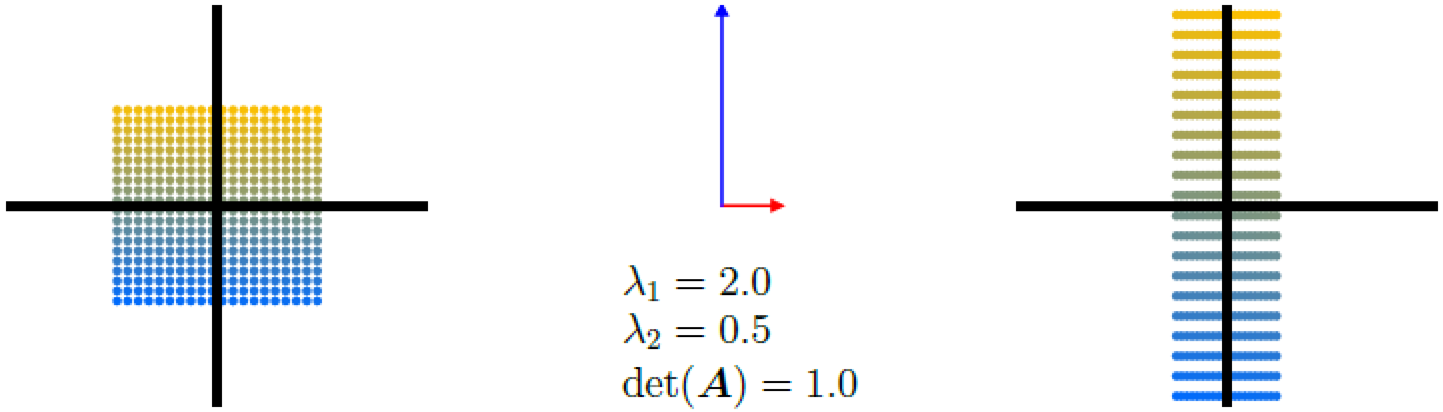
› Eigen Decomposition:

- $A\textcolor{red}{x} = V \sum V^{-1} \textcolor{red}{x}$
- $\sum = V^{-1} A V$



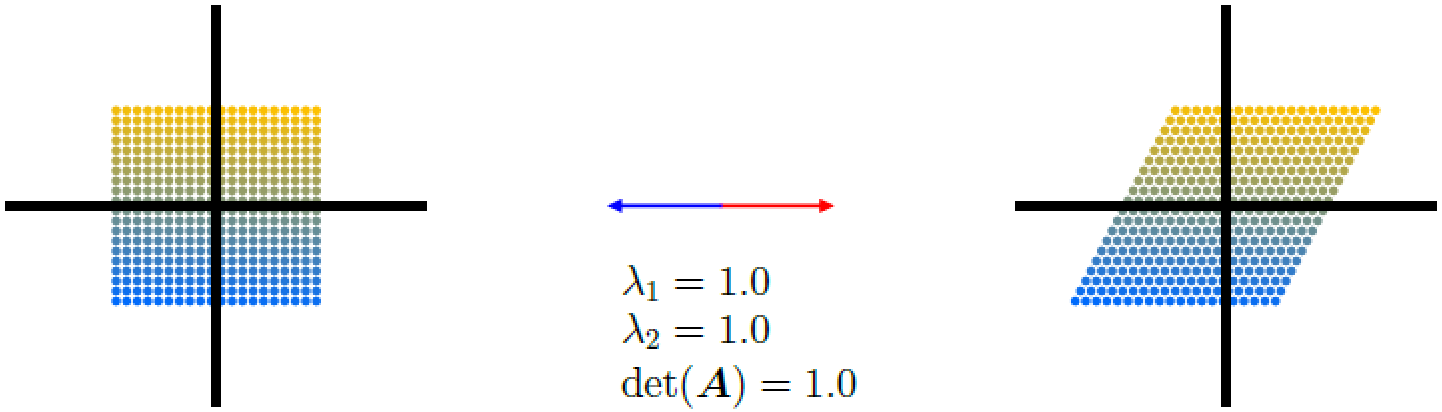
Matrix and Vectors

› Eigenvalues and Eigenvectors Mapping Effect



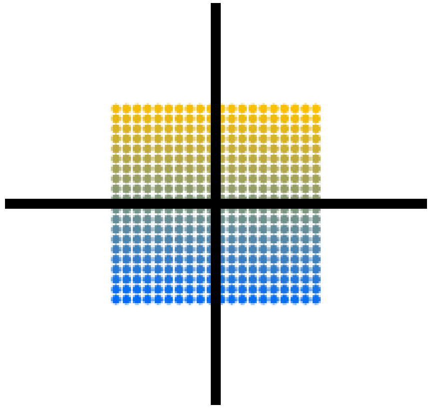
Matrix and Vectors

› Eigenvalues and Eigenvectors Mapping Effect

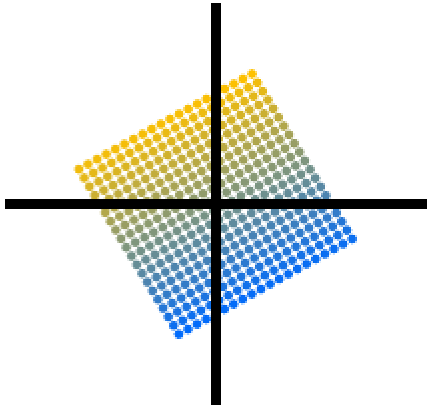


Matrix and Vectors

› Eigenvalues and Eigenvectors Mapping Effect

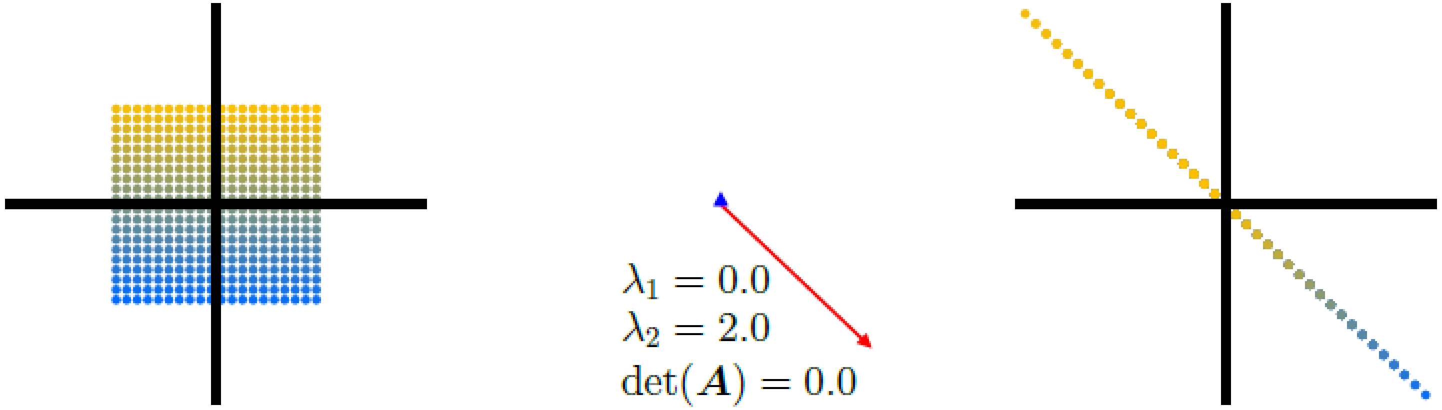


$\lambda_1 = (0.87\text{-}0.5j)$
$\lambda_2 = (0.87+0.5j)$
$\det(\boldsymbol{A}) = 1.0$



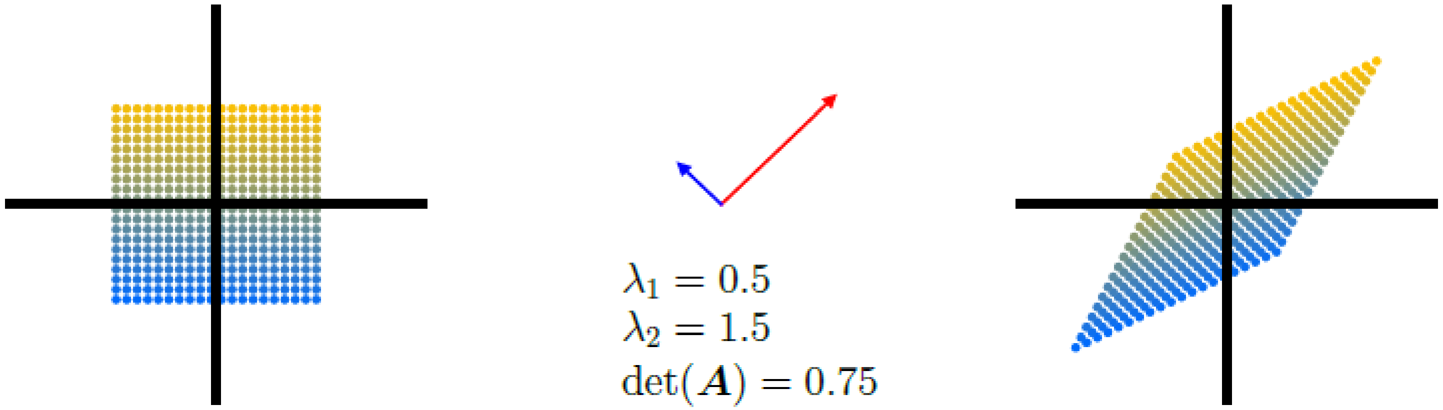
Matrix and Vectors

› Eigenvalues and Eigenvectors Mapping Effect



Matrix and Vectors

› Eigenvalues and Eigenvectors Mapping Effect



Matrix and Vectors

› Hermitian Matrix ($A=A^H$):

- All $\{\lambda_i\}_{i=1}^{n}$ are real
- $v_i^H v_j = 0, \quad v_i^H v_i = \|v_i\|^2 = 1$ *Orthonormal Vector Space*
- $A = V \Sigma V^H, VV^H = I$
- For **real** Hermitian matrix: $\{v_i\}_{i=1}^{n}$ are real
- $A = V \Sigma V^H = \sum_{i=1}^{n} \lambda_i v_i v_i^H, \quad A^{-1} = V \Sigma^{-1} V^H = \sum_{i=1}^{n} \frac{1}{\lambda_i} v_i v_i^H$

Matrix and Vectors

› Unitary Matrix ($A^{-1}=A^H$):

- Any Hermitian matrix: $V=\left[v_1 \middle| v_2 \middle| \cdots \middle| v_n \right]$ is unitary ($VV^H=V^HV=I$)

Matrix and Vectors

› Singular Value Decomposition (SVD):

- For any m×n Matrix:
  - $A = U \Sigma V^H$
  - $UU^H = I_{m\times m}$: Left Singular Vectors (AA^H)
  - $VV^H = I_{n\times n}$: Right Singular Vectors ($A^H A$)
  - Σ : Rectangular Diagonal (m×n) **nonnegative real**

Matrix and Vectors

- Positive Definite Matrix (pdm):

$$\textbf{Hermitian} \quad x^H A x > 0, \quad \textbf{any } x$$

Matrix and Vectors

› System of Linear Equations:

- $A_{m\times n} x_{n\times 1} = b_{m\times 1}$
- $m = n \Rightarrow x = A^{-1}b$
- $m > n \Rightarrow x^* = \arg\min_a \|Ax - b\|_2^2 = \left(A^H A\right)^{-1} A^H b$
- $m < n \Rightarrow x^* = \arg\min_a \|x\|_2^2 , \quad s.t.\ Ax = b \Rightarrow x^* = A^H \left(AA^H\right)^{-1} b$

Matrix and Vectors

› Useful Gradient:

- $$\frac{\partial x^T A x}{\partial x} = \left(A + A^T\right) x$$
- $$\frac{\partial b^T x}{\partial x} = b$$
- **Get *The Matrix Cookbook***

Random Vectors

$$\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$$

- Autocorrelation Matrix: $R_{xx} = E\{\mathbf{x}\mathbf{x}^H\} = [r_{ij}]$
- Covariance Matrix: $C_{xx} = E\{(\mathbf{x} - \mathbf{m}_\mathbf{x})(\mathbf{x} - \mathbf{m}_\mathbf{x})^H\} = [\sigma_{ij}]$
- Cross-Correlation Matrix: $R_{xy} = E\{\mathbf{x}\mathbf{y}^H\}$
- Cross-Covariance Matrix: $C_{xy} = E\{(\mathbf{x} - \mathbf{m}_\mathbf{x})(\mathbf{y} - \mathbf{m}_\mathbf{y})^H\} = R_{xx} - \mathbf{m}_\mathbf{x}\mathbf{m}_y^H$
- Orthogonal Vectors: $R_{xy} = \mathbf{0}$
- Uncorrelated Vectors: $C_{xy} = \mathbf{0}$

Random Vectors

› Properties of Covariance Matrix:

- $R_{xx} = R_{xx}^H, \quad C_{xx} = C_{xx}^H, \quad R_{xy} = R_{yx}^H, \quad C_{xy} = C_{yx}^H$
- $\{\lambda_i\}_{i=1}^n : real, \quad \{v_i\}_{i=1}^n : orthonormal$
- For real x, $\{v_i\}_{i=1}^n$ are real
- Non Negative Definite Matrix (NNPD):
  - $\forall z: \quad z^H C_{xx} z \geq 0$
  - $\{\lambda_i\}_{i=1}^n$: Non-Negative real
- $C_{xx} = LL^H : \quad L = V\Sigma^{0.5}$

Random Vectors

› Whitening:

- White Stochastic Vector:
  - $R_{ww} = C_{ww} = diag\left(\sigma^2 I\right)$
  - $\mathbf{m}_w = 0$
- Whitening: $\mathbf{w} = A\mathbf{x} + b \Rightarrow b = -A\mathbf{m}_x \Rightarrow \mathbf{w} = A\left(\mathbf{x} - \mathbf{m}_x\right)$
  - $A = L^{-1} = \Gamma$: Whitener/Innovation Matrix
  - $\mathbf{x} = Lw + \mathbf{m}_x$

Random Vectors

› Principal Component Analyses

$$\mathbf{x} = \in \Re^n$$

$$\mathbf{m}_x = E\{\mathbf{x}\} \approx \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k, \quad N: \text{ \# of observations}$$

$$\mathbf{C}_x = E\left\{(\mathbf{x}-\mathbf{m}_x)(\mathbf{x}-\mathbf{m}_x)^T\right\} \approx \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k\mathbf{x}_k^T - \mathbf{m}_x\mathbf{m}_x^T$$

$$\mathbf{C}_x: \text{ Real Positive Definiter Matrix}$$

$$\mathbf{C}_x v = \lambda v: \quad \lambda_i \geq 0, \quad v_i \perp v_j, i \neq j$$

$$\mathbf{A} = \left[v_1 \middle| v_2 \middle| \cdots v_n \middle|\right]^T, \quad \lambda_1 \geq \lambda_2 \cdots \geq \lambda_n \Rightarrow A^T = A^{-1}$$

Random Vectors

› Principal Component Analyses

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m_x}) \Rightarrow \mathbf{m_y} = \mathbf{0}, \quad \mathbf{C_y} = \mathbf{A}\mathbf{C_x}\mathbf{A^T} = diag\left(\begin{bmatrix}\lambda_1 & \lambda_2 & \cdots & \lambda_n\end{bmatrix}\right)$$

› Complete Synthesis:

$$\mathbf{x} = \mathbf{A}^T\mathbf{y} + \mathbf{m_x}$$

› Optimal Synthesis:

$\hat{\mathbf{A}}_k^T$: Form Using k eigenvectors corresponding to k-largest eigenvalues

$\hat{\mathbf{y}}_k$: Form Using k element correosponding to k-largest eigenvalues

$$\hat{\mathbf{x}} = \hat{\mathbf{A}}_k^T\hat{\mathbf{y}}_k + \mathbf{m_x}$$

$$e_{rms} = E\left\{\left\|\mathbf{x} - \hat{\mathbf{x}}\right\|^2\right\} = \sum_{j=1}^{n}\lambda_j - \sum_{j=1}^{k}\lambda_j = \sum_{j=k+1}^{n}\lambda_j$$

$$\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ j & k & l & m \\ n & o & p & q \end{bmatrix}\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix}$$

Random Vectors

- References:
  - Mathematics for Machine Learning, https://mml-book.github.io/
  - Matrix Cookbook, https://www2.imm.dtu.dk/pubdb/pubs/3274-full.html