CS365 Written Assignment 1

Khoa Cao

Due

Question 1

1. Let $\mathbf{X} = \mathbf{AB}$ where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$. Let $\vec{a_i}$ be the i^{th} row of $\mathbf{A}$ and $\vec{b_j}$ be the j^{th} column of $\mathbf{B}$. Then, the ij^{th} entry of $\mathbf{X}$ is given by

$$X_{ij} = \vec{a_i} \cdot \vec{b_j} \qquad \text{definition of matrix multiplication}$$

$$X_{ij} = \sum_{k=1}^{n} A_{ik} \cdot B_{kj} \qquad \text{definition of dot product}$$

Differentiating w.r.t. A_{lp} gives:

$$\frac{\partial X_{ij}}{\partial A_{lp}} = \frac{\partial (\sum_{k=1}^{n} A_{ik} \cdot B_{kj})}{\partial A_{lp}} \qquad \text{from previous result}$$

$$= \begin{cases} B_{pj} & \text{if } i = l \text{ and } k = p \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow \frac{\partial \mathbf{X}}{\partial A_{lp}} = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ B_{p1} & \cdots & B_{pj} & \cdots & B_{pn} \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \qquad \text{writing derivative in matrix form}$$

$$\Rightarrow \frac{\partial \mathbf{X}}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial \mathbf{X}}{\partial A_{11}} & \frac{\partial \mathbf{X}}{\partial A_{12}} & \cdots & \frac{\partial \mathbf{X}}{\partial A_{1n}} \\ \frac{\partial \mathbf{X}}{\partial A_{21}} & \frac{\partial \mathbf{X}}{\partial A_{22}} & \cdots & \frac{\partial \mathbf{X}}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{X}}{\partial A_{m1}} & \frac{\partial \mathbf{X}}{\partial A_{m2}} & \cdots & \frac{\partial \mathbf{X}}{\partial A_{mn}} \end{bmatrix} \qquad \text{writing derivative in matrix form}$$

$$= \begin{bmatrix} \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ B_{11} & \cdots & B_{1j} & \cdots & B_{1n} \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} & \cdots \\ \vdots & \ddots \end{bmatrix} \qquad \text{substituting previous result}$$

2. Let $\mathbf{X} = \mathbf{AB}$ where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$. Let $\vec{a_i}$ be the i^{th} row of $\mathbf{A}$ and $\vec{b_j}$ be the j^{th} column of $\mathbf{B}$. Then, the ij^{th} entry of $\mathbf{X}$ is given by

$$X_{ij} = \vec{a_i} \cdot \vec{b_j} \qquad \text{definition of matrix multiplication}$$

$$X_{ij} = \sum_{k=1}^{n} A_{ik} \cdot B_{kj} \qquad \text{definition of dot product}$$

Differentiating w.r.t. B_{lp} gives:

$$\begin{aligned}
\frac{\partial X_{ij}}{\partial B_{lp}} &= \frac{\partial(\sum_{k=1}^{n} A_{ik} \cdot B_{kj})}{\partial B_{lp}} && \text{from previous result} \\
&= \begin{cases} A_{il} & \text{if } k = l \text{ and } j = p \\ 0 & \text{otherwise} \end{cases} \\
\Rightarrow \frac{\partial \mathbf{X}}{\partial B_{lp}} &= \begin{bmatrix} 0 & \cdots & A_{1l} & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & A_{il} & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & A_{ml} & \cdots & 0 \end{bmatrix} && \text{writing derivative in matrix form} \\
\Rightarrow \frac{\partial \mathbf{X}}{\partial \mathbf{A}} &= \begin{bmatrix} \frac{\partial \mathbf{X}}{\partial B_{11}} & \frac{\partial \mathbf{X}}{\partial B_{12}} & \cdots & \frac{\partial \mathbf{X}}{\partial B_{1n}} \\ \frac{\partial \mathbf{X}}{\partial B_{21}} & \frac{\partial \mathbf{X}}{\partial B_{22}} & \cdots & \frac{\partial \mathbf{X}}{\partial B_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{X}}{\partial B_{m1}} & \frac{\partial \mathbf{X}}{\partial B_{m2}} & \cdots & \frac{\partial \mathbf{X}}{\partial B_{mn}} \end{bmatrix} && \text{writing derivative in matrix form} \\
&= \begin{bmatrix} \begin{bmatrix} 0 & \cdots & A_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & A_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & A_{m1} & \cdots & 0 \end{bmatrix} & \cdots \\ \vdots & \ddots \end{bmatrix} && \text{substituting previous result}
\end{aligned}$$

3. First, we find the scalar equation for the L_p norm of $\vec{x}$.

$$\begin{aligned}
||\vec{x}||_p^p &= \left[\left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \right]^p && \text{definition of } L_p \text{ norm} \\
&= \sum_{i=1}^{n} |x_i|^p && \text{simplifying}
\end{aligned}$$

Then, differentiating w.r.t. x_i gives:

$$
\begin{aligned}
\frac{\partial ||\vec{x}||_p^p}{\partial x_i} &= p|x_i|^{p-1} \cdot sign(x_i) && \text{chain rule} \\
\Rightarrow \frac{\partial ||\vec{x}||_p}{\partial \vec{x}} &= \begin{bmatrix} \frac{\partial ||\vec{x}||_p^p}{\partial x_1} & \frac{\partial ||\vec{x}||_p^p}{\partial x_2} & \cdots & \frac{\partial ||\vec{x}||_p^p}{\partial x_n} \end{bmatrix} && \text{writing derivative in vector form} \\
&= \begin{bmatrix} p|x_1|^{p-1} \cdot sign(x_1) & \cdots & p|x_n|^{p-1} \cdot sign(x_n) \end{bmatrix} && \text{substituting previous result} \\
&\text{where } sign(x_i) = \begin{cases} 1 & \text{if } x_i > 0 \\ -1 & \text{if } x_i < 0 \\ 0 & \text{if } x_i = 0 \end{cases}
\end{aligned}
$$

4. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$. First, we find the scalar equation for $\vec{x}^T \mathbf{A} \vec{x}$.

$$
\begin{aligned}
\vec{x}^T \mathbf{A} \vec{x} &= \begin{bmatrix} \vec{x} \cdot \vec{a_1} & \vec{x} \cdot \vec{a_2} & \cdots & \vec{x} \cdot \vec{a_n} \end{bmatrix} \cdot \vec{x} && \text{definition of matrix multiplication} \\
&= \sum_{i=1}^{n} (\vec{x} \cdot \vec{a_i}) x_i && \text{definition of dot product} \\
&= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} A_{ji} x_j \right) x_i && \text{expanding } \vec{x} \cdot \vec{a_i} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ji} x_j x_i && \text{simplifying}
\end{aligned}
$$

Then, differentiating w.r.t. x_k gives:

$$
\begin{aligned}
\frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial x_k} &= \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ji} \frac{\partial (x_j x_i)}{\partial x_k} && \text{linearity of diffe} \\
&= \sum_{i=1}^{n} A_{ki} x_i + \sum_{j=1}^{n} A_{jk} x_j && \text{only terms with } i = k \text{ or } j = \\
\Rightarrow \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial \vec{x}} &= \begin{bmatrix} \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial x_1} & \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial x_2} & \cdots & \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial x_n} \end{bmatrix} && \text{writing derivative in v} \\
&= \begin{bmatrix} \sum_i A_{1i} x_i + \sum_j A_{j1} x_j & \cdots & \sum_i A_{ni} x_i + \sum_j A_{jn} x_j \end{bmatrix} && \text{substituting prev} \\
&= \begin{bmatrix} \sum_i A_{1i} x_i & \cdots & \sum_i A_{ni} x_i \end{bmatrix} + \begin{bmatrix} \sum_j A_{j1} x_j & \cdots & \sum_j A_{jn} x_j \end{bmatrix} && \text{separating} \\
&= \mathbf{A} \vec{x} + \mathbf{A}^T \vec{x} && \text{rewriting in m} \\
&= (\mathbf{A} + \mathbf{A}^T) \vec{x} && \text{facto}
\end{aligned}
$$

5. Using the result from part (d), we have:

$$
\vec{x}^T \mathbf{A} \vec{x} = \sum_i \sum_j A_{ji} x_j x_i \qquad \text{from part (d)}
$$

Differentiating w.r.t. A_{kl} gives:

$$\begin{aligned}
\frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{kl}} &= \sum_i \sum_j \frac{\partial (A_{ji} x_j x_i)}{\partial A_{kl}} && \text{linearity of differentiation} \\
&= x_l x_k && \text{only the term with } j = k \text{ and } i = l \text{ remains} \\
\Rightarrow \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial \mathbf{A}} &= \begin{bmatrix} \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{11}} & \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{12}} & \cdots & \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{1n}} \\ \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{21}} & \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{22}} & \cdots & \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{n1}} & \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{n2}} & \cdots & \frac{\partial \vec{x}^T \mathbf{A} \vec{x}}{\partial A_{nn}} \end{bmatrix} && \text{writing derivative in matrix form} \\
&= \begin{bmatrix} x_1 x_1 & x_1 x_2 & \cdots & x_1 x_n \\ x_2 x_1 & x_2 x_2 & \cdots & x_2 x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n x_1 & x_n x_2 & \cdots & x_n x_n \end{bmatrix} && \text{substituting previous result} \\
&= \vec{x} \vec{x}^T && \text{rewriting in vector form}
\end{aligned}$$

Question 2

1. Given:

$$Q(X|\theta,\pi) = \sum_{\vec{x_i}\in X}\sum_{j=1}^{k}\gamma_{ij}\left(\log(\pi_j) + \log(Pr[\vec{x_i}|\vec{\mu}_j,\Sigma_j])\right) s.t. \sum_{j=1}^{k}\pi_{ij} \qquad = 1$$

We will use Lagrange multipliers to find the value of μ_m that maximizes $Q(X|\theta,\pi)$. Consider:

$$\begin{aligned}
Q(X|\theta,\pi) &= \sum_{\vec{x_i}\in X}\sum_{j=1}^{k}\gamma_{ij}\left(\log(\pi_j) + \log(Pr[\vec{x_i}|\vec{\mu}_j,\Sigma_j])\right) - \lambda\left(\sum_{j=1}^{k}\pi_{ij} - 1\right) && \text{adding constra}\\
\Rightarrow \frac{\partial Q}{\partial \mu_m} &= \sum_{\vec{x_i}\in X}\sum_{j=1}^{k}\gamma_{ij}\left(\log(\pi_j) + \log(Pr[\vec{x_i}|\vec{\mu}_j,\Sigma_j])\right) && \text{constant terms disapp}\\
&= \frac{\partial}{\partial\vec{\mu_m}}\sum_{\vec{x_i}\in X}\sum_{j=1}^{k}\gamma_{ij}\log\pi_j + \gamma_{ij}\log(Pr[\vec{x_i}|\vec{\mu}_j,\Sigma_j]) && \text{foi}\\
&= \sum_{\vec{x_i}\in X}\sum_{j=1}^{k}\gamma_{ij}\frac{\partial}{\partial\vec{\mu_m}}\log(Pr[\vec{x_i}|\vec{\mu_j},\Sigma_j]) && \text{constant terms disapp}\\
&= \sum_{\vec{x_i}\in X}\gamma_{im}\frac{\partial}{\partial\vec{\mu_m}}\log(Pr[\vec{x_i}|\vec{\mu_m},\Sigma_m]) && \text{only terms with } j=m \text{ rem}\\
&= \sum_{\vec{x_i}\in X}\gamma_{im}\frac{\partial}{\partial\vec{\mu_m}}\log\frac{e^{-\frac{1}{2}(\vec{x_i}-\vec{\mu_m})^T\mathbf{\Sigma}^{-1}(\vec{x_i}-\vec{\mu_m})}}{\sqrt{(2\pi)^k\,|\Sigma|}} && \text{substituting}\\
&= \sum_{\vec{x_i}\in X}\gamma_{im}\frac{\partial}{\partial\vec{\mu_m}}\log e^{-\frac{1}{2}(\vec{x_i}-\vec{\mu_m})^T\mathbf{\Sigma}^{-1}(\vec{x_i}-\vec{\mu_m})} - \log\left(\sqrt{(2\pi)^k\,|\Sigma|}\right) && \text{using log r}\\
&= \sum_{\vec{x_i}\in X}\gamma_{im}\frac{\partial}{\partial\vec{\mu_m}} - \frac{1}{2}(\vec{x_i}-\vec{\mu_m})^T\mathbf{\Sigma}^{-1}(\vec{x_i}-\vec{\mu_m}) && \text{constant terms disapp}\\
&= \sum_{\vec{x_i}\in X} -\frac{\gamma_{im}}{2}\left(\left[\mathbf{\Sigma}^{-1} + \mathbf{\Sigma}^{-1T}\right](\vec{x_i}-\vec{\mu_m})\right)\frac{\partial}{\partial\vec{\mu_m}}(\vec{x_i}-\vec{\mu_m}) && \text{result from}\\
&= \sum_{\vec{x_i}\in X} -\frac{\gamma_{im}}{2}\left(-2\mathbf{\Sigma}^{-1}(\vec{x_i}-\vec{\mu_m})\right) && \text{since } \mathbf{\Sigma} \text{ is symme}\\
&= \sum_{\vec{x_i}\in X}\gamma_{im}\mathbf{\Sigma}^{-1}(\vec{x_i}-\vec{\mu_m}) && \text{simplify}
\end{aligned}$$

Setting the equation above to 0 and solving for $\vec{\mu_m}$ gives:

$$
\begin{aligned}
&\Rightarrow \frac{\partial Q}{\partial \vec{\mu_m}} = 0 && \text{setting derivative to 0 for maximization}\\
&\Rightarrow \sum_{\vec{x_i} \in X} \gamma_{im} \mathbf{\Sigma}^{-1}(\vec{x_i} - \vec{\mu_m}) = 0 && \text{from previous result}\\
&\Rightarrow \mathbf{\Sigma}^{-1} \sum_{\vec{x_i} \in X} \gamma_{im}(\vec{x_i} - \vec{\mu_m}) = 0 && \text{factoring out } \mathbf{\Sigma}^{-1}\\
&\Rightarrow \sum_{\vec{x_i} \in X} \gamma_{im}(\vec{x_i} - \vec{\mu_m}) = 0 && \text{columns of } \mathbf{\Sigma}^{-1} \text{ are linearly independent}\\
&\Rightarrow \sum_{\vec{x_i} \in X} \gamma_{im}\vec{x_i} - \sum_{\vec{x_i} \in X} \gamma_{im}\vec{\mu_m} = 0 && \text{distributing the sum}\\
&\Rightarrow \sum_{\vec{x_i} \in X} \gamma_{im}\vec{x_i} - \vec{\mu_m} \sum_{\vec{x_i} \in X} \gamma_{im} = 0 && \text{factoring out } \vec{\mu_m}\\
&\Rightarrow \vec{\mu_m} = \frac{\sum_{\vec{x_i} \in X} \gamma_{im}\vec{x_i}}{\sum_{\vec{x_i} \in X} \gamma_{im}} && \text{solving for } \vec{\mu_m}
\end{aligned}
$$

2. Similarly, we will solve for the MLE estimate, $\mathbf{\Sigma}_m^*$. Consider:

$$
\begin{aligned}
\frac{\partial Q}{\partial \mathbf{\Sigma}_m} &= \sum_{\vec{x_i} \in X} \gamma_{im} \frac{\partial}{\partial \mathbf{\Sigma}_m} \left[\log e^{-\frac{1}{2}(\vec{x_i} - \vec{\mu_m})^T \mathbf{\Sigma}^{-1}(\vec{x_i} - \vec{\mu_m})} - \log\left(\sqrt{(2\pi)^k |\mathbf{\Sigma}|}\right)\right] && \text{from part}\\
&= \sum_{\vec{x_i} \in X} \gamma_{im} \frac{\partial}{\partial \mathbf{\Sigma}_m} \left[-\frac{1}{2}(\vec{x_i} - \vec{\mu_m})^T \mathbf{\Sigma}^{-1}(\vec{x_i} - \vec{\mu_m}) - \frac{k}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}|\right] && \text{log ru}\\
&= \sum_{\vec{x_i} \in X} -\frac{\gamma_{im}}{2} \frac{\partial}{\partial \mathbf{\Sigma}_m} \left[(\vec{x_i} - \vec{\mu_m})^T \mathbf{\Sigma}^{-1}(\vec{x_i} - \vec{\mu_m}) + k\log(2\pi) + \log|\mathbf{\Sigma}|\right] && \text{linearity of differentiati}\\
&= \sum_{\vec{x_i} \in X} -\frac{\gamma_{im}}{2} \left[\frac{\partial}{\partial \mathbf{\Sigma}_m}(\vec{x_i} - \vec{\mu_m})^T \mathbf{\Sigma}^{-1}(\vec{x_i} - \vec{\mu_m}) + \frac{\partial}{\partial \mathbf{\Sigma}_m}\log|\mathbf{\Sigma}|\right] && \text{linearity of differentiati}\\
&= \sum_{\vec{x_i} \in X} -\frac{\gamma_{im}}{2} \left[-\mathbf{\Sigma}^{-T}(\vec{x_i} - \vec{\mu_m})(\vec{x_i} - \vec{\mu_m})^T \mathbf{\Sigma}^{-T} + \mathbf{\Sigma}^{-T}\right] && \text{matrix cookbo}\\
&= \sum_{\vec{x_i} \in X} \frac{\gamma_{im}}{2} \left[\mathbf{\Sigma}^{-1}(\vec{x_i} - \vec{\mu_m})(\vec{x_i} - \vec{\mu_m})^T \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\right] && \text{since } \mathbf{\Sigma} \text{ is symmet}
\end{aligned}
$$

Setting the equation above to 0 and solving for $\mathbf{\Sigma}_m$ gives:

$$\begin{aligned}
\frac{\partial Q}{\partial \mathbf{\Sigma}_m} &= 0 && \text{setting derivative to 0 for ma} \\
\Rightarrow \sum_{\vec{x_i} \in X} \frac{\gamma_{im}}{2} \left[\mathbf{\Sigma}^{-1} (\vec{x_i} - \vec{\mu_m})(\vec{x_i} - \vec{\mu_m})^T \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1} \right] &= 0 && \text{from prev} \\
\Rightarrow \sum_{\vec{x_i} \in X} \gamma_{im} \mathbf{\Sigma}^{-1} (\vec{x_i} - \vec{\mu_m})(\vec{x_i} - \vec{\mu_m})^T \mathbf{\Sigma}^{-1} &= \sum_{\vec{x_i} \in X} \gamma_{im} \mathbf{\Sigma}^{-1} && \text{r} \\
\Rightarrow \mathbf{\Sigma}^{-1} \left(\sum_{\vec{x_i} \in X} \gamma_{im} (\vec{x_i} - \vec{\mu_m})(\vec{x_i} - \vec{\mu_m})^T \right) \mathbf{\Sigma}^{-1} &= \left(\sum_{\vec{x_i} \in X} \gamma_{im} \right) \mathbf{\Sigma}^{-1} && \text{factorin} \\
\Rightarrow \sum_{\vec{x_i} \in X} \gamma_{im} (\vec{x_i} - \vec{\mu_m})(\vec{x_i} - \vec{\mu_m})^T &= \mathbf{\Sigma} \left(\sum_{\vec{x_i} \in X} \gamma_{im} \right) && \text{multiplying both s} \\
\Rightarrow \frac{\sum_{\vec{x_i} \in X} \gamma_{im} (\vec{x_i} - \vec{\mu_m})(\vec{x_i} - \vec{\mu_m})^T}{\sum_{\vec{x_i} \in X} \gamma_{im}} &= \mathbf{\Sigma}_m^* && \text{solvi}
\end{aligned}$$

Question 3

1. Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Suppose $\mathbf{M}$ has eigenvectors $\vec{v_1}, \vec{v_2}$ with distinct eigenvalues λ_1, λ_2. We will show that these eigenvectors are orthogonal, i.e. $\vec{v_1} \cdot \vec{v_2} = 0$. Consider:

$$\begin{aligned}
\mathbf{M}\vec{v_1} \cdot \vec{v_2} &= \vec{v_1} \cdot \mathbf{M}^T\vec{v_2} && \text{given lemma}\\
\Rightarrow \mathbf{M}\vec{v_1} \cdot \vec{v_2} &= \vec{v_1} \cdot \mathbf{M}\vec{v_2} && \text{since } \mathbf{M} \text{ is symmetric}\\
\Rightarrow \lambda_1\vec{v_1} \cdot \vec{v_2} &= \lambda_2\vec{v_1} \cdot \vec{v_2} && \text{definition of eigenvector}\\
\Rightarrow \lambda_1\vec{v_1} \cdot \vec{v_2} - \lambda_2\vec{v_1} \cdot \vec{v_2} &= 0 && \text{rearranging}\\
\Rightarrow (\lambda_1 - \lambda_2)(\vec{v_1} \cdot \vec{v_2}) &= 0 && \text{factoring}\\
\Rightarrow \vec{v_1} \cdot \vec{v_2} &= 0 && \text{since } \lambda_1 \neq \lambda_2
\end{aligned}$$

Therefore, eigenvectors of a square symmetric matrix with distinct eigenvalues are orthogonal.

2. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be any matrix. We will show that $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$ and $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ are symmetric. Consider:

$$\begin{aligned}
(\mathbf{A}\mathbf{A}^T)^T &= (\mathbf{A}^T)^T\mathbf{A}^T && \text{transpose of a matrix product}\\
&= \mathbf{A}\mathbf{A}^T && \text{double transpose}\\
&\Rightarrow \mathbf{A}\mathbf{A}^T \text{ is symmetric by definition}
\end{aligned}$$

Similarly, consider:

$$\begin{aligned}
(\mathbf{A}^T\mathbf{A})^T &= \mathbf{A}^T(\mathbf{A})^T && \text{transpose of a matrix product}\\
&= \mathbf{A}^T\mathbf{A} && \text{double transpose}\\
&\Rightarrow \mathbf{A}^T\mathbf{A} \text{ is symmetric by definition}
\end{aligned}$$

3. Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be the SVD of $\mathbf{A}$, $\vec{v}$ be an eigenvector of $\mathbf{A}\mathbf{A}^T$ with eigenvalue λ. Consider $\mathbf{A}\mathbf{A}^T$:

$$\begin{aligned}
\mathbf{A}\mathbf{A}^T &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T && \text{substituting SVD of } \mathbf{A}\\
&= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T && \text{transpose of a matrix product}\\
&= \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^T && \text{since } \mathbf{V}^T\mathbf{V} = \mathbf{I}\\
&= \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T && \text{since } \mathbf{\Sigma} \text{ is diagonal}\\
\Rightarrow \mathbf{A}\mathbf{A}^T\mathbf{U} &= \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T\mathbf{U} && \text{multiplying both sides by } \mathbf{U}\\
\Rightarrow \mathbf{A}\mathbf{A}^T\mathbf{U} &= \mathbf{U}\mathbf{\Sigma}^2 && \text{since } \mathbf{U}^T\mathbf{U} = \mathbf{I}\\
\Rightarrow \begin{bmatrix}\mathbf{A}\mathbf{A}^T\vec{u_1} & \mathbf{A}\mathbf{A}^T\vec{u_2} & \cdots & \mathbf{A}\mathbf{A}^T\vec{u_m}\end{bmatrix} &= \begin{bmatrix}\sigma_1^2\vec{u_1} & \sigma_2^2\vec{u_2} & \cdots & \sigma_m^2\vec{u_m}\end{bmatrix} && \text{writing in terms of columns}
\end{aligned}$$

Therefore, the eigenvectors of $\mathbf{A}\mathbf{A}^T$ are the columns of $\mathbf{U}$ and the corresponding eigenvalues are the squares of the singular values in $\mathbf{\Sigma}$.

4. Similarly, consider $\mathbf{A}^T\mathbf{A}$:

$$\begin{aligned}
\mathbf{A}^T\mathbf{A} &= (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T && \text{substituting SVD of } \mathbf{A} \\
&= \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T && \text{transpose of a matrix product} \\
&= \mathbf{V}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{V}^T && \text{since } \mathbf{U}^T\mathbf{U} = \mathbf{I} \\
&= \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T && \text{since } \mathbf{\Sigma} \text{ is diagonal} \\
\Rightarrow \mathbf{A}^T\mathbf{A}\mathbf{V} &= \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{V} && \text{multiplying both sides by } \mathbf{V} \\
\Rightarrow \mathbf{A}^T\mathbf{A}\mathbf{V} &= \mathbf{V}\mathbf{\Sigma}^2 && \text{since } \mathbf{V}^T\mathbf{V} = \mathbf{I} \\
\Rightarrow \begin{bmatrix}\mathbf{A}^T\mathbf{A}\vec{v_1} & \mathbf{A}^T\mathbf{A}\vec{v_2} & \cdots & \mathbf{A}^T\mathbf{A}\vec{v_n}\end{bmatrix} &= \begin{bmatrix}\sigma_1^2\vec{v_1} & \sigma_2^2\vec{v_2} & \cdots & \sigma_n^2\vec{v_n}\end{bmatrix} && \text{writing in terms of columns}
\end{aligned}$$

Therefore, the eigenvectors of $\mathbf{A}^T\mathbf{A}$ are the columns of $\mathbf{V}$ and the corresponding eigenvalues are the squares of the singular values in $\mathbf{\Sigma}$.

Extra Credit

In lecture, we saw that the probability of the Monte-carlo Johnson-Lindenstrauss algorithm succeeding is $1-\delta$ where $\delta < 1$. Let X be the number of runs of the Las Vegas version of this algorithm. Then, the expected number of times we need to run the Monte-carlo algorithm until it succeeds is given by:

$$\begin{aligned}
E[X] &= \frac{1}{P[\text{success}]} && \text{definition of expected value for geometric distribution} \\
&= \frac{1}{1-\delta} && \text{substituting probability of success}
\end{aligned}$$

Therefore, the expected number of runs of the Monte-carlo Johnson-Lindenstrauss algorithm until it succeeds is $\frac{1}{1-\delta}$. Consider the expression relating the target dimension k to the number of points n, distortion ϵ and failure probability δ:

$$\begin{aligned}
k &\geq \frac{24}{3\epsilon^2 - 2\epsilon^3} \log \frac{n}{\delta} && \text{from lecture} \\
\Rightarrow \log \frac{n}{\delta} &\leq \frac{k(3\epsilon^2 - 2\epsilon^3)}{24} && \text{rearranging} \\
\Rightarrow \frac{n}{\delta} &\leq e^{\frac{k(3\epsilon^2 - 2\epsilon^3)}{24}} && \text{exponentiating both sides} \\
\Rightarrow \frac{1}{\delta} &\leq \frac{e^{\frac{k(3\epsilon^2 - 2\epsilon^3)}{24}}}{n} && \text{dividing both sides by } n \\
\Rightarrow \delta &\geq \frac{n}{e^{\frac{k(3\epsilon^2 - 2\epsilon^3)}{24}}} && \text{taking reciprocal} \\
\Rightarrow 1 - \delta &\leq 1 - \frac{n}{e^{\frac{k(3\epsilon^2 - 2\epsilon^3)}{24}}} && \text{subtracting from 1} \\
\Rightarrow \frac{1}{1-\delta} &\geq \frac{1}{1 - \frac{n}{e^{\frac{k(3\epsilon^2 - 2\epsilon^3)}{24}}}} && \text{taking reciprocal}
\end{aligned}$$

Therefore, the expected number of runs of the Monte-carlo Johnson-Lindenstrauss algorithm until it succeeds is at most

$$\frac{1}{1 - \frac{n}{e^{\frac{k(3\epsilon^2 - 2\epsilon^3)}{24}}}}.$$