Recitation 1

DS-GA 1013 Mathematical Tools for Data Science

1. If a matrix $U \in \mathbf{R}^{n\times n}$ is orthogonal, must its transpose also be orthogonal?

> **Solution:** Yes. As U is orthogonal, we have $U^TU = I$. Thus U, U^T are inverses of each other, so $UU^T = I$. This shows $(U^T)^TU^T = I$ proving U^T is orthogonal.

2. Suppose $A \in \mathbf{R}^{m\times n}$ with $trace(AA^T) = 0$. What can be said about A?

> **Solution:** We must have $A = 0$.
> Note that,
> $$trace(AA^T) = \langle A^T, A^T\rangle = \|A^T\|_F^2 = 0,$$
> implying $A^T = 0$.

3. Prove or disprove: If $A, B \in \mathbf{R}^{n\times n}$ with $trace(A) = 0 = trace(B)$ then $trace(AB) = 0$.

> **Solution:** False. Let
> $$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
> and note that $trace(A) = 0$ and $trace(A^2) = trace(I) = 2$.

4. Prove the converse to the Pythagorean theorem holds in a real inner product space: If $\|\vec{x}\|^2 + \|\vec{y}\|^2 = \|\vec{x} + \vec{y}\|^2$ then $\langle \vec{x}, \vec{y}\rangle = 0$.

> **Solution:** Note that
> $$\begin{aligned} \|\vec{x}+\vec{y}\|^2 = \langle \vec{x}+\vec{y}, \vec{x}+\vec{y}\rangle &= \langle \vec{x}, \vec{x}+\vec{y}\rangle + \langle \vec{y}, \vec{x}+\vec{y}\rangle && (1)\\ &= \langle \vec{x}, \vec{x}\rangle + \langle \vec{x}, \vec{y}\rangle + \langle \vec{y}, \vec{x}\rangle + \langle \vec{y}, \vec{y}\rangle && (2)\\ &= \|\vec{x}\|^2 + 2\langle \vec{x}, \vec{y}\rangle + \|\vec{y}\|^2. && (3)\end{aligned}$$
> By assumption, the last line is equal to $\|\vec{x}\|^2 + \|\vec{y}\|^2$ proving $\langle \vec{x}, \vec{y}\rangle = 0$.

5. Prove Bessel's inequality: Let $\vec{x}, \vec{b}_1, \ldots, \vec{b}_n \in V$ where V is a (real or complex) inner product space. Then
$$\sum_{i=1}^{n} |\langle \vec{x}, \vec{b}_i\rangle|^2 \le \|\vec{x}\|^2$$
if $\vec{b}_1, \ldots,_n$ are orthonormal.

Solution: Let $\mathcal{S} = \operatorname{span}(\vec{b}_1, \ldots, \vec{b}_n)$. Then

$$\mathcal{P}_{\mathcal{S}}\vec{x} = \sum_{i=1}^{n} \langle \vec{x}, {}_i \rangle {}_i .$$

By the Pythagorean theorem, we have

$$\|\mathcal{P}_{\mathcal{S}}\vec{x}\|^2 + \|\mathcal{P}_{\mathcal{S}^\perp}\vec{x}\|^2 = \|\vec{x}\|^2.$$

Thus

$$\|\vec{x}\|^2 \geq \|\mathcal{P}_{\mathcal{S}}\vec{x}\|^2 = \sum_{i=1}^{n} |\langle \vec{x}, {}_i \rangle|^2.$$

6. For any $\vec{x} \in \mathbb{R}^n$ show that

$$\|\vec{x}\|_\infty \leq \|\vec{x}\|_2 \leq \|\vec{x}\|_1.$$

Solution: For the first inequality, let m be such that $|\vec{x}[m]| = \|\vec{x}\|_\infty$. Then

$$\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^{n} \vec{x}[i]^2} \geq \sqrt{|\vec{x}[m]|^2} = |\vec{x}[m]| = \|\vec{x}\|_\infty.$$

For the second inequality, note that

$$\|\vec{x}\|_1^2 = \left(\sum_{i=1}^{n} |\vec{x}[i]|\right)^2 = \sum_{i=1}^{n} |\vec{x}[i]|^2 + \sum_{i \neq j} |\vec{x}[i]\vec{x}[j]| \geq \sum_{i=1}^{n} |\vec{x}[i]|^2 = \|\vec{x}\|_2^2.$$

For an alternative proof of the second inequality, let $|\vec{x}| \in \mathbb{R}^n$ be the vector with $|\vec{x}|[i] = |\vec{x}[i]|$. Then

$$\|\vec{x}\|_2^2 = |\vec{x}|^T|\vec{x}| = \operatorname{trace}(|\vec{x}|^T|\vec{x}|) = \operatorname{trace}(|\vec{x}||\vec{x}|^T) \leq \vec{1}^T|\vec{x}||\vec{x}|^T\vec{1} = \|\vec{x}\|_1^2,$$

where $\vec{1} \in \mathbb{R}^n$ is the vector with 1 in every coordinate.

7. If $X \sim \mathcal{N}(0,1)$ then we say that $X^2 \sim \chi_1^2$ (called a chi-squared distribution). Give the pdf, mean, and variance of the χ_1^2 distribution.

Solution: Let $Y = X^2$. To compute the pdf we use the cdf $F_Y(y)$ of Y for $y \geq 0$:

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \mathbb{P}(-\sqrt{y} < X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

The pdf of Y is given by

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} - \frac{f_X(-\sqrt{y})}{-2\sqrt{y}} = \frac{f_X(\sqrt{y})}{\sqrt{y}} = \frac{e^{-y/2}}{\sqrt{2\pi y}},$$

for $y > 0$ and 0 otherwise.

8. Let $\vec{\mathbf{x}}$ denote the random vector in $\mathbb{R}^n$ where each coordinate is i.i.d., taking the values $-1, 0, +1$ with equal probability (1/3 each).

    1. Compute $E[\|\vec{\mathbf{x}}\|_2^2]$.
    2. Compute $E[\|\vec{\mathbf{x}}\|_\infty]$.
    3. Compute the covariance matrix of $\vec{\mathbf{x}}$.

Solution:

1. $E[\|\vec{\mathbf{x}}\|_2^2] = \sum_{k=1}^n E[\vec{\mathbf{x}}[i]^2] = 2n/3$.
2. $E[\|\vec{\mathbf{x}}\|_\infty] = 1 - 1/3^n$.
3. Let $\Sigma = \mathrm{Cov}(\vec{\mathbf{x}})$. Then $\Sigma[i,i] = 2/3$ and $\Sigma[i,j] = 0$ for $i \neq j$ by independence.

9. Below let $\mathbf{x}$ be a random variable with finite mean $\mu = E[\mathbf{x}]$ and finite variance $\sigma^2 = \mathrm{Var}(\mathbf{x})$. Define $\bar{\mathbf{x}}_n$ to be the sample mean of $\mathbf{x}_1, \ldots, \mathbf{x}_n$:

$$\bar{\mathbf{x}}_n := \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i,$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are i.i.d. draws from the same distribution as $\mathbf{x}$.

1. Order the following quantities from smallest to largest: $E[\mu^2]$, $E[\mathbf{x}^2]$, $E[\bar{\mathbf{x}}_n^2]$.
2. ($\star$) Order the following quantities from smallest to largest: $E[e^\mu]$, $E[e^{\mathbf{x}}]$, $E[e^{\bar{\mathbf{x}}_n}]$.
3. Prove that the sample variance $\hat{\sigma}_n^2$ defined by

$$\hat{\sigma}_n^2 := \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_n)^2$$

is an unbiased estimator of the population variance σ^2.

Solution:

1. We prove

$$E[\mu^2] \leq E[\bar{\mathbf{x}}_n^2] \leq E[\mathbf{x}^2].$$

For a random variable $\mathbf{y}$ we have $E[\mathbf{y}^2] = E[\mathbf{y}]^2 + \mathrm{Var}(\mathbf{y})$. Thus we can order them by their variance. Note that

$$\mathrm{Var}(\mu) = 0, \quad \mathrm{Var}(\mathbf{x}) = \sigma^2, \quad \text{and} \quad \mathrm{Var}(\bar{\mathbf{x}}_n) = \sigma^2/n.$$

2. We prove
$$e^{\mu} \leq E[e^{\bar{\mathbf{x}}_n}] \leq E[e^{\mathbf{x}}].$$
Recall Jensen's inequality which states that if $f : \mathbb{R} \to \mathbb{R}$ is convex then
$$E[f(\mathbf{x})] \geq f(E[\mathbf{x}]).$$
Letting $f(t) = e^t$, which is convex, we obtain
$$E[e^{\bar{\mathbf{x}}_n}] = E[f(\bar{\mathbf{x}}_n)] \geq f(E[\bar{\mathbf{x}}_n]) = e^{\mu}.$$
For the other inequality, we use the same f and note
$$\begin{aligned}
E[e^{\bar{\mathbf{x}}_n}] &= E[f(\bar{\mathbf{x}}_n)] \\
&= E\left[f\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\right)\right] \\
&\leq E\left[\frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}_i)\right] && \text{(Jensen on empirical distribution)} \\
&= \frac{1}{n}\sum_{i=1}^{n} E[f(\mathbf{x}_i)] \\
&= E[f(\mathbf{x})] && \text{(Identical marginals)} \\
&= E[e^{\mathbf{x}}].
\end{aligned}$$
Our above proofs hold for any f that is convex.

3. Note that
$$\begin{aligned}
E\left[\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}}_n)^2\right] &= E\left[\sum_{i=1}^{n}\mathbf{x}_i^2 + n\bar{\mathbf{x}}_n^2 - 2\bar{\mathbf{x}}_n\sum_{i=1}^{n}\mathbf{x}_i\right] \\
&= nE[\mathbf{x}^2] - nE[\bar{\mathbf{x}}_n^2] \\
&= nE[\mathbf{x}^2] - \mathrm{Var}(\mathbf{x}) + nE[\mathbf{x}]^2 \\
&= (n-1)\,\mathrm{Var}(\mathbf{x}).
\end{aligned}$$
Dividing both sides by $n - 1$ gives $E[^2_n] = \sigma^2$.

10. Let X be a random vector taking values in $\mathbb{R}^n$ with mean $\vec{\mu} \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$. If $A \in \mathbb{R}^{m \times n}$ and $\vec{b} \in \mathbb{R}^m$ what are the mean and covariance matrix of $AX + \vec{b}$?

Solution: The mean is $A\vec{\mu} + \vec{b}$ and the covariance is given by $A\Sigma A^T$.

To see the mean, note that
$$E[A_{i,:}X] = E[\sum_{k=1}^{n} A_{ik}X[k]] = \sum_{k=1}^{n} A_{ik}E[X[k]] = A_{i,:}E[X],$$

by the linearity of expectation. Applying this to every row shows $E[AX] = AE[X]$.
To see the covariance, recall that $\mathrm{Cov}(X) = E[(X - \vec{\mu})(X - \vec{\mu})^T]$. Thus we have

$$\begin{aligned}
\mathrm{Cov}(AX + \vec{b}) &= E[(AX + \vec{b} - (A\vec{\mu} + \vec{b}))(AX + \vec{b} - (A\vec{\mu} + \vec{b}))^T] \\
&= E[(A(X - \vec{\mu}))(A(X - \vec{\mu}))^T] \\
&= E[A(X - \vec{\mu})(X - \vec{\mu})^T A^T] \\
&= AE[(X - \vec{\mu})(X - \vec{\mu})^T]A^T \\
&= A\Sigma A^T,
\end{aligned}$$

by linearity of expectation twice (for A on the left and A^T on the right).