Exercise 1 **Inner Products induce Norms** 20 credits

Let V be a vector space, and let $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ be an inner product on V. Define $||\mathbf{x}|| := \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$. Prove that $||\cdot||$ is a norm.

(Hint: To prove the triangle inequality holds, you may need the Cauchy-Schwartz inequality, $\langle \mathbf{x}, \mathbf{y}\rangle \leq ||\mathbf{x}||||\mathbf{y}||$.)

① Absolutely homogeneous

$$||\lambda x|| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\lambda^2 \langle x, x\rangle} = |\lambda|\sqrt{\langle x,x\rangle} = |\lambda|\,||x||$$

② positive definite

for $x \neq 0$: $\because ||x|| = \sqrt{\langle x,x\rangle}$ $\because \langle x,x\rangle > 0$ $\therefore \sqrt{\langle x,x\rangle} > 0$

if $x = 0$, clearly $\langle x,x\rangle = 0$ $\therefore \sqrt{\langle x,x\rangle} = 0$

if $\sqrt{\langle x,x\rangle} = 0$, which means $\langle x,x\rangle = 0$, so $x = 0$

③ triangle inequality: $||x+y|| \leq ||x|| + ||y||$

$$\sqrt{\langle x+y, x+y\rangle} = \sqrt{\langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle} = \sqrt{\langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle}$$

$\because \langle x,y\rangle \leq ||x||\,||y||$

$$\therefore \sqrt{\langle x+y, x+y\rangle} \leq \sqrt{||x||^2 + 2||x||\,||y|| + ||y||^2} \leq ||x|| + ||y||$$

$\therefore ||x+y|| \leq ||x|| + ||y||$

Exercise 2 **Vector Calculus Identities** 10+10 credits

1. Let $\mathbf{x}, \mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. Prove that $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{a}\mathbf{b}^T\mathbf{x}) = \mathbf{a}^T\mathbf{x}\mathbf{b}^T + \mathbf{b}^T\mathbf{x}\mathbf{a}^T$.
2. Let $\mathbf{B} \in \mathbb{R}^{n\times n}, \mathbf{x} \in \mathbb{R}^n$. Prove that $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{B}\mathbf{x}) = \mathbf{x}^T(\mathbf{B} + \mathbf{B}^T)$.

1. $f(x) = x^T a b^T x \in R$ $\quad f \in R^n \to R$

$$\nabla_x(f(x)) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_n}\right] \in R^{1\times n}$$

for $k \in \{1 \dots n\}$, $\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k}\left[\left(\sum_{i=1}^n x_i b_i\right)\cdot\left(\sum_{i=1}^n x_i a_i\right)\right] = b_k\left(\sum_{i=1}^n x_i a_i\right) + a_k\left(\sum_{i=1}^n x_i b_i\right)$

$$\therefore \nabla_x(f(x)) = \left[b_1\left(\sum_{i=1}^n x_i a_i\right) \quad b_2\left(\sum_{i=1}^n x_i a_i\right) \cdots b_n\left(\sum_{i=1}^n x_i a_i\right)\right] + \left[a_1\left(\sum_{i=1}^n x_i b_i\right) \quad a_2\left(\sum_{i=1}^n x_i b_i\right) \cdots a_n\left(\sum_{i=1}^n x_i b_i\right)\right]$$

$$= [a_1 \cdots a_n]\begin{bmatrix} x_1 b_1 & x_1 b_2 & \cdots & x_1 b_n \\ x_2 b_1 & & & \vdots \\ \vdots & & \ddots & \\ x_n b_1 & \cdots & & x_n b_n \end{bmatrix} + [b_1 \cdots b_n]\begin{bmatrix} x_1 a_1 & x_1 a_2 & \cdots & x_1 a_n \\ x_2 a_1 & & & \vdots \\ \vdots & & \ddots & \\ x_n a_1 & \cdots & & x_n a_n \end{bmatrix}$$

$$= a^T x b^T + b^T x a^T$$

2. $f(x) = x^T B x \in R$ $\quad f \in R^n \to R$

$$\nabla_x(f(x)) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_n}\right] \in R^{1\times n}$$

for $k \in \{1 \dots n\}$ $\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k}\left(\sum_{i=1}^n \sum_{j=1}^n x_i B_{ij} x_j\right) = \left(\sum_{j=1}^n B_{kj} x_j + x_k B_{kk}\right) + \sum_{j=1}^n x_j B_{jk} - x_k B_{kk}$

$$= \sum_{j=1}^n x_j (B_{kj} + B_{jk})$$

$$\sum_{j=1}^n x_j (B_{kj} + B_{jk}) = [x_1 \ x_2 \cdots x_n]\begin{bmatrix} B_{11}+B_{11} & B_{12}+B_{21} & \cdots & B_{1n}+B_{n1} \\ B_{21}+B_{12} & & & \vdots \\ \vdots & & \ddots & \\ B_{n1}+B_{1n} & \cdots & & B_{nn}+B_{nn} \end{bmatrix} = x^T(B + B^T)$$

Exercise 3 **Properties of Symmetric Positive Definiteness** 10 credits

Let $\mathbf{A}, \mathbf{B}$ be symmetric positive definite matrices. [1] Prove that for any $p, q > 0$ that $p\mathbf{A} + q\mathbf{B}$ is also symmetric and positive definite.

Since A and B is symmetric positive definite matrices, so the following holds.

① for any $x \neq \vec{0}$, $x^TAx > 0$ and $x^TBx > 0$

② qA and pB are symmetric clearly.

③ $A+B$ are symmetric (from lecture slide Linear Algebra 1, page 15)

so $pA+qB$ is also symmetric

$\because x^T(pA+qB)x = x^T(pAx + qBx) = x^TpAx + x^TqBx$

since q, B is scalar so $x^T(pA+qB)x = p \cdot x^TAx + q \cdot x^TBx$

$\because p, q > 0$, $x^TAx > 0$, $x^TBx > 0$ so $p \cdot x^TAx + q \cdot x^TBx > 0$

$\therefore$ it is positive definite

$\therefore$ $pA+qB$ is symmetric positive definite metrices.

Exercise 4 **General Linear Regression with Regularisation** (10+10+10+10+10 credits)

Let $\mathbf{A} \in \mathbb{R}^{N \times N}, \mathbf{B} \in \mathbb{R}^{D \times D}$ be symmetric, positive definite matrices. From the lectures, we can use symmetric positive definite matrices to define a corresponding inner product, as shown below. From the previous question, we can also define a norm using the inner products.

1. Find the gradient $\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}, \mathbf{c})$. $\mathcal{L}_{\mathbf{A},\mathbf{B},\mathbf{y},\mathbf{X}}(\boldsymbol{\theta}, \mathbf{c}) = ||\mathbf{y} - \boldsymbol{X\theta} - \mathbf{c}||^2_{\mathbf{A}} + ||\boldsymbol{\theta}||^2_{\mathbf{B}} + \|\mathbf{c}\|^2_{\mathbf{B}}$

$$L(\theta, c) = \|y - X\theta - c\|_A^2 + \|\theta\|_B^2 + \|c\|_B^2 = (y - X\theta - c)^TA(y - X\theta - c) + \theta^TB\theta + \|c\|_B^2$$

$$L(\theta, c) = (y^T - \theta^TX^T - c^T)(Ay - AX\theta - Ac) + \theta^TB\theta + \|c\|_B^2$$

$$= \left[(y^T - c^T) - \theta^Tx^T\right]\left[(Ay - Ac) - AX\theta\right] + \theta^TB\theta + \|c\|_B^2$$

$$= (y^T - c^T)(Ay - Ac) - \theta^Tx^T(Ay - Ac) - (y^T - c^T)AX\theta + \theta^Tx^TAX\theta + \theta^TB\theta + \|c\|_B^2$$

$$\nabla_\theta L(\theta, c) = \frac{d}{d\theta}\left[(y^T - c^T)(Ay - Ac) - \theta^Tx^T(Ay - Ac) - (y^T - c^T)AX\theta + \theta^Tx^TAX\theta + \theta^TB\theta + \|c\|_B^2\right)$$

$$= \frac{d}{d\theta}\left[-\theta^Tx^T(Ay - Ac)\right] + \frac{d}{d\theta}\left[-(y^T - c^T)AX\theta\right] + \frac{d}{d\theta}\left[\theta^Tx^TAX\theta\right] + \frac{d}{d\theta}(\theta^TB\theta)$$

$\because \dfrac{\partial x^Ta}{\partial x} = a^T$, $\dfrac{\partial a^Tx}{\partial x} = a^T$ (from lecture slide: Vector calculus: page 27).

$$\frac{\partial x^TBx}{\partial x} = x^T(B + B^T) \text{ (already proved it in Q2.2)}$$

$$\therefore \nabla_\theta L(\theta, c) = -[\underbrace{x^T(Ay - Ac)}_{(y^TA^T - c^TA^T)x}]^T - (y^T - c^T)AX + \theta^TX^TAX + \theta^TX^TA^Tx + \theta^TB + \theta^TB^T$$

$$= \underline{-y^TA^TX} + \underline{c^TA^TX} - \underline{y^TAX} + \underline{c^TAX} + \underline{\theta^TX^TAX} + \underline{\theta^Tx^TA^TX} + \underline{\theta^TB} + \underline{\theta^TB^T}$$

$\because A, B$ are symmetric positive definite matric, so $A + A^T = 2A$, $B + B^T = 2B$

$$\therefore \nabla_\theta L(\theta, c) = -2y^TAX + 2c^TAX + 2\theta^TX^TAX + 2\theta^TB = 2(\theta^TX^T + c^T - y^T)AX + 2\theta^TB$$

$$\therefore \nabla_\theta L(\theta, c) = 2(\theta^TX^T + c^T - y^T)AX + 2\theta^TB$$

2. Let $\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}, \mathbf{c}) = \mathbf{0}$, and solve for $\boldsymbol{\theta}$. If you need to invert a matrix to solve for $\boldsymbol{\theta}$, you should prove the inverse exists.

$$\nabla_\theta L(\theta, c) = 2(\theta^T X^T + c^T - y^T) A X + 2\theta^T B = \underline{2\theta^T X^T A X} + 2c^T A X - 2y^T A X + \underline{2\theta^T B}$$

$$= \theta^T(2X^T A X + 2B) + 2c^T A X - 2y^T A X = 0$$

$$\theta^T(X^T A X + B) = (y^T - c^T) A X$$

$$(X^T A X + B^T)\,\theta = X^T A^T (y - c)$$

$$\theta = (X^T A X + B^T)^{-1} X^T A^T (y - c)$$

3. Find the gradient $\nabla_{\mathbf{c}}\mathcal{L}(\boldsymbol{\theta}, \mathbf{c})$.

We now compute the gradient with respect to $\mathbf{c}$. $\mathcal{L}_{\mathbf{A},\mathbf{B},\mathbf{y},\mathbf{X}}(\boldsymbol{\theta}, \mathbf{c}) = ||\mathbf{y} - \boldsymbol{X}\boldsymbol{\theta} - \mathbf{c}||^2_{\mathbf{A}} + ||\boldsymbol{\theta}||^2_{\mathbf{B}} + ||\mathbf{c}||^2_{\mathbf{B}}$

$$L(\theta, c) = ||y - X\theta - c||_A^2 + ||\theta||_B^2 + ||c||_B^2 = (y - X\theta - c)^T A (y - X\theta - c) + ||\theta||_B^2 + c^T A c$$

$$L(\theta, c) = (y^T - \theta^T X^T - c^T)(Ay - AX\theta - Ac) + c^T A c + ||\theta||_B^2$$

$$L(\theta, c) = [(y^T - \theta^T X^T) - c^T][(Ay - AX\theta) - Ac] + c^T A c + ||\theta||_B^2$$

$$= (y^T - \theta^T X^T)(Ay - AX\theta) - c^T(Ay - AX\theta) - (y^T - \theta^T X^T)Ac + 2c^T A c + ||\theta||_B^2$$

$$\nabla_c L(\theta, c) = \frac{d}{dc}(-c^T(Ay - AX\theta)) + \frac{d}{dc}((\theta^T X^T - y^T)Ac) + \frac{d}{dc}(2c^T A c)$$

$$= (AX\theta - Ay)^T + (\theta^T X^T - y^T)A + 4c^T A$$

$$= \underline{\theta^T X^T A} - y^T A + \underline{\theta^T X^T A} - y^T A + 4c^T A$$

$$= (2\theta^T X^T - 2y^T + 4c^T)A$$

the statement of properties is claimed at Q4.1

4. Let $\nabla_{\mathbf{c}}\mathcal{L}(\boldsymbol{\theta}) = \mathbf{0}$, and solve for $\mathbf{c}$. If you need to invert a matrix to solve for $\mathbf{c}$, you should prove the inverse exists.

$$\nabla_c L(\theta, c) = 2\theta^T X^T A - 2y^T A + 4c^T A = 0$$

$$4c^T A = 2y^T A - 2\theta^T X^T A$$

$$c^T = \frac{1}{2}(y^T A - \theta^T X^T A)A^{-1}$$

$$c^T = \frac{1}{2}(y^T - \theta^T X^T)$$

$$c = \frac{1}{2}(y - X\theta)$$

the inverse of A exists because A is a symmetric positive definite matric

And a positive definite matric has the property:

a symmetric positive definite matrix is invertible.

5. Show that if we set $\mathbf{A} = \mathbf{I}, \mathbf{c} = \mathbf{0}, \mathbf{B} = \lambda\mathbf{I}$, where $\lambda \in \mathbb{R}$, your answer for 4.2 agrees with the analytic solution for the standard least squares regression problem with L2 regularization, given by

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\mathbf{y}.$$

In 4.2, $\theta = (X^T A X + B^T)^{-1} X^T A^T (y - c)$ and $A = I$, $c = 0$, $B = \lambda I$

$$\theta = (X^T I X + \lambda I^T)^{-1} X^T I^T (y - 0) = (X^T X + \lambda I)^{-1} X^T y$$

— Yixi Rao U6826541