1. Math Refresher

Mittwoch, 23. Oktober 2019 11:09

Problem 1: Let $\boldsymbol{x} \in \mathbb{R}^M$, $\boldsymbol{y} \in \mathbb{R}^N$ and $\boldsymbol{Z} \in \mathbb{R}^{P\times Q}$. The function $f : \mathbb{R}^M \times \mathbb{R}^N \times \mathbb{R}^{P\times Q} \to \mathbb{R}$ is defined as

$$f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}) = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{y} + \boldsymbol{B}\boldsymbol{x} - \boldsymbol{y}^T \boldsymbol{C}\boldsymbol{Z}\boldsymbol{D} - \boldsymbol{y}^T \boldsymbol{E}^T \boldsymbol{y} + \boldsymbol{F}.$$

What should be the dimensions (shapes) of the matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}, \boldsymbol{F}$ for the expression above to be a valid mathematical expression?

Problem 2: Let $\boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{M} \in \mathbb{R}^{N\times N}$. Express the function $f(\boldsymbol{x}) = \sum_{i=1}^N \sum_{j=1}^N x_i x_j M_{ij}$ using **only** matrix-vector multiplications.

$$f(\vec{x}) = \sum_i x_i \sum_j M_{ij} x_j = \vec{x}^T M \vec{x}$$

Wrong:

$$\underbrace{\underbrace{\underbrace{\vec{x}}_{N\times 1}\;\underbrace{\vec{x}^T}_{1\times N}}_{N\times N}\;\underbrace{M}_{N\times N}}_{N\times N}$$

Problem 3: Let $\boldsymbol{A} \in \mathbb{R}^{M\times N}, \boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{b} \in \mathbb{R}^M$. We are interested in solving the following system of linear equations for $\boldsymbol{x}$

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{1}$$

a) Under what conditions does the system of linear equations have a **unique** solution $\boldsymbol{x}$ for any choice of $\boldsymbol{b}$?

$M < N$: Solution not always unique

$N < M$: Solution does not always exist

$\Rightarrow M = N$. Solve: $A\vec{x} = \vec{b} \quad | A^{-1}\cdot$

$\vec{x} = A^{-1}\vec{b}$ $\boxed{A \text{ is invertible}}$ $\Leftrightarrow$ A has full rank

b) Assume that $M = N = 5$ and that $\boldsymbol{A}$ has the following eigenvalues: $\{-5, 0, 1, 1, 3\}$. Does Equation 1 have a unique solution $\boldsymbol{x}$ for any choice of $\boldsymbol{b}$? Justify your answer.

A is invertible $\Leftrightarrow$ only non-zero eigenvalues

$\Rightarrow$ NO

Problem 4: Let $\boldsymbol{A} \in \mathbb{R}^{N\times N}$. Assume that there exists a matrix $\boldsymbol{B} \in \mathbb{R}^{N\times N}$ such that $\boldsymbol{B}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{B} = \boldsymbol{I}$. What can you say about the eigenvalues of $\boldsymbol{A}$? Justify your answer.

$B = A^{-1} \Rightarrow$ only non-zero eigenvalues

Problem 5: A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ is positive semi-definite (PSD) if and only if for any $\boldsymbol{x} \in \mathbb{R}^N$ it holds that $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geq 0$. Prove that a symmetric matrix $\boldsymbol{A}$ is PSD if and only if it has no negative eigenvalues.

No negative EV $\Rightarrow$ PSD: $\qquad A \Rightarrow B \qquad \neg A \not\Rightarrow B$

A sym. $\Rightarrow$ choose orthonormal eigenvectors, express

$$\vec{x} = \sum_i w_i \underbrace{\vec{v}_i}_{\text{eigenvectors}}$$

$$x^T A x = \sum_i \sum_j w_i w_j \vec{v}_i^{\,T} A \vec{v}_j = \sum_i \sum_j w_i w_j \underbrace{\vec{v}_i^{\,T}}_{\delta_{ij}} \lambda_j \underbrace{\vec{v}_j} = \sum_i \sum_j w_i w_j \lambda_j \delta_{ij}$$

$$= \sum_i w_i^2 \lambda_i \geq 0$$

PSD $\Rightarrow$ no neg. EV:

$\lambda, \vec{v}$ eigval, eigvector of A

$$0 \leq \vec{v}^{\,T} A \vec{v} = \vec{v}^{\,T} \lambda \vec{v} = \lambda \underbrace{\|\vec{v}\|_2^2}_{\geq 0}$$

$$\Rightarrow \lambda \geq 0$$

Problem 6: Let $\boldsymbol{A} \in \mathbb{R}^{M \times N}$. Prove that the matrix $\boldsymbol{B} = \boldsymbol{A}^T \boldsymbol{A}$ is positive semi-definite for any choice of $\boldsymbol{A}$.

$$\vec{x}^{\,T} B \vec{x} = \vec{x}^{\,T} A^T A \vec{x} = (A\vec{x})^T A \vec{x} = \|A\vec{x}\|_2^2 \geq 0$$

$\square$

Problem 7: Consider the following function $f : \mathbb{R} \to \mathbb{R}$

$$f(x) = \frac{1}{2} a x^2 + b x + c$$

We are interested in solving the following optimization problem

$$\min_{x \in \mathbb{R}} \; f(x)$$

a) Under what conditions does this optimization problem have (i) a unique solution, (ii) infinitely many solutions or (iii) no solution? Justify your answer.

a) Assume that the optimization problem has a unique solution. Write down the closed-form expression for $x^\star$ that minimizes the objective function, i.e. find $x^\star = \arg\min_{x \in \mathbb{R}} f(x)$.

Problem 8: Consider the following function $g : \mathbb{R}^N \to \mathbb{R}$

$$g(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$$

where $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is a symmetric, PSD matrix, $\boldsymbol{b} \in \mathbb{R}^N$ and $c \in \mathbb{R}$.

We are interested in solving the following optimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \; g(\boldsymbol{x})$$

a) Compute the Hessian $\nabla^2 g(\boldsymbol{x})$ of the objective function. Under what conditions does this optimization problem have a unique solution?

$$\partial_{x_\ell}\partial_{x_k} g(\vec{x}) = \partial_{x_\ell}\partial_{x_k}\left(\frac{1}{2}\sum_i\sum_j x_i A_{ij} x_j + \sum_i b_i x_i + c\right) =$$

$$= \partial_{x_\ell}\left(\frac{1}{2}\sum_i \underbrace{A_{ik}}_{\text{sym} \Rightarrow A_{ki}} x_i + \frac{1}{2}\sum_j A_{kj} x_j + b_k\right) \overset{*}{=}$$

$$= \partial_{x_\ell}\left(\sum_j A_{kj} x_j + b_k\right) = A_{k\ell} = A_{\ell k}$$

$$\nabla^2 g(\vec{x}) = A$$

Unique solution (minimum) if Hessian in pos. def. $\Leftrightarrow$ A pos. def.

Non-symmetric: $\nabla^2 g(\vec{x}) = \frac{1}{2}(A + A^T)$

b) Why is it necessary for a matrix $\boldsymbol{A}$ to be PSD for the optimization problem to be well-defined? What happens if $\boldsymbol{A}$ has a negative eigenvalue?

Logic: $A \Rightarrow B \Leftrightarrow \neg B \Rightarrow \neg A$

PSD $\Leftarrow$ well-defined $\Leftrightarrow$ $\neg$ PSD $\Rightarrow$ $\neg$ well-defined

$\neg$ PSD $\Rightarrow$ neg. eigenvalue λ, corresponding eigenvector $\vec{v}$, $a \in \mathbb{R}$:

$$a\vec{v}^T A a\vec{v} + \mathcal{O}(\|a\|) = a^2 \vec{v}^T \lambda \vec{v} + \mathcal{O}(\|a\|) =$$

$$= a^2 \lambda \|\vec{v}\|_2^2 + \mathcal{O}(\|a\|)$$

$\Rightarrow$ as we take $a \to \infty$ this will keep decreasing. $\Rightarrow$ no minimum!

$$\left[f \in \mathcal{O}(g) : \forall c > 0 \; \exists x_0 > 0 \; \forall x > x_0 : |f(x)| \le c \cdot |g(x)| \right]$$

c) Assume that the matrix $\boldsymbol{A}$ is positive definite (PD). Write down the closed-form expression for $\boldsymbol{x}^\star$ that minimizes the objective function, i.e. find $\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} g(\boldsymbol{x})$.

$$\partial_{x_k} g(\vec{x}) = \sum_j A_{kj} x_j + b_k$$

$$\vec{\nabla} g(\vec{x}) = A\vec{x} + \vec{b}$$

$$\vec{\nabla} g(\vec{x}) = A\vec{x} + \vec{b}$$

Solution: $A\vec{x} + \vec{b} = 0 \iff \vec{x} = -A^{-1}\vec{b}$

A invertible!

Problem 9: Prove or disprove the following statement

$$p(a|b,c) = p(a|c) \Rightarrow p(a|b) = p(a)$$

A is random variable

Disprove by counter example

Coin: fair or unfair? $\to C$

$p(A=T) = 0{,}5$ (fair), $p(A=T) = 1$ (unfair)

$$p(C=F) = p(C=U) = 0{,}5$$

A, B coin tosses $\Rightarrow$ independent: $p(a|c) = p(a|b,c)$

But: $p(A=T) = p(A=T|C=F)\,p(C=F) + p(A=T|C=U)\,p(C=U) = 3/4$

$$p(A=T|B=T) = \frac{p(A=T, B=T)}{p(B=T)} =$$

$$= \frac{p(A=T,B=T|C=F)\,p(C=F) + p(A=T,B=T|C=U)\,p(C=U)}{p(B=T)} =$$

$$= \frac{p(A=T|B=T,C=T)\,p(B=T|C=T)\,p(C=T) + p(A=T|B=T,C=U)\,p(B=T|C=U)\,p(C=U)}{p(B=T)} =$$

$$= \frac{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + 1\cdot 1\cdot\frac{1}{2}}{3/4} = \frac{5/8}{3/4} = 5/6$$

$p(A=T) \neq p(A=T|B=T)$ $\square$

Problem 10: Prove or disprove the following statement

$$p(a|b) = p(a) \Rightarrow p(a|b,c) = p(a|c)$$

Disprove by counterexample:

A, B dice rolls

$C = A + B$

$p(A=1|C=3) = 1/2$

$p(A=1|C=3, B=2) = 1$ $\square$

Problem 11: You are given the joint PDF $p(a, b, c)$ of three continuous random variables. Show how the following expressions can be obtained using the rules of probability

1. $p(a)$
2. $p(c|a, b)$
3. $p(b|c)$

1. $p(a) = \int\int p(a,b,c)\, db\, dc$

2. $p(c|a,b) = \frac{p(a,b,c)}{p(a,b)} = \frac{p(a,b,c)}{\int p(a,b,c)\, dc}$

3. $p(b|c) = \frac{p(b,c)}{p(c)} = \frac{\int p(a,b,c)\, da}{\int\int p(a,b,c)\, da\, db}$

Problem 12: Researchers have developed a test which determines whether a person has a rare disease. The test is fairly reliable: if a person is sick, the test will be positive with 95% probability, if a person is healthy, the test will be negative with 95% probability. It is known that $\frac{1}{1000}$ of the population have this rare disease. A person (chosen uniformly at random from the population) takes the test and obtains a positive result. What is the probability that the person has the disease?

Problem 13: Let $X \sim \mathcal{N}(\mu, \sigma^2)$, and $f(x) = ax + bx^2 + c$. What is $\mathbb{E}[f(x)]$?

variance $\sigma^2 = \mathbb{E}[x^2] - \mathbb{E}[x]^2 \Leftrightarrow \mathbb{E}[x^2] = \sigma^2 + \mu^2$

mean

$\mathbb{E}[f(x)] = \mathbb{E}[ax + bx^2 + c] = a\mathbb{E}[x] + b\,\mathbb{E}[x^2] + c =$

$= a\mu + b(\sigma^2 + \mu^2) + c$

Problem 14: Let $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $g(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ (where $\boldsymbol{A} \in \mathbb{R}^{N \times N}$). What are the values of the following expressions:

- $\mathbb{E}[g(\boldsymbol{x})]$,
- $\mathbb{E}[g(\boldsymbol{x})g(\boldsymbol{x})^T]$,
- $\mathbb{E}[g(\boldsymbol{x})^T g(\boldsymbol{x})]$,
- the covariance matrix $\mathrm{Cov}[g(\boldsymbol{x})]$.

$\mathbb{E}[g(\vec{x})] = \mathbb{E}[A\vec{x}] = A\,\mathbb{E}[\vec{x}] = A\vec{\mu}$

$\mathbb{E}[g(\vec{x})g(\vec{x})^T] = \mathbb{E}[A\vec{x}(A\vec{x})^T] = \mathbb{E}[A\vec{x}\vec{x}^T A^T] =$

$= A\,\underbrace{\mathbb{E}[\vec{x}\vec{x}^T]}\,A^T = A(\Sigma + \vec{\mu}\vec{\mu}^T)A^T$

$\Sigma = \mathbb{E}[\vec{x}\vec{x}^T] - \mathbb{E}[\vec{x}]\,\mathbb{E}[\vec{x}]^T$

$$\mathbb{E}[g(\vec{x})^T g(\vec{x})] = \mathbb{E}[(A\vec{x})^T A\vec{x}] = \mathbb{E}[\vec{x}^T A^T A \vec{x}] =$$

$$= \mathbb{E}[\mathrm{Tr}[\vec{x}^T A^T A \vec{x}]] = \mathbb{E}[\mathrm{Tr}[A\vec{x}\vec{x}^T A^T]] =$$

Trace of scalar = scalar

Tr is invariant under cyclic permutations

$$\mathrm{Tr}[ABC] = \mathrm{Tr}[CAB] = \mathrm{Tr}[BCA]$$

$$= \mathrm{Tr}[\mathbb{E}[A\vec{x}\vec{x}^T A^T]] =$$

Trace is a sum

$$= \mathrm{Tr}[A(\Sigma + \vec{\mu}\vec{\mu}^T)A^T]$$

$$\mathrm{Cov}[g(\vec{x})] = \mathbb{E}[g(\vec{x})g(\vec{x})^T] - \mathbb{E}[g(\vec{x})]\mathbb{E}[g(\vec{x})]^T =$$

$$= A(\Sigma + \vec{\mu}\vec{\mu}^T)A^T - A\mu(A\mu)^T =$$

$$= A\Sigma A^T + A\cancel{\vec{\mu}\vec{\mu}^T}A^T - A\cancel{\mu\mu^T}A^T =$$

$$= A\Sigma A^T$$