Exercise

01

TUM Department of Informatics

Supervised by Prof. Dr. Stephan Günnemann

Informatics 3 - Professorship of Data Mining and Analytics

Submitted by Marcel Bruckner

Julian Hohenadel

Kevin Bein (03707775)

Submission date Munich, October 17, 2019

Linear Algebra

Problem 1: Dimensions of matrices A, B, C, D, E, F

$$A \in \mathbb{R}^{M\times N}, \qquad B \in \mathbb{R}^{1\times M}, \qquad C \in \mathbb{R}^{N\times P} \tag{1}$$

$$D \in \mathbb{R}^{Q}, \qquad B \in \mathbb{R}^{N\times N}, \qquad C \in \mathbb{R}^{1} \tag{2}$$

Problem 2: $f(x) = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j M_{ij}$ using only matrix-vector multiplications.

$$f(x) = x^T M x \tag{3}$$

Problem 3:

(a) Conditions for unique solution x for any choice of b in $Ax = b$

$$rank(A) = M, \qquad det(A) \neq 0, \qquad ker(A) = \{0\}$$

(b) Unique solution x for any choice of b in $Ax = b$ with eigenvalues of A: $\{-5, 0, 1, 1, 3\}$

$det(A) = \prod_i \lambda_i = -5 * 0 * 1 * 1 * 3 = 0 \implies$ No unique solution

Problem 4: Properties of eigenvalues of A in $BA = AB = I$

$$BA = AB = I \implies B = A^{-1} \tag{4}$$

A has to be invertable $\implies det(A) \neq 0 \implies \forall i : \lambda_i \neq 0$

Problem 5: A is PSD if and only if it has no negative eigenvalues

Definition of eigenvalue: $Ax = \lambda x$

$$PSD \Leftrightarrow x^T A x \geq 0 \tag{5}$$

$$PSD \Leftrightarrow x^T A x = x^T \lambda x = \lambda x^T x = \lambda \sum_i x_i^2 \geq 0 \tag{6}$$

$$\sum_i x_i^2 \geq_{always} 0 \implies \forall \lambda : \lambda \geq 0 \tag{7}$$

Problem 6: $B = A^T A$ is PSD for any A

$$B = A^T A \implies Bx = \lambda_B x = A^T A x = \lambda_A \lambda_A x = \lambda_A^2 x \tag{8}$$

$$\lambda_B = \lambda_A^2 \implies \lambda_B \geq_{always} 0 \tag{9}$$

B has to be PSD for any choice of A. $\square$

Calculus

Problem 7:

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

   (i) The function got a global minimum.

   (ii) The function got multiple local minima.

   (iii) The function is not bounded below.

(b) Assume that the optimization problem has a unique solution. Write down the closed-form expression for $x^\star$ that minimizes the objective function.

$$f'(x) \overset{!}{=} 0$$

Problem 8:

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

$$g(x) = \frac{1}{2} \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b^T \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c$$

$$\implies \nabla^2 g(x) = \begin{bmatrix} A_{11} & 0 & \dots & 0 \\ 0 & A_{22} & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & A_{nn} \end{bmatrix}$$

Unique solution: No entry on the principle diagonal of the Hessian can be 0, else the determinant would be 0.

(b) Why is it necessary for a matrix **A** to be PSD for the optimization problem to be well-defined? What happens if A has a negative eigenvalue?

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

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

$g(x) = \frac{1}{2}\lambda_A \sum_i x_i^2 + \sum_i b_i^T x_i + c$

Curvatures are the same in all dimensions, else a good approximation would not be possible.

(c) Assume that the matrix **A** is positive definite (PD). Write doen the closed-form expression for $x^\star$ that minimizes the objective function.

$$g'(x) \stackrel{!}{=} 0 \implies$$

$\frac{1}{2}x^T A + b^T = 0$

$x^T A = -2b^T$

$x^T = -2b^T A^{-1}$

$x = (-2b^T A^{-1})^T$

Probability Theory

Problem 9: *Missing counter example*

Problem 10:

$$P(A) = P(A|B) \overset{Bayes}{=} \frac{P(A,B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

$$P(A|B,C) = \frac{P(A,B|C)}{P(B|C)} = \frac{P(A|C)P(B|C)}{P(B|C)} = P(A|C)$$

Problem 11:

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

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

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

Problem 12:

Extracting the variables from the Text:

$P(T)$:= Test positive	$P(S)$:= sick	$P(T|S) = 0.95$	$P(\neg T|\neg S) = 0.95$	$P(S) = \frac{1}{1000} = 0.001$
$P(\neg T)$:= Test negative	$P(\neg S)$:= healthy	$P(T|\neg S) = 0.05$	$P(\neg T|S) = 0.05$	$P(\neg S) = 0.999$

Calculation of $P(S|T)$:

$$\begin{aligned} P(S|T) &= \frac{P(T|S)P(S)}{P(T|S)P(S) + P(T|\neg S)(\neg S)} \\ &= \frac{0.95 \cdot 0.00.1}{0.95 \cdot 0.001 + 0.05 \cdot 0.999} \\ &\approx 0.019 \end{aligned}$$

Problem 13:

Given

- $\mathbb{E}[x - \mu] = \mathbb{E}[\mu] = \mu - \mu = 0$
- $Var[x] = \sigma^2 = [(x-\mu)^2]a$

the expected values of $f(x)$ becomes the following:

$$
\begin{aligned}
\mathbb{E}[f(x)] &= \mathbb{E}[ax + bx^2 + c] && (10)\\
&= \mathbb{E}[ax] + \mathbb{E}[bx^2] + \mathbb{E}[c] && (11)\\
&= a\mathbb{E}[x] + b\mathbb{E}[x^2] + c && (12)\\
&= a\mu + b\mathbb{E}[(x-\mu+\mu)^2] + c && (13)\\
&= a\mu + c + b(\mathbb{E}[((x-\mu)+\mu)((x-\mu)+\mu)]) \qquad \phi := x - \mu && (14)\\
&= a\mu + c + b(\mathbb{E}[(\phi+\mu)^2]) && (15)\\
&= a\mu + c + b(\mathbb{E}[\phi^2 + 2\phi\mu + \mu^2]) && (16)\\
&= a\mu + c + b(\mathbb{E}[\phi^2] + \mathbb{E}[2\phi\mu] + \mathbb{E}[\mu^2]) && (17)\\
&= a\mu + c + b(\mathbb{E}[(x-\mu)^2] + 2\mu\mathbb{E}[x-\mu] + \mathbb{E}[\mu^2]) && (18)\\
&= a\mu + c + b(\sigma^2 + 0 + \mu^2) && (19)\\
&= a\mu + c + b\sigma^2 + b\mu^2 && (20)
\end{aligned}
$$

Problem 14:

- $$
\begin{aligned}
\mathbb{E}[g(x)] &= \mathbb{E}[Ax]\\
&= A\mathbb{E}[x]\\
&= A\mu
\end{aligned}
$$

- $$
\begin{aligned}
\mathbb{E}[g(x)g(x)^T] &= \mathbb{E}[Ax(Ax)^T]\\
&= A\mathbb{E}[x(ax)^T]\\
&= A\mathbb{E}[xx^TA^T]\\
&= A\mathbb{E}[xx^T]A^T\\
&= A(\Sigma + \mu\mu^T)A^T\\
&= AA^T\Sigma + AA^T\mu\mu^T
\end{aligned}
$$

- $$\begin{aligned}
\mathbb{E}[g(x)^T g(x)] &= \mathbb{E}[(Ax)^T Ax] \\
&= \mathbb{E}[x^T A^T Ax] && B := A^T A \\
&= \mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N} B_{i,j} x_i x_j] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} B_{i,j}\mathbb{E}[x_i x_j] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} B_{i,j}(\sigma_{i,j} + \mu_i\mu_j) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} B_{i,j}\sigma_{i,j} + \sum_{i=1}^{N}\sum_{j=1}^{N} B_{i,j}\mu_i\mu_j \\
&= \sum_{i=1}^{N}(B\Sigma)_{i,i} = \mu^T B\mu \\
&= tr(A^T A\Sigma) + \mu^T A^T A\mu
\end{aligned}$$

- $$\begin{aligned}
Cov[g(x)] &= Cov[Ax] \\
&= \mathbb{E}[(Ax - \mathbb{E}[Ax])(Ax - \mathbb{E}[Ax])^T] \\
&= \mathbb{E}[(Ax - A\mathbb{E}[x])(Ax - A\mathbb{E}[x])^T] \\
&= \mathbb{E}[A(x - A\mathbb{E}[x])(x - \mathbb{E}[x])^T A^T] \\
&= A\mathbb{E}[(x - A\mathbb{E}[x])(x - \mathbb{E}[x])^T]A^T \\
&= A\,Cov(x)A^T
\end{aligned}$$

Appendix

We confirm that the submitted solution is original work and was written by us without further assistance. Appropriate credit has been given where reference has been made to the work of others.

Munich, October 17, 2019, Signature Marcel Bruckner

Munich, October 17, 2019, Signature Julian Hohenadel

Munich, October 17, 2019, Signature Kevin Bein (03707775)