

# Statistical Machine Learning (STAT W4400)

## Homework 0

Linan Qiu

1q2137

September 16, 2015

Let

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

1.  $B_{2,1}$  is 3

$$2. A + B = \begin{bmatrix} 2 & 4 \\ 5 & 8 \end{bmatrix}$$

$$3. AB = \begin{bmatrix} 1 * 1 + 2 * 3 & 1 * 2 + 2 * 4 \\ 2 * 1 + 4 * 3 & 2 * 2 + 4 * 4 \end{bmatrix} = \begin{bmatrix} 7 & 10 \\ 14 & 20 \end{bmatrix}$$

4.  $\text{rank}(A) = 1$

$$Av = \lambda v$$

$$(A - \lambda I)v = 0$$

$$\det(A - \lambda I) = 0$$

$$= (1 - \lambda)(4 - \lambda) - 4$$

$$= \lambda^2 - 5\lambda$$

$$= \lambda(\lambda - 5)$$

5. Largest eigenvalue of  $A$  is 5

$$\begin{aligned}
(A - \lambda I)v &= 0 \\
(A - 5I)v &= 0 \\
\begin{bmatrix} 1-5 & 2 \\ 2 & 4-5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= 0
\end{aligned}$$

Then,

$$\begin{aligned}
-4v_1 + 2v_2 &= 0 \\
2v_1 - 1v_2 &= 0
\end{aligned}$$

Then, let  $v_2 = t$ ,  $v_1 = -2t$ . Then, the eigenspace corresponding to  $\lambda = 5$  is given by the span of  $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ .

6. Eigenvector associated is the span of  $v = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

7.  $|B| = 1 * 4 - 2 * 3 = -2$

8.  $x^T A x = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 16$

9.  $x^T x = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 5$

10.  $x x^T = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$

11.  $\|x\|_2 = \sqrt{2^2 + 1^2} = \sqrt{5}$

$$\nabla_x x^T A x = x^T A^T + x^T A = x^T (A^T + A)$$

For the special case that  $A$  is symmetric (as it is given),

$$\nabla_x x^T A x = 2x^T A$$

12.  $\nabla_x x^T A x = 2x^T A$

$$\nabla_x^2 x^T A x = \nabla_x (x^T A^T + x^T A) = A + A^T$$

For the special case that  $A$  is symmetric (as it is given),

$$\nabla_x^2 x^T A x = 2A$$

13.  $\nabla_x^2 x^T A x = 2A$

14.  $n = 2$

$$\begin{aligned} \text{Var}(y) &= E(y)^2 - E(y^2) \\ E(y^2) &= E(y)^2 - \text{Var}(y) \\ &= \mu^2 - \sigma^2 \end{aligned}$$

15.  $E(y^2) = \mu^2 - \sigma^2$

16.  $y + w \sim N(2.7 + 3.1, \sqrt{8^2 + 15^2}) = N(5.8, 17)$

17. The normalizing constant is  $(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}$  where  $p = 2$  and  $|\Sigma|$  is the determinant of the  $\Sigma$  given.

18. The support of a Bernoulli random variable is  $\{0, 1\}$

19. Define the Bernoulli process as  $N(k, n)$ . Then  $N(k, n) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$

$$\begin{aligned} h(x_1) &= \frac{1}{3}x_1^3 - \frac{1}{2}x_1^2 - 6x_1 + \frac{27}{2} \\ h'(x_1) &= x_1^2 - x_1 - 6 \\ h''(x_1) &= 2x_1 - 1 \end{aligned}$$

Minima/maxima is at

$$\begin{aligned} h'(x_1) &= x_1^2 - x_1 - 6 = 0 \\ (x_1 - 3)(x_1 + 2) &= 0 \end{aligned}$$

$x_1 = 3$  is a minima since  $h''(3) > 0$ .  $x_1 = -2$  is a maxima since  $h''(-2) < 0$ . To compare with boundary values,

$$\begin{aligned} h(3) &= 0 \\ h(-2) &= 20.83 \\ h(-4) &= 8.17 \\ h(4) &= 2.83 \end{aligned}$$

20.  $x_1 = -2$

21.  $x_1 = 3$

The indefinite integral

$$\int \frac{1}{Z} \left( \frac{1}{3}x_1^3 - \frac{1}{2}x_1^2 - 6x_1 + \frac{27}{2} \right) = \frac{1}{Z} \left( \frac{1}{12}x^4 - \frac{1}{6}x^3 - 3x^2 + \frac{27}{2}x + C \right)$$

Set  $C$  and  $Z$  such that

$$\begin{aligned}\int_0^1 \frac{1}{Z} \left( \frac{1}{3}x_1^3 - \frac{1}{2}x_1^2 - 6x_1 + \frac{27}{2} \right) &= 1 \\ \frac{1}{Z} \left( \frac{1}{12} - \frac{1}{6} - 3 + \frac{27}{2} \right) &= 1 \\ Z &= \frac{1}{12} - \frac{1}{6} - 3 + \frac{27}{2}\end{aligned}$$

22.  $Z = \frac{1}{12} - \frac{1}{6} - 3 + \frac{27}{2}$

$$\begin{aligned}\int_A b(x) &= \int_0^3 \int_0^2 x_1 x_2^3 dx_2 dx_1 \\ &= \int_0^3 \left[ x_1 \frac{1}{4} x_2^4 \right]_0^2 dx_1 \\ &= \int_0^3 4x_1 dx_1 \\ &= [2x_1^2]_0^3 \\ &= 18\end{aligned}$$

23. 18

$$L(x_1, x_2; \lambda) = x_1 + \sqrt{3}x_2 + \lambda(x_1^2 + x_2^2 - 1)$$

$$\begin{aligned}\frac{\delta L}{\delta x_1} &= 1 + \lambda 2x_1 \\ \frac{\delta L}{\delta x_2} &= \sqrt{3} + \lambda 2x_2 \\ \frac{\delta L}{\delta \lambda} &= x_1^2 + x_2^2 - 1\end{aligned}$$

Using first 2 equations,  $x_1 = \frac{1}{\sqrt{3}}x_2$ . Solving trivially, we have

$$\begin{aligned}x_1 &= \pm \frac{1}{2} \\ x_2 &= \pm \frac{\sqrt{3}}{2}\end{aligned}$$

24. The  $x$  that maximizes  $c(x)$  subject to the constraints specified is  $x = \left[ \begin{array}{c} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{array} \right]$

$$L(x_1, x_2; \lambda) = -x_1 \log x_1 - x_2 \log x_2 + \lambda(x_1 + x_2 - 1)$$

$$\frac{\delta L}{\delta x_1} = -1 - \log x_1 + \lambda$$

$$\frac{\delta L}{\delta x_2} = -1 - \log x_2 + \lambda$$

$$\frac{\delta L}{\delta \lambda} = x_1 + x_2 - 1$$

By visual inspection, if  $\log x_1 = \log x_2$ , then  $x_1 = x_2 = 0.5$

25.  $x = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$