Yang, Ya-Chu
11571276
qwerthjkl45@gmail.com

Homework Assignment 1
Machine Learning 1, 17/18

2017-09-13

1 Basic Linear Algebra and Derivatives

1.1

1. Compute Ab

$$Ab = \begin{bmatrix} -7 & 8 & 1 \\ -4 & 3 & 5 \\ 7 & 7 & -8 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} (-7 \times 1) + (8 \times 2) + (1 \times 5) \\ (-4 \times 1) + (3 \times 2) + (5 \times 5) \\ (7 \times 1) + (7 \times 2) + (-8 \times 5) \end{bmatrix} = \begin{bmatrix} 14 \\ 27 \\ -19 \end{bmatrix}$$

2. Compute $b^T A$

$$b^T = \begin{bmatrix} 1 & 2 & 5 \end{bmatrix}$$

$$b^T A = \begin{bmatrix} 1 & 2 & 5 \end{bmatrix} \times \begin{bmatrix} -7 & 8 & 1 \\ -4 & 3 & 5 \\ 7 & 7 & -8 \end{bmatrix} = \begin{bmatrix} (-7 \times 1) + (-4 \times 2) + (7 \times 5) \\ (8 \times 1) + (3 \times 2) + (7 \times 5) \\ (1 \times 1) + (5 \times 2) + (-8 \times 5) \end{bmatrix} = \begin{bmatrix} 20 \\ 49 \\ -29 \end{bmatrix}$$

3.Compute the vector c for which Ac = b through elimination

$$\begin{cases} -7x_1 + 8x_2 + x_3 = 1 \\ -4x_1 + 3x_2 + 5x_3 = 2 \\ 7x_1 + 7x_2 - 8x_3 = 5 \end{cases} \Rightarrow \begin{cases} x_1 - \frac{8}{7}x_2 - \frac{1}{7}x_3 = 1 \\ -4x_1 + 3x_2 + 5x_3 = 2 \\ 7x_1 + 7x_2 - 8x_3 = 5 \end{cases} \Rightarrow \begin{cases} x_1 - \frac{8}{7}x_2 - \frac{1}{7}x_3 = 1 \\ \frac{-11}{7}x_2 + \frac{31}{7}x_3 = 2 \\ 15x_2 - 7x_3 = 6 \end{cases} \Rightarrow \begin{cases} x_1 - \frac{8}{7}x_2 - \frac{1}{7}x_3 = 1 \\ \frac{-11}{7}x_2 + \frac{31}{7}x_3 = 2 \\ 15x_2 - 7x_3 = 6 \end{cases} \Rightarrow$$

$$\begin{cases} x_1 - \frac{37}{11}x_3 = \frac{13}{11} \\ x_2 - \frac{31}{11}x_3 = \frac{-10}{11} \\ x_3 = \frac{54}{97} \end{cases} \Rightarrow \begin{cases} x_1 - \frac{37}{11}x_3 = \frac{13}{11} \\ x_2 - \frac{31}{11}x_3 = \frac{-10}{11} \\ x_3 = \frac{54}{97} \end{cases} \Rightarrow \begin{cases} x_1 = \frac{67}{97} \\ x_2 = \frac{64}{97} \\ x_3 = \frac{54}{97} \end{cases} \Rightarrow \begin{bmatrix} \frac{67}{97} \\ \frac{64}{97} \\ \frac{54}{97} \end{bmatrix}$$

4.

$$\left[\begin{array}{ccc|ccc} -7 & 8 & 1 & 1 & 0 & 0 \\ -4 & 3 & 5 & 0 & 1 & 0 \\ 7 & 7 & -8 & 0 & 0 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 1 & -\frac{8}{7} & -\frac{1}{7} & -\frac{1}{7} & 0 & 0 \\ -4 & 3 & 5 & 0 & 1 & 0 \\ 7 & 7 & -8 & 0 & 0 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 1 & -\frac{8}{7} & -\frac{1}{7} & -\frac{1}{7} & 0 & 0 \\ 0 & 1 & -\frac{31}{11} & \frac{4}{11} & -\frac{7}{11} & 0 \\ 0 & 15 & -7 & 1 & 0 & 1 \end{array}\right] \Rightarrow$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -\frac{37}{11} & \frac{3}{11} & -\frac{8}{11} & 0 \\ 0 & 1 & -\frac{31}{11} & \frac{4}{11} & -\frac{7}{11} & 0 \\ 0 & 0 & \frac{388}{11} & -\frac{49}{11} & \frac{105}{11} & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & -\frac{37}{11} & \frac{3}{11} & -\frac{8}{11} & 0 \\ 0 & 1 & -\frac{31}{11} & \frac{4}{11} & -\frac{7}{11} & 0 \\ 0 & 0 & 1 & -\frac{49}{388} & \frac{105}{388} & \frac{11}{388} \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{59}{388} & -\frac{71}{388} & \frac{37}{388} \\ 0 & 1 & 0 & \frac{3}{388} & \frac{49}{388} & \frac{31}{388} \\ 0 & 0 & 1 & -\frac{49}{388} & \frac{105}{388} & \frac{11}{388} \end{array}\right]$$

$$A^{-1} = \begin{bmatrix} -\frac{59}{388} & -\frac{71}{388} & \frac{37}{388} \\ \frac{3}{388} & \frac{49}{388} & \frac{31}{388} \\ -\frac{49}{388} & \frac{105}{388} & \frac{11}{388} \end{bmatrix}$$

5.

$$A^{-1}b = \begin{bmatrix} -\frac{59}{388} & -\frac{71}{388} & \frac{37}{388} \\ \frac{3}{388} & \frac{49}{388} & \frac{31}{388} \\ -\frac{49}{388} & \frac{105}{388} & \frac{11}{388} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} \frac{67}{97} \\ \frac{64}{97} \\ \frac{54}{97} \end{bmatrix}$$

1.2

1.

$$\frac{d}{dx}\left[\left(\frac{2}{x^2} + x^{-7} + x^3\right)^2\right] = 2(\frac{2}{x^2} + x^{-7} + x^3)\left(\frac{\partial\left(\frac{2}{x^2} + x^{-7} + x^3\right)}{\partial x}\right) = 2(2x^{-2} + x^{-7} + x^3)(-4x^{-3} - 7x^{-8} + 3x^2)$$
$$= 6x^5 + 4 - 24x^{-5} - 36x^{-10} - 14x^{-15}$$

2.

$$\frac{d}{dx}\left[\left(x^2\sqrt{e^{-\sqrt[3]{x}}}\right)\right] = \frac{d}{dx}\left[x^2\left(e^{-\frac{x^3}{2}}\right)^{\frac{1}{2}}\right] = \frac{d}{dx}\left[x^2\left(e^{-\frac{x^3}{4}}\right)\right] = 2xe^{-\frac{x^3}{4}} + x^2\left(-\frac{3x^2}{4}e^{-\frac{x^3}{4}}\right) = 2xe^{-\frac{x^3}{4}} - \frac{3x^4}{4}e^{-\frac{x^3}{4}}$$

3.

$$\frac{d}{dx}\left[x + ln\left(x\right)\right] = 1 + \frac{1}{x}$$

4.

$$\frac{d}{dx}\left[xln(\sqrt{x})\right] = \frac{d(x)}{dx}ln(\sqrt{x}) + x\frac{d(ln(\sqrt{x})))}{dx} = ln(\sqrt{x}) + x\frac{1}{\sqrt{x}}\frac{1}{2\sqrt{x}} = ln(\sqrt{x}) + \frac{1}{2} = \frac{1}{2}ln(x) + \frac{1}{2}$$

5.

$$\frac{d}{dx}\left[6(x^2-1)\sin x\right] = 6\sin x\left[\frac{d(x^2-1)}{dx}\right] + 6(x^2-1)\left[\frac{d(\sin x)}{dx}\right] = 12x\sin x + 6(x^2-1)\cos x$$

6.

$$\frac{d}{dx}\left[ln\left(\sqrt[3]{\frac{e^{3x}}{1+3^{3x}}}\right)\right] = \frac{d}{dx}\left[ln(e^x) - \frac{1}{3}ln(1+e^{3x})\right] = 1 - \frac{1\times(e^{3x}]\times 3}{3\times(1+e^{3x})} = \frac{1}{1+e^{3x}}$$

7.
8.

$$\begin{aligned}
\frac{\partial\left(2ln\left(y - exp(x^{-1}) - \sin(zx^2)\right)\right)}{\partial x} &= 2\frac{\frac{\partial\left(y-exp(x^{-1})-\sin(zx^2)\right)}{\partial x}}{(ln\left(y - exp(x^{-1}) - \sin(zx^2)\right))} = \frac{2\times\left(e^{\frac{1}{x}}x^{-2} - 2zx\cos(zx^2\left(\right)\right.}{(y - exp(x^{-1}) - \sin(zx^2))} \\
\frac{\partial\left(2ln\left(y - exp(x^{-1}) - \sin(zx^2)\right)\right)}{\partial y} &= 2\frac{\frac{\partial\left(y-exp(x^{-1})-\sin(zx^2)\right)}{\partial y}}{(ln\left(y - exp(x^{-1}) - \sin(zx^2)\right))} = \frac{2}{(ln\left(y - exp(x^{-1}) - \sin(zx^2)\right))} \\
\frac{\partial\left(2ln\left(y - exp(x^{-1}) - \sin(zx^2)\right)\right)}{\partial z} &= 2\frac{\frac{\partial\left(y-exp(x^{-1})-\sin(zx^2)\right)}{\partial z}}{(ln\left(y - exp(x^{-1}) - \sin(zx^2)\right))} = \frac{-\cos(zx^2)}{(ln\left(y - exp(x^{-1}) - \sin(zx^2)\right))}
\end{aligned}$$

9.

$$\begin{aligned}
\frac{\partial\left(ln\left(\sqrt[\gamma]{z^\alpha y^\beta x^\gamma}\right)\right)}{\partial x} &= \frac{\partial\left(ln\left(x\sqrt[\gamma]{z^\alpha y^\beta}\right)\right)}{\partial x} = \frac{\frac{\partial\left(x\sqrt[\gamma]{z^\alpha y^\beta}\right)}{\partial x}}{x\sqrt[\gamma]{z^\alpha y^\beta}} = \frac{1}{x} \\
\frac{\partial\left(ln\left(\sqrt[\gamma]{z^\alpha y^\beta x^\gamma}\right)\right)}{\partial y} &= \frac{\partial\left(ln(xz^{\frac{\alpha}{\gamma}}y^{\frac{\beta}{\gamma}})\right)}{\partial y} = \frac{\frac{\beta}{\gamma}xz^{\frac{\alpha}{\gamma}}y^{\frac{\beta-\gamma}{\gamma}}}{ln(xz^{\frac{\alpha}{\gamma}}y^{\frac{\beta}{\gamma}})} \\
\frac{\partial\left(ln\left(\sqrt[\gamma]{z^\alpha y^\beta x^\gamma}\right)\right)}{\partial z} &= \frac{\partial\left(ln(xz^{\frac{\alpha}{\gamma}}y^{\frac{\beta}{\gamma}})\right)}{\partial z} = \frac{\frac{\alpha}{\gamma}xz^{\frac{\alpha-\gamma}{\gamma}}y^{\frac{\beta}{\gamma}}}{ln(xz^{\frac{\alpha}{\gamma}}y^{\frac{\beta}{\gamma}})}
\end{aligned}$$

1.3

1.

$$\begin{aligned}
&(x-\mu)^T\Sigma^{-1}(x-\mu) + (\mu-\mu_0)^T S^{-1}(\mu-\mu_0) \\
=&(x^T-\mu^T)\Sigma^{-1}(x-\mu) + (\mu^T-\mu_0^T)S^{-1}(\mu-\mu_0) \\
=&x^T\Sigma^{-1}x - 2x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu + \mu^T S^{-1}\mu - 2\mu^T S^{-1}\mu_0 + \mu_0^T S^{-1}\mu_0
\end{aligned}$$

2.

$$
\begin{aligned}
&x^T\Sigma^{-1}x - \mu^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu + \mu^T S^{-1}\mu - \mu_0^T S^{-1}\mu - \mu^T S^{-1}\mu_0 + \mu_0^T S^{-1}\mu_0 \\
=&x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu - x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu + \mu^T S^{-1}\mu - \mu_0^T S^{-1}\mu - \mu_0^T S^{-1}\mu + \mu_0^T S^{-1}\mu_0 \\
=&x^T\Sigma^{-1}x + \mu_0^T S^{-1}\mu_0 - 2\mu^T S^{-1}\mu_0 + \left[-2x^T\Sigma^{-1} + \mu^T(\Sigma^{-1} + S^{-1})\right]\mu
\end{aligned}
$$

3.

$$
\begin{aligned}
&\frac{\partial\left(x^T\Sigma^{-1}x - 2x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu + \mu^T S^{-1}\mu - 2\mu^T S^{-1}\mu_0 + \mu_0^T S^{-1}\mu_0\right)}{\partial\mu} \\
=&-\frac{\partial}{\partial\mu}2x^T\Sigma^{-1}\mu + \frac{\partial}{\partial x}\mu^T(\Sigma^{-1} + S^{-1})\mu \\
=&(-2x^T\Sigma^{-1})^T + ((\Sigma^{-1} + S^{-1}) + (\Sigma^{-1} + S^{-1})^T)\mu \\
=&-2\Sigma^{-1}x^T + 2(\Sigma^{-1} + S^{-1})\mu \\
=&0 \\
&\therefore \mu = \frac{2\Sigma^{-1}x^T}{2(\Sigma^{-1} + S^{-1})} = \frac{\Sigma^{-1}x^T}{(\Sigma^{-1} + S^{-1})}
\end{aligned}
$$

2 Probqability theroy

2.1

1.
- the city where I am staying now, defined as **C** can take value 0(staying in Rotterdam), 1(staying in Amsterdam)
- the weather in the city you are staying, defined as **W** can take value 0(raining), 1(not raining)

2.

$$P(W = 1 \mid C = 0) = 1 - P(W = 0 \mid C = 0) = 1 - 0.75 = 0.25$$

3.

$$
\begin{aligned}
P(W = 0) &= P(W = 0, C = 0) + P(W = 0, C = 1) \\
&= P(W = 0|C = 0)P(C = 0) + P(W = 0|C = 1)P(C = 1) \\
&= (0.75 \times 0.2) + (0.5 \times 0.8) \\
&= 0.55
\end{aligned}
$$

4.

$$P(C = 1|W = 0) = \frac{0.5 \times 0.8}{(0.5 \times 0.8) + (0.75 \times 0.2)} = \frac{0.4}{0.55} = 0.\overline{72}$$

2.2

1.
- Being pregnant, defined as **C** can take value 0(non-pregnant), 1(pregnant)
- test result, defined as **T** can take value 0(negative result), 1(positive result)
- person conducts test, defined as **E** can take value 0(professional), 1(non-professional)

2.

$$
\begin{aligned}
GroupA : P(T = 1, E = 0) \times 2000 &= (P(T = 1, E = 0, C = 0) + P(T = 1, E = 0, C = 1)) \times 2000 \\
&= [P(T = 1, E = 0|C = 0)P(C = 0) + P(T = 1, E = 0|C = 1)P(C = 1)] \times 2000 \\
&= [(0.01\% \times 50\%) + ((100\% - 2.6\%) \times 50\%)] \times 2000 \\
&\simeq 975 \\
GroupB : P(T = 1, E = 1) \times 8000 &= (P(T = 1, E = 1, C = 0) + P(T = 1, E = 1, C = 1)) \times 8000 \\
&= [P(T = 1, E = 1|C = 0)P(C = 0) + P(T = 1, E = 1|C = 1)P(C = 1)] \times 8000 \\
&= [(0.01\% \times 50\%) + ((100\% - 25\%) \times 50\%)] \times 8000 \\
&\simeq 3000
\end{aligned}
$$

3.

$$
\begin{aligned}
GroupA : &[P(C=0,T=1,E=0)+P(C=1,T=0,E=0)]\times 2000\\
&=[P(T=1,E=0|C=0)P(C=0)+P(T=0,E=0|C=1)P(C=1)]\times 2000\\
&=[(0.01\%\times 50\%)+(2.6\%\times 50\%)]\times 2000\\
&\simeq 26\\
GroupB : &[P(C=0,T=1,E=1)+P(C=1,T=0,E=1)]\times 8000\\
&=[P(T=1,E=1|C=0)P(C=0)+P(T=0,E=1|C=1)P(C=1)]\times 8000\\
&=[(0.01\%\times 50\%)+(25\%\times 50\%)]\times 8000\\
&\simeq 1000
\end{aligned}
$$

2.3

1.

$$
\begin{aligned}
&p(\theta|D)=\frac{p(D|\theta)p(\theta)}{p(D)}\\
&prior : p(\theta), likelihood : p(D|\theta), evidence : p(D), posterir : p(\theta|D)
\end{aligned}
$$

2.

$$
\begin{aligned}
&p(D|\theta)=p(x_1,x_2,...,x_N|\mu,\sigma^2)=N(X|\mu,\sigma^2)\\
&p(\theta)=N(\mu|\mu_0,\sigma_0^2)\\
&p(\theta|D)=\frac{p(D|\theta)p(\theta)}{p(D)}=\frac{N(X|\mu,\sigma^2)N(\mu|\mu_0,\sigma_0^2)}{\sum(p(D|\theta)p(\theta))}=\frac{N(X|\mu,\sigma^2)N(\mu|\mu_0,\sigma_0^2)}{\int(N(X|\mu,\sigma^2)N(\mu|\mu_0,\sigma_0^2))d\mu}
\end{aligned}
$$