Machine Learning 1 - Homework 1

Pascal Mattia Esser

September 13th, 2017

1 Basic Linear Algebra and Derivatives

1.1 Matrix Operations

Assume for this question:

$$\boldsymbol{A} = \begin{bmatrix} -7 & 8 & 1 \\ -4 & 3 & 5 \\ 7 & 7 & -8 \end{bmatrix} \textbf{ and } \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} \tag{1}$$

1.1.1

Calculate $\boldsymbol{Ab}$

$$\boldsymbol{Ab} = \begin{bmatrix} -7*1+8*2+1*5 \\ -4*1+3*2+5*5 \\ 7*1+7*2-8*5 \end{bmatrix} = \begin{bmatrix} 14 \\ 27 \\ -19 \end{bmatrix} \tag{2}$$

1.1.2

Calculate $\boldsymbol{b}^T\boldsymbol{A}$

$$\boldsymbol{b}^T\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} -7 & 8 & 1 \\ -4 & 3 & 5 \\ 7 & 7 & -8 \end{bmatrix} \tag{3}$$

$$= \begin{bmatrix} 1*-7+2*-4+5*7 & 1*8+2*3+5*7 & 1*1+2*5-5*8 \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} 20 & 49 & -29 \end{bmatrix} \tag{5}$$

1.1.3

Compute the vector c for which $\boldsymbol{Ac} = \boldsymbol{b}$ through elimination. Goal:

$$\boldsymbol{D} = \left[\begin{array}{ccc|c} -7 & 8 & 1 & 1 \\ -4 & 3 & 5 & 2 \\ 7 & 7 & -8 & 5 \end{array}\right] \overset{\text{elementary matrix transformations}}{\rightsquigarrow} \left[\begin{array}{ccc|c} 1 & 0 & 0 & c_1 \\ 0 & 1 & 0 & c_2 \\ 0 & 0 & 1 & c_3 \end{array}\right] \tag{6}$$

or more general:

$$\left[\boldsymbol{A} \mid \boldsymbol{b}\right] \rightsquigarrow \left[\boldsymbol{I}_4 \mid \boldsymbol{c}\right] \tag{7}$$

Calculation: For easter reading the rows are marked with $I := \boldsymbol{d}_{1j}, II := \boldsymbol{d}_{2j}, III := \boldsymbol{d}_{3j}$ applied in the listed order.

$$\left[\begin{array}{ccc|c} -7 & 8 & 1 & 1 \\ -4 & 3 & 5 & 2 \\ 7 & 7 & -8 & 5 \end{array}\right] \tag{8}$$

$$\underset{\rightsquigarrow}{I+III \to III} \left[\begin{array}{ccc|c} -7 & 8 & 1 & 1 \\ -4 & 3 & 5 & 2 \\ 0 & 15 & -7 & 6 \end{array}\right] \tag{9}$$

$$\underset{\rightsquigarrow}{I*-4 \to I, II*7 \to II} \left[\begin{array}{ccc|c} 28 & -32 & -4 & -4 \\ -28 & 21 & 35 & 14 \\ 0 & 15 & -7 & 6 \end{array}\right] \tag{10}$$

$$\underset{\rightsquigarrow}{II+I \to II} \left[\begin{array}{ccc|c} 28 & -32 & -4 & -4 \\ 0 & -11 & 31 & 10 \\ 0 & 15 & -7 & 6 \end{array}\right] \tag{11}$$

$$\underset{\rightsquigarrow}{I*1/28 \to I, II*15 \to II, III*11 \to III} \left[\begin{array}{ccc|c} 1 & -8/7 & -1/7 & -1/7 \\ 0 & -165 & 465 & 150 \\ 0 & 165 & -77 & 66 \end{array}\right] \tag{12}$$

$$\underset{\rightsquigarrow}{III+II \to IIII} \left[\begin{array}{ccc|c} 1 & -8/7 & -1/7 & -1/7 \\ 0 & -165 & 465 & 150 \\ 0 & 0 & 388 & 216 \end{array}\right] \tag{13}$$

$$\underset{\rightsquigarrow}{II*-8/1155} \left[\begin{array}{ccc|c} 1 & -8/7 & -1/7 & -1/7 \\ 0 & 8/7 & -248/77 & -80/77 \\ 0 & 0 & 388 & 216 \end{array}\right] \tag{14}$$

$$\underset{\rightsquigarrow}{I+II \to I, II*7/8 \to II} \left[\begin{array}{ccc|c} 1 & 0 & -37/11 & -13/11 \\ 0 & 1 & -31/11 & -10/11 \\ 0 & 0 & 388 & 216 \end{array}\right] \tag{15}$$

$$\underset{\rightsquigarrow}{III*1/388 \to III} \left[\begin{array}{ccc|c} 1 & 0 & -37/11 & -13/11 \\ 0 & 1 & -31/11 & -10/11 \\ 0 & 0 & 1 & 54/97 \end{array}\right] \tag{16}$$

$$\underset{\rightsquigarrow}{II+31/11*III \to II, I+37/11*III \to III} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 67/97 \\ 0 & 1 & 0 & 64/97 \\ 0 & 0 & 1 & 54/97 \end{array}\right] \tag{17}$$

Therefore

$$\boldsymbol{c} = \begin{bmatrix} 67/97 \\ 64/97 \\ 54/97 \end{bmatrix} \tag{18}$$

1.1.4

Compute the vector c for which $\boldsymbol{Ac} = \boldsymbol{b}$ through elimination. Goal:

$$\left[\boldsymbol{A} \;\middle|\; \boldsymbol{I}_4\right] \rightsquigarrow \left[\boldsymbol{I}_4 \;\middle|\; \boldsymbol{A}^{-1}\right] \tag{19}$$

Calculation

$$\left[\begin{array}{ccc|ccc} -7 & 8 & 1 & 1 & 0 & 0 \\ -4 & 3 & 5 & 0 & 1 & 0 \\ 7 & 7 & -8 & 0 & 0 & 1 \end{array}\right] \tag{20}$$

$$\underset{\leadsto}{I+III\to III} \left[\begin{array}{ccc|ccc} -7 & 8 & 1 & 1 & 0 & 0 \\ -4 & 3 & 5 & 0 & 1 & 0 \\ 0 & 15 & -7 & 1 & 0 & 1 \end{array}\right] \tag{21}$$

$$\underset{\leadsto}{I*-4\to I, II*7\to II} \left[\begin{array}{ccc|ccc} 28 & -32 & -4 & -4 & 0 & 0 \\ -28 & 21 & 35 & 0 & 7 & 0 \\ 0 & 15 & -7 & 1 & 0 & 1 \end{array}\right] \tag{22}$$

$$\underset{\leadsto}{II+I\to II} \left[\begin{array}{ccc|ccc} 28 & -32 & -4 & -4 & 0 & 0 \\ 0 & -11 & 31 & -4 & 7 & 0 \\ 0 & 15 & -7 & 1 & 0 & 1 \end{array}\right] \tag{23}$$

$$\underset{\leadsto}{I*1/28\to I, II*15\to II, III*11\to III} \left[\begin{array}{ccc|ccc} 1 & -8/7 & -1/7 & -1/7 & 0 & 0 \\ 0 & -165 & 465 & -60 & 150 & 0 \\ 0 & 165 & -77 & 11 & 0 & 11 \end{array}\right] \tag{24}$$

$$\underset{\leadsto}{III+II\to IIII} \left[\begin{array}{ccc|ccc} 1 & -8/7 & -1/7 & -1/7 & 0 & 0 \\ 0 & -165 & 465 & -60 & 150 & 0 \\ 0 & 0 & 388 & -49 & 105 & 11 \end{array}\right] \tag{25}$$

$$\underset{\leadsto}{II*-8/1155} \left[\begin{array}{ccc|ccc} 1 & -8/7 & -1/7 & -1/7 & 0 & 0 \\ 0 & 8/7 & -248/77 & 32/77 & 8/11 & 0 \\ 0 & 0 & 388 & -49 & 150 & 11 \end{array}\right] \tag{26}$$

$$\underset{\leadsto}{I+II\to I, II*7/8\to II} \left[\begin{array}{ccc|ccc} 1 & 0 & -37/11 & 3/11 & -8/11 & 0 \\ 0 & 1 & -31/11 & 4/11 & -7/11 & 0 \\ 0 & 0 & 388 & -49 & 105 & 11 \end{array}\right] \tag{27}$$

$$\underset{\leadsto}{III*1/388\to III} \left[\begin{array}{ccc|ccc} 1 & 0 & -37/11 & 3/11 & -8/11 & 0 \\ 0 & 1 & -31/11 & 4/11 & -7/11 & 0 \\ 0 & 0 & 1 & -49/388 & 105/388 & 11/388 \end{array}\right] \tag{28}$$

$$\underset{\leadsto}{II+31/11*III\to II, I+37/11*III\to III} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -59/388 & 71/388 & 37/388 \\ 0 & 1 & 0 & 3/388 & 49/388 & 31/388 \\ 0 & 0 & 1 & -49/388 & 105/388 & 11/388 \end{array}\right] \tag{29}$$

therefore

$$\boldsymbol{A^{-1}} = \begin{bmatrix} -59/388 & 71/388 & 37/388 \\ 3/388 & 49/388 & 31/388 \\ -49/388 & 105/388 & 11/388 \end{bmatrix} \tag{30}$$

1.1.5

Show the special case: $\boldsymbol{A^{-1}b} = \boldsymbol{c}$

$$\boldsymbol{A^{-1}b} = \begin{bmatrix} -59/388 & 71/388 & 37/388 \\ 3/388 & 49/388 & 31/388 \\ -49/388 & 105/388 & 11/388 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} \tag{31}$$

$$= \begin{bmatrix} (-59/388) * 1 + (71/388) * 2 + (37/388) * 5 \\ (3/388) * 1 + (49/388) * 2 + (31/38)8 * 5 \\ (-49/388) * 1 + (105/388) * 2 + (11/388) * 5 \end{bmatrix} \tag{32}$$

$$= \begin{bmatrix} 67/97 \\ 64/97 \\ 54/97 \end{bmatrix} \tag{33}$$

General: if $\boldsymbol{Ac} = \boldsymbol{b}$ is solved for $\boldsymbol{c}$, than $\boldsymbol{A^{-1}b} = \boldsymbol{c}$ holds.

Proof. Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ be nonsingular, than $\boldsymbol{A}^{-1}$, the inverse, is defined as the $n \times n$ Matrix, so that

$$\boldsymbol{AA}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I} \tag{34}$$

holds, where $\boldsymbol{I}$ is the identity matrix. Furthermore $\boldsymbol{Ac} = \boldsymbol{b}$, with $\boldsymbol{b} \in \mathbb{C}^{n \times m}$ and $\boldsymbol{c} \in \mathbb{C}^{n \times m}$ is given. now show $\boldsymbol{A^{-1}b} = \boldsymbol{c}$ holds in general:

$$\boldsymbol{Ac} = \boldsymbol{b} \Leftrightarrow \boldsymbol{AA}^{-1}\boldsymbol{c} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{35}$$

$$\Leftrightarrow \boldsymbol{A}^{-1}\boldsymbol{Ac} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{36}$$

$$\Leftrightarrow \boldsymbol{Ic} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{37}$$

$$\Leftrightarrow \boldsymbol{c} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{38}$$

$$\tag{39}$$

□

1.2 Calculating Derivatives

1.2.1

$$\frac{d}{dx}\left(\left(\frac{2}{x^2} + x^{-7} + x^3\right)^2\right) \tag{40}$$

$$= 2\left(\frac{2}{x^2} + x^{-7} + x^3\right)\frac{d}{dx}\left(\frac{2}{x^2} + x^{-7} + x^3\right) \tag{41}$$

$$= 2\left(\frac{2}{x^2} + x^{-7} + x^3\right)\left(\frac{4}{x^3} - \frac{7}{x^{-8}} + 3x^2\right) \tag{42}$$

$$= 6x^5 - \frac{24}{x^5} - \frac{36}{x^{10}} - \frac{14}{x^{15}} + 4 \tag{43}$$

1.2.2

$$\frac{d}{dx}\left(x^2\sqrt{e^{-\sqrt[3]{x}}}\right) \tag{44}$$
$$= \frac{d}{dx}\left(x^2\right)\sqrt{e^{-\sqrt[3]{x}}} + \frac{d}{dx}\left(\sqrt{e^{-\sqrt[3]{x}}}\right)x^2 \tag{45}$$
$$\frac{d}{dx}\left(x^2\right) = 2x \tag{46}$$
$$\text{let } s = e^{-\sqrt[3]{x}} \tag{47}$$
$$\frac{d}{dx}\left(\sqrt{e^{-\sqrt[3]{x}}}\right) \tag{48}$$
$$= \frac{d}{ds}\left(\sqrt{s}\right)\frac{d}{dx}\left(e^{-\sqrt[3]{x}}\right) \tag{49}$$
$$= \left(\frac{1}{2\sqrt{s}}\right)\frac{d}{dx}\left(e^{-\sqrt[3]{x}}\right) \tag{50}$$
$$= \left(\frac{1}{2\sqrt{s}}\right)e^{-\sqrt[3]{x}}\left(-\frac{1}{3x^{2/3}}\right) \tag{51}$$
$$\text{substitute s back} \tag{52}$$
$$= \left(\frac{1}{2\sqrt{e^{-\sqrt[3]{x}}}}\right)e^{-\sqrt[3]{x}}\left(-\frac{1}{3x^{2/3}}\right) \tag{53}$$
$$= \left(\frac{1}{2e^{\frac{-\sqrt[3]{x}}{2}}}\right)\left(-\frac{e^{-\sqrt[3]{x}}}{3x^{2/3}}\right) \tag{54}$$
$$= -\frac{1}{6e^{\frac{\sqrt[3]{x}}{2}}x^{2/3}} \tag{55}$$
$$\text{putting the parts back into Equation 45 gives:} \tag{56}$$
$$2x\sqrt{e^{-\sqrt[3]{x}}} + -\frac{1}{6e^{\frac{\sqrt[3]{x}}{2}}x^{2/3}}x^2 \tag{57}$$
$$= 2xe^{\frac{-\sqrt[3]{x}}{2}} - \frac{1}{6e^{\frac{\sqrt[3]{x}}{2}}x^{2/3}}x^2 \tag{58}$$
$$= 2xe^{\frac{-\sqrt[3]{x}}{2}} - \frac{x^{4/3}}{6e^{\frac{\sqrt[3]{x}}{2}}} \tag{59}$$

1.2.3

$$\frac{d}{dx}\left(x + ln(x)\right) = 1 + \frac{1}{x} \tag{60}$$

1.2.4

$$\frac{d}{dx}\left(xln(\sqrt{x})\right) \tag{61}$$
$$= \frac{d}{dx}\left(x)ln(\sqrt{x})\right) + x\frac{d}{dx}\left(ln(\sqrt{x})\right) \tag{62}$$
$$= ln(\sqrt{x}) + x\frac{1}{\sqrt{x}}\frac{1}{2\sqrt{x}} \tag{63}$$
$$= ln(\sqrt{x}) + 1/2 \tag{64}$$

1.2.5

$$\frac{d}{dx} 6\left(x^2-1\right) sin(x) \tag{65}$$

$$= 6\left(\frac{d}{dx}\left(x^2-1\right) sin(x) + \left(x^2-1\right)\frac{d}{dx} sin(x)\right) \tag{66}$$

$$= 6\left(2xsin(x) + cos(x)(x^2-1)\right) \tag{67}$$

1.2.6

$$\text{in the following assume for substituetion:} \tag{68}$$

$$s = \sqrt[3]{\frac{e^{3x}}{1+e^{3x}}}, s' = \frac{e^{3x}}{1+e^{3x}} \tag{69}$$

$$\frac{d}{dx}\left(ln\left(\sqrt[3]{\frac{e^{3x}}{1+e^{3x}}}\right)\right) \tag{70}$$

$$= \frac{d}{ds}\left(ln(s)\right)\frac{d}{dx}\left(\sqrt[3]{\frac{e^{3x}}{1+e^{3x}}}\right) \tag{71}$$

$$\frac{d}{ds}\left(ln(s)\right) = \frac{1}{s} \tag{72}$$

$$\frac{d}{dx}\left(\sqrt[3]{\frac{e^{3x}}{1+e^{3x}}}\right) \tag{73}$$

$$= \frac{d}{ds}\left(\sqrt[3]{s'}\right)\frac{d}{dx}\left(\frac{e^{3x}}{1+e^{3x}}\right) \tag{74}$$

$$\frac{d}{ds}\left(\sqrt[3]{s'}\right) = \frac{1}{3s'^{2/3}} \tag{75}$$

$$\frac{d}{dx}\left(\frac{e^{3x}}{1+e^{3x}}\right) \tag{76}$$

$$= \frac{e^{3x}3(1+e^{3x}) - 3^{3x}e^{3x}}{(1+e^{3x})^2} \tag{77}$$

$$= \frac{3e^{3x}}{(1+e^3x)^2} \tag{78}$$

$$\text{substitute s' back} \tag{79}$$

$$\frac{1}{3\left(\frac{e^{3x}}{1+e^{3x}}\right)^{2/3}}\frac{3e^{3x}}{(1+e^3x)^2} \tag{80}$$

$$= \frac{e^{3x}}{\frac{e^{3x}}{(1+e^{3x})^{2/3}}\left(1+e^{3x}\right)^2} \tag{81}$$

$$= \frac{e^{3x}}{e^{2x}\left(1+e^{3x}\right)^{4/3}} \tag{82}$$

$$= \frac{e^{x}}{\left(1+e^{3x}\right)^{4/3}} \tag{83}$$

$$\text{substitute s back} \tag{84}$$

$$\frac{1}{\sqrt[3]{\frac{e^{3x}}{1+e^{3x}}}}\frac{e^x}{\left(1+e^{3x}\right)^{4/3}} \tag{85}$$

$$= \frac{\sqrt[3]{1+e^{3x}}}{e^x}\frac{e^x}{\left(1+e^{3x}\right)^{4/3}} \tag{86}$$

$$= \frac{1}{1+e^{3x}} \tag{87}$$

1.2.7

The partial derivative $\frac{\partial f}{\partial x}$ of a function f, of several variables gives the derivative with respect to one variable (x) while keeping the other variables constant.

1.2.8

$$\text{set } s = (y - exp(x^{-1}) - sin(zx^2) \tag{88}$$
$$\frac{\partial}{\partial x}\left(2ln(y - exp(x^{-1}) - sin(zx^2))\right) \tag{89}$$
$$= 2\frac{\partial}{\partial s}\left(ln(s)\right)\frac{\partial}{\partial x}\left(y - exp(x^{-1}) - sin(zx^2)\right) \tag{90}$$
$$= 2\frac{1}{s}\left(exp(x^{-1})\left(-\frac{1}{x^2}\right) + cos(zx^2)2xz\right) \tag{91}$$
$$= \frac{1}{(y - exp(x^{-1}) - sin(zx^2)}\left(\left(-\frac{exp(x^{-1})}{x^2}\right) + cos(zx^2)2xz\right) \tag{92}$$
$$= \frac{2\left(exp(x^{-1}) - cos(zx^2)2x^3z\right)}{x^2(y - exp(x^{-1}) - sin(zx^2)} \tag{93}$$

$$\text{set } s = (y - exp(x^{-1}) - sin(zx^2) \tag{94}$$
$$\frac{\partial}{\partial y}\left(2ln(y - exp(x^{-1}) - sin(zx^2))\right) \tag{95}$$
$$= 2\frac{\partial}{\partial s}\left(ln(s)\right)\frac{\partial}{\partial y}\left(y - exp(x^{-1}) - sin(zx^2)\right) \tag{96}$$
$$= 2\frac{1}{s}\frac{\partial}{\partial y}\left(y - exp(x^{-1}) - sin(zx^2)\right) \tag{97}$$
$$= 2\frac{1}{s}1 \tag{98}$$
$$= \frac{2}{(y - exp(x^{-1}) - sin(zx^2)} \tag{99}$$

$$\text{set } s = (y - exp(x^{-1}) - sin(zx^2) \tag{100}$$
$$\frac{\partial}{\partial z}\left(2ln(y - exp(x^{-1}) - sin(zx^2))\right) \tag{101}$$
$$= 2\frac{\partial}{\partial s}\left(ln(s)\right)\frac{\partial}{\partial z}\left(y - exp(x^{-1}) - sin(zx^2)\right) \tag{102}$$
$$= 2\frac{1}{s}\left(-x^2cos(x^2z)\right) \tag{103}$$
$$= -\frac{2x^2cos(x^2z)}{(y - exp(x^{-1}) - sin(zx^2)} \tag{104}$$

1.2.9

$$\frac{\partial}{\partial x}\left(ln\left(\left(z^\alpha y^\beta x^\gamma\right)^{1/\gamma}\right)\right) \tag{105}$$
$$= \frac{\partial}{\partial x}\left(\frac{\alpha ln(z)}{\gamma} + \frac{\beta ln(y)}{\gamma} + ln(x)\right) \tag{106}$$
$$= \frac{1}{x} \tag{107}$$

$$\frac{\partial}{\partial y}\left(ln\left(\left(z^{\alpha}y^{\beta}x^{\gamma}\right)^{1/\gamma}\right)\right) \tag{108}$$
$$= \frac{\partial}{\partial y}\left(\frac{\alpha ln(z)}{\gamma}+\frac{\beta ln(y)}{\gamma}+ln(x)\right) \tag{109}$$
$$= \frac{\beta}{\gamma y} \tag{110}$$

WLOG we can interchange y and z and use $\frac{\partial}{\partial z}$ to get $\frac{\partial}{\partial y}$:

$$\frac{\partial}{\partial z} = \frac{\alpha}{\gamma z} \tag{111}$$

1.3 Manipulating Vectors and Matrices and finding Derivatives

1.3.1

Starting with the following expression:

$$f = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) + (\boldsymbol{\mu}-\boldsymbol{\mu}_0)^T\boldsymbol{S}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) \tag{112}$$

expand the formula to:

$$f = \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{S}^{-1}\boldsymbol{\mu} \tag{113}$$
$$-\boldsymbol{\mu_0}^T\boldsymbol{S}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^T\boldsymbol{S}^{-1}\boldsymbol{\mu_0} - \boldsymbol{\mu_0}^T\boldsymbol{S}^{-1}\boldsymbol{\mu_0} \tag{114}$$

1.3.2

because of:

$$\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \tag{115}$$
$$\boldsymbol{\mu_0}^T\boldsymbol{S}^{-1}\boldsymbol{\mu} = \boldsymbol{\mu}^T\boldsymbol{S}^{-1}\boldsymbol{\mu_0} \tag{116}$$

we can write:

$$f = \boldsymbol{\mu}^T(\boldsymbol{\Sigma}^{-1}+\boldsymbol{S}^{-1})\boldsymbol{\mu} \qquad \text{depends quadratic on } \boldsymbol{\mu} \tag{117}$$
$$-2\boldsymbol{\mu}^T(\boldsymbol{\Sigma}^{-1}\boldsymbol{x}+\boldsymbol{S}^{-1}\boldsymbol{\mu_0}) \qquad \text{depends liniar on } \boldsymbol{\mu} \tag{118}$$
$$-\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{\mu_0}^T\boldsymbol{S}^{-1}\boldsymbol{\mu_0} \qquad \text{does not depend on } \boldsymbol{\mu} \tag{119}$$

Substitute:

$$\boldsymbol{A} = \boldsymbol{\Sigma}^{-1}+\boldsymbol{S}^{-1} \tag{120}$$

$$\boldsymbol{B} = \boldsymbol{\Sigma}^{-1}\boldsymbol{x}+\boldsymbol{S}^{-1}\boldsymbol{\mu_0} \tag{121}$$

to get a function of the from:

$$f' = \boldsymbol{\mu}^T \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{B}^T \boldsymbol{\mu} \tag{122}$$

Now for this function, the gradient can be calculated as:

$$\nabla_\mu f' = \frac{\partial f}{\partial x} = (\boldsymbol{A} + \boldsymbol{A}^T)\boldsymbol{\mu} + \boldsymbol{B}^T \tag{123}$$

Apply this on f under consideration of the symmetry of the matrices $\boldsymbol{A}$ and $\boldsymbol{S}$:

$$\nabla_\mu f = ((\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}) + (\boldsymbol{\Sigma}^{-1} - \boldsymbol{S}^{-1})^T)\boldsymbol{\mu} - 2(\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{S}^{-1}\boldsymbol{\mu_0}) \tag{124}$$

$$\nabla_\mu f = 2(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1})\boldsymbol{\mu} - 2(\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{S}^{-1}\boldsymbol{\mu_0}) \tag{125}$$

setting

$$\nabla_\mu f = 0 \tag{126}$$

solving for $\boldsymbol{\mu}$ under the assumption, that $(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1})$ is invertible:

$$\boldsymbol{\mu} = (\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{S}^{-1}\boldsymbol{\mu_0})(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1})^{-1} \tag{127}$$

2 Probability Theory

in the following use $\perp$ for 'false' and $\top$ for 'true'.

2.1 Weather in Amsterdam

2.1.1 Defining the variables

Defining the variables: rain: $r = \{\perp, \top\}$, location: $l = \{A, R\}$, with A for Amsterdam and R for Rotterdam. The probabilities that are given:

$$p(r = \top | l = A) = 0.5 \tag{128}$$

$$p(r = \top | l = R) = 0.75 \tag{129}$$

$$p(l = A) = 0.8 \tag{130}$$

2.1.2 Probability for no rain Rotterdam

probability for no rain Rotterdam

$$p(r = \perp | l = R) = 1 - p(r = \top | l = R) = 0.25 \tag{131}$$

2.1.3 Rain at your current location

let $p(k)$ be the probability for rain at your current location

$$
\begin{aligned}
& \quad && (132)\\
& \text{[using the sum rule]} && (133)\\
p(k) &= \sum_{i=\{A,R\}} p(\top, l=i) && (134)\\
& \text{[using the product rule]} && (135)\\
&= \sum_{i=\{A,R\}} p(\top|l=i)p(l=i) && (136)\\
&= p(\top|A)p(A) + p(\top|R)p(R) && (137)\\
&= 0.5 * 0.8 + 0.75 * 0.2 && (138)\\
&= 0.55 && (139)
\end{aligned}
$$

2.1.4 It rains, are you in Amsterdam?

It rains, what is the probability, that you are in Amsterdam? Using Bayes:

$$
\begin{aligned}
p(l=A|r=\top) &= \frac{p(\top|A)p(A)}{p(r=\top)} && (140)\\
&= \frac{p(\top|A)p(A)}{p(k)} && (141)\\
&= \frac{0.5 * 0.8}{0.55} && (142)\\
&= 8/11 \sim 0.72 && (143)
\end{aligned}
$$

2.2 Pregnancy Test

2.2.1 Defining all the Variables

Defining the variables: test: $t = \{\perp, \top\}$ pregnant: $r = \{\perp, \top\}$ tested by an expert: $e = \{\perp, \top\}$ The probabilities that are given:

$$
\begin{aligned}
& p(r=\top) = 0.5 && (144)\\
& p(t=\top, r=\perp|e=\top) = 0.026 && (145)\\
& p(t=\top, r=\perp|e=\perp) = 0.25 && (146)\\
& p(t=\perp, r=\top) = 0.0001 && (147)
\end{aligned}
$$

number of test subjects in each group:

$$
\begin{aligned}
& && (148)\\
n_{e=\top} &= 2000 && (149)\\
n_{e=\perp} &= 8000 && (150)
\end{aligned}
$$

2.2.2 Calculating how many woman are tested pregnant

let n be the number of subjects in the respective groups

$$\begin{aligned}
\mathbb{E}[t = \top]_e = & \qquad (151)\\
& [\text{wrong negative}] \qquad (152)\\
& n_e * p(r = \top)p(t = \perp, r = \top) \qquad (153)\\
& [\text{wrong positive}] \qquad (154)\\
& - n_e * p(r = \top)p(t = \top, r = \perp|e) \qquad (155)\\
& [\text{expected value}] \qquad (156)\\
& + n_e * p(r = \top) \qquad (157)
\end{aligned}$$

putting in the numbers for $e = \top$ gives $\mathbb{E}[t = \top]_{e=\top} = 974.1 \sim 974$ and $e = \top$ gives $\mathbb{E}[t = \top]_{e=\perp} = 3000.4 \sim 3000$

2.2.3 How many false results are there?

Use subsubsection 2.2.2 and subtract the number of expected values so only the wrong positives and wrong negatives remain.

$$\begin{aligned}
\mathbb{E}[t = \top|r = \perp \vee t = \perp|r = \top]_e = & n_e * p(r = \top)p(t = \perp, r = \top) \qquad (158)\\
& - n_e * p(r = \top)p(t = \top, r = \perp|e) \qquad (159)
\end{aligned}$$

putting in the numbers for $e = \top$ give $\mathbb{E}[t = \top|r = \perp \vee t = \perp|r = \top]_e = \top = 26.1 \sim 26$ and $e = \top$ gives $\mathbb{E}[t = \top|r = \perp \vee t = \perp|r = \top]_e = \perp = 1000.4 \sim 1000$

2.3 Distribution Parameters

2.3.1 General expression for a posterior distribution

The posterior distribution can be written as:

$$p(\Theta|D) = \frac{p(D|\Theta)p(\Theta)}{p(D)} \qquad (160)$$

Where:

$$p(\Theta) \qquad (161)$$

is the prior distribution

$$p(D|\Theta) \qquad (162)$$

is the likelihood

$$p(D) \qquad (163)$$

is the evidence

2.3.2 Write as function of the given variables

Starting again with:

$$p(\Theta|D) = \frac{p(D|\Theta)p(\Theta)}{p(D)} \qquad (164)$$

and using:

$$p(\Theta) = p(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2) \tag{165}$$

$$p(D|\Theta) = p(D|\mu) = \prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \sigma^2) \tag{166}$$

$$p(D) = p(D) = \int p(D|\mu)p(\mu)d\mu \tag{167}$$

where Θ is replaced by μ because this is give parameter. Assuming a known σ_0^2, gives:

$$p(\mu|D) = \frac{\prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \sigma^2)\mathcal{N}(\mu|\mu_0, \sigma_0^2)}{\int p(D|\mu)p(\mu)d\mu} \tag{168}$$

and replacing the expressions in the integral in the denominator gives:

$$p(\mu|D) = \frac{\prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \sigma^2)\mathcal{N}(\mu|\mu_0, \sigma_0^2)}{\int \prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \sigma^2)\mathcal{N}(\mu|\mu_0, \sigma_0^2)d\mu} \tag{169}$$