TARUN KRISHNA
11593040
krishnatarun7@gmail.com

Homework Assignment 1
Machine Learning 1, 17/18

2017-09-16

1 First Exercise

1.1 First Subtask

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

1. Compute **Ab**

$$\mathbf{Ab} = \begin{bmatrix} -7 & 8 & 1 \\ -4 & 3 & 5 \\ 7 & 7 & -8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

following Row * Cols convention in general.

$$\mathbf{Ab} = \begin{bmatrix} -7+16+5 \\ -4+6+25 \\ 7+14-40 \end{bmatrix}$$
$$= \begin{bmatrix} 14 \\ 27 \\ -19 \end{bmatrix}$$

2. Compute $b^T A$

$$b^T A = \begin{bmatrix} 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} 7 & 8 & 1 \\ -4 & 3 & 5 \\ 7 & 7 & -8 \end{bmatrix}$$
$$= \begin{bmatrix} -7-8+35 & 8+6+35 & 1+10-40 \end{bmatrix}$$
$$= \begin{bmatrix} 20 & 49 & -29 \end{bmatrix}$$

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

$$c = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$\begin{bmatrix} -7 & 8 & 1 \\ -4 & 3 & 5 \\ 7 & 7 & -8 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

Create an augmented matrix $[A|b]$

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

Applying elementary row operations to get **c**. Will convert the matrix in Reduced Roe Echelon Form.
Step 1. $\Rightarrow R_1 \rightarrow -7R_1$

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

Step 2. $\Rightarrow R_2 \to R_2 - (-4)R_1$

$$\left[\begin{array}{ccc|c} 1 & 8/11 & -1/7 & -1/7 \\ 0 & -11/7 & 31/7 & 10/7 \\ 7 & 7 & -8 & 5 \end{array}\right]$$

Step 3. $\Rightarrow R_3 \to R_3 - 7R_1$

$$\left[\begin{array}{ccc|c} 1 & -8/7 & -1/7 & \frac{-}{1}/7 \\ 0 & -11/7 & 31/7 & 10/7 \\ 0 & 15 & -7 & 6 \end{array}\right]$$

Step 4. $\Rightarrow R_2 \to -7/11R_2$

$$\left[\begin{array}{ccc|c} 1 & -8/7 & -1/7 & -1/7 \\ 0 & 1 & -31/11 & -10/11 \\ 0 & 15 & -7 & 6 \end{array}\right]$$

Step 5 $\Rightarrow R_1 \to R_1 - (-8/7)R_2$

$$\left[\begin{array}{ccc|c} 1 & 0 & -37/11 & -13/11 \\ 0 & 1 & -31/11 & -10/11 \\ 0 & 15 & -7 & 6 \end{array}\right]$$

Step 6.$\Rightarrow R_3 \to R_3 - 15R_1$

$$\left[\begin{array}{ccc|c} 1 & 0 & -37/11 & -13/11 \\ 0 & 1 & -31/11 & -10/11 \\ 0 & 0 & 388/11 & 216/11 \end{array}\right]$$

Step 7. $\Rightarrow R_3 \to 11/388R_3$

$$\left[\begin{array}{ccc|c} 1 & 0 & -37/11 & -13/11 \\ 0 & 1 & -31/11 & -10/11 \\ 0 & 0 & 1 & 54/97 \end{array}\right]$$

Step 7.$\Rightarrow R_1 \to R_1 - (-37/11)R_3$ Step 8.$\Rightarrow R_2 \to R_2 - (-31/11)R_3$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 67/97 \\ 0 & 1 & 0 & 64/97 \\ 0 & 0 & 1 & 54/97 \end{array}\right]$$

Solution comes to be **x=67/97,y =64/97, z=54/97**

4. Compute the inverse of A

Will Use gauss-Jordan method of matrix inverse

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

Using the same Steps as in previous question. With same steps I mean elemntary operation that are done in solving $Ac = b$ had fetched us identity ,matrix(A) at the end.Using exactly same Row operations again, will get:

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

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

5.

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

Well, the multiplication satisfies the fact $A^{-1}b = c$. In general we can also prove.Let us suppose that

$$\begin{aligned} Ac &= b \\ A^{-1}Ac &= A^{-1}b \\ c &= A^{-1}b \end{aligned}$$

1.2 Second Subtask

Find the derivatives for the following function with respect to x:
1. $f(x) = (\frac{2}{x^2} + x^-7 + x^3)^2$

$$\begin{aligned} \frac{df(x)}{dx} &= \frac{d(\frac{2}{x^2} + x^{-7} + x^3)^2}{dx} \\ &= 2\left(3x^2 - \frac{4}{x^3} - \frac{7}{x^8}\right)\left(x^3 + \frac{2}{x^2} + \frac{1}{x^7}\right) \\ &= \frac{6x^{20} + 4x^{15} - 24x^{10} - 36x^5 - 14}{x^{15}} \end{aligned}$$

2. $f(x) = x^2\sqrt{e^{-\sqrt[3]{x}}}$

$$\frac{df(x)}{dx} = \frac{d(x^2\sqrt{e^{-\sqrt[3]{x}}})}{dx}$$

Using product rule $(fg)' = f'g + fg'$

$$\begin{aligned} &= 2x\mathrm{e}^{-\frac{\sqrt[3]{x}}{2}} - \frac{x^{\frac{4}{3}}\mathrm{e}^{-\frac{\sqrt[3]{x}}{2}}}{6} \\ &= \frac{\left(12x - x^{\frac{4}{3}}\right)\mathrm{e}^{-\frac{\sqrt[3]{x}}{2}}}{6} \end{aligned}$$

3. $f(x) = x + ln(x)$

$$\begin{aligned} \frac{df(x)}{dx} &= \frac{d[x + ln(x))]}{dx} \\ &= \frac{d[x]}{dx} +] + \frac{d[ln(x)]}{dx} \\ &= 1 + \frac{1}{x} \end{aligned}$$

4. $f(x) = xln(\sqrt{x})$

function can be re-written as :

$$\begin{aligned} f(x) &= \frac{xln(x)}{2} \\ \frac{df(x)}{dx} &= \frac{dx[ln(x)]}{2dx} \end{aligned}$$

Using product rule as mentioned above

$$\begin{aligned} &= \frac{1}{2} * \frac{dx}{dx} ln(x) + \frac{1}{2} * x \frac{dln(x)}{dx} \\ &= \frac{1}{2}(ln(x) + x\frac{1}{x}) \\ &= \frac{ln(x) + 1}{2} \end{aligned}$$

5. $f(x) = 6(x^2 - 1)sin(x)$

$$\begin{aligned} \frac{df(x)}{dx} &= \frac{d[6(x^2 - 1)sin(x)]}{dx} \\ &= 6(2x)sin(x) + 6(x^2 - 1)cox(x) \\ &= 12xsin(x) + 6(x^2 - 1)cox(x) \end{aligned}$$

6. $f(x) = ln(\sqrt[3]{\frac{e^{3x}}{1+e^{3x}}})$

function can be further simplified and re-written as :

$$\begin{aligned} f(x) &= xln(e) - \frac{1}{3}(1 + e^{3x}) \\ \frac{df(x)}{dx} &= 1 - \frac{1}{3} * \frac{3e^{3x}}{1 + e^{3x}} \\ &= \frac{(1 + e^{3x}) - e^{3x}}{1 + e^{3x}} \\ &= \frac{1}{1 + e^{3x}} \end{aligned}$$

7. For a multivariable function, like $f(x, y) = x^2y$ computing partial derivatives looks something like this:

$$\begin{aligned} \frac{\partial f(x,y)}{\partial x} &= \underbrace{\frac{\partial x^2 y}{\partial x}}_{\text{Treat y as constant}} = 2xy \\ \frac{\partial f(x,y)}{\partial y} &= \underbrace{\frac{\partial x^2 y}{\partial y}}_{\text{Treat x as constant}} = x^2 * 1 \end{aligned}$$

The partial derivative generalizes the notion of the derivative to higher dimensions. A partial derivative of a multivariable function is a derivative with respect to one variable with all other variables held constant. In order to find maxima and minima ,the story is very similar for multivariable functions. When the function is continuous and differentiable, all the partial derivatives will be **0** at a local maximum or minimum point.

$$\begin{aligned} \underbrace{f_x(x, y, z, ...)}_{\text{Partial with respect to } x} &= 0 \\ \underbrace{f_y(x, y, z, ...)}_{\text{Partial with respect to } y} &= 0 \\ \underbrace{f_z(x, y, z, ...)}_{\text{Partial with respect to } z} &= 0 \end{aligned}$$

And so on for rest of other variables as well.
Saying that all the partial derivatives are zero at a point is the same as saying the **gradient** at that point is the zero vector:

$$\nabla(f(x,y,z,....)) = \begin{bmatrix} f_x(x,y,z,...) \\ f_y(x,y,z,...) \\ f_z(x,y,z,...) \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

8. $f(x,y,z) = 2ln(y - e^{x^-1} - sin(zx^2))$

$$\begin{aligned} \frac{\partial f(x,y,z)}{\partial x} &= 2\frac{\partial(ln(y - e^{x^-1} - sin(zx^2)))}{\partial x} \\ &= 2\frac{e^{x^-1} * (\frac{-1}{x^2}) - cos(zx^2) * 2xz}{y - e^{x-1} - sin(zx^2)} \\ &= 2(\frac{\frac{e^{x^-1}}{x^2} - 2xzcos(zx^2)}{y - e^{x-1} - sin(zx^2)}) \end{aligned}$$

$$\begin{aligned} \frac{\partial f(x,y,z)}{\partial y} &= 2\frac{\partial(ln(y - e^{x^-1} - sin(zx^2)))}{\partial y} \\ &= \frac{2}{y - e^{x-1} - sin(zx^2)} \end{aligned}$$

$$\begin{aligned} \frac{\partial f(x,y,z)}{\partial z} &= 2\frac{\partial(ln(y - e^{x^-1} - sin(zx^2)))}{\partial z} \\ &= 2\frac{x^2 cos(zx^2)}{y - e^{x-1} - sin(zx^2)} \end{aligned}$$

9. $f(x,y,z) = ln(\sqrt[\gamma]{z^\alpha y^\beta x^\gamma})$
function can be further simplified and can be re-written as:

$$\begin{aligned} f(x,y,z) &= \frac{ln(z^\alpha y^\beta x^\gamma)}{\gamma} \\ &= \frac{ln(z^\alpha) + ln(y^\beta) + ln(x^\gamma)}{\gamma} \\ &= \frac{\alpha ln(z) + \beta ln(y) + \gamma ln(x)}{\gamma} \\ \frac{\partial f(x,y,z)}{\partial x} &= \frac{\gamma}{x\gamma} \\ &= \frac{1}{x} \\ \frac{\partial f(x,y,z)}{\partial y} &= \frac{\beta}{y\gamma} \\ \frac{\partial f(x,y,z)}{\partial z} &= \frac{\alpha}{z\gamma} \end{aligned}$$

1.3 Third Subtask

1. Given the following expression:

$$(x-\mu)^T\Sigma^{-1}(x-\mu) + (\mu-\mu_o)^T S^{-1}(\mu-\mu_o)$$

Given a Symmetric matrix A, $A = A^T$.
Solving and expanding the terms:

$$(x-\mu)^T\Sigma^{-1T}(x-\mu) + (\mu-\mu_o)^T S^{-1T}(\mu-\mu_o)$$
$$(\Sigma^{-1}(x-\mu))^T(x-\mu) + (S^{-1}(\mu-\mu_o))^T(\mu-\mu_o)$$
$$(\Sigma^{-1}x-\Sigma^{-1}\mu)^T(x-\mu) + (S^{-1}\mu-S^{-1}\mu_o)^T(\mu-\mu_o)$$

Expanding and multiplying.

$$x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu + \mu^T S^{-1}\mu - \mu^T S^{-1}\mu_o - \mu_o^T S^{-1}\mu + \mu_o^T S^{-1}\mu_o$$

$$x^T\Sigma^{-1}x - 2\mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu + \mu^T S^{-1}\mu - 2\mu^T S^{-1}\mu_o + \mu_o^T S^{-1}\mu_o \tag{1}$$

equation can be further simplified into:

$$x^T\Sigma^{-1}x + \mu_o^T S^{-1}\mu_o + \mu^T(\Sigma^{-1}+S^{-1})\mu - 2\mu^T(\Sigma^{-1}x + S^{-1}\mu_o) \tag{2}$$

2. Collecting terms depending on μ, can easily seen in equation (2), which can be again modified as:

$$x^T\Sigma^{-1}x + \mu_o^T S^{-1}\mu_o + \underbrace{\mu^T(\Sigma^{-1}+S^{-1})\mu - 2\mu^T(\Sigma^{-1}x + S^{-1}\mu_o)}_{\text{depends on } \mu}$$

3. For derivative considering only term depending on μ.so derivative can be written as:

$$2\mu^T(\Sigma^{-1}+S^{-1}) - 2(\Sigma^{-1}x + S^{-1}\mu_o)^T = 0$$
$$\frac{(\Sigma^{-1}x + S^{-1}\mu_o)^T}{(\Sigma^{-1}+S^{-1})} = \mu^T$$

taking transpose both the sides

$$\frac{(\Sigma^{-1}x + S^{-1}\mu_o)}{(\Sigma^{-1}+S^{-1})} = \mu$$

2 Second Exercise

Probability Theory

2.1 First Subtask

1. There are two random variable. First is Rain denoted by r and second Location, denoted by L.

$$r = \{0,1\} \Rightarrow 0 : \text{not raining, } 1 : \text{raining}$$
$$L = \{A, R\} \Rightarrow \text{ A:Amsterdam, P:Rotterdam}$$

2.Probability that it does not rain when you are in Rotterdam.

$$\begin{aligned} p(r=0|L=R) &= 1-0.75 \\ &= 0.25 \end{aligned}$$

3.Probability that it rains at your current location.

$$\begin{aligned} p(r=1) &= p(r=1|L=A)p(L=A) + p(r=1|L=R)p(L=R) \\ p(r=1) &= 0.8*0.5 + 0.75*0.2 \\ &= 0.55 \end{aligned}$$

4. Probaility of you are in Amsterdama given that it's raining i.e $p(L=A|r=1)$.
Using Bayes's rule:

$$\begin{aligned} p(L=A|r=1) &= \frac{p(r=1|L=A)*p(L=A)}{p(r=1)} \\ &= \frac{0.5*0.8}{.55} \\ &= \frac{8}{11} \end{aligned}$$

2.2 Second Subtask

1. Considering two Random variables Test(T) and Pregnant(G) defined as below:

$$
\begin{aligned}
T &= 0,1 \Rightarrow 0 : \text{-ve test, } 1 : \text{+ve test} \\
G &= 0,1 \Rightarrow 0 : \text{not pregnant, } 1 : \text{pregnant}
\end{aligned}
$$

2.Pregnant women in each group *A* and *B* respectively is sum of women who are actually pregnant and women who were falsely identified as pregnant.

$$p(G = 1) = 1/2, p(G = 0) = 1/2$$

Group A :

$$
\begin{aligned}
&= \underbrace{2000(p(G=1) * (1 - \underbrace{.026}_{\text{false -ve}}))}_{\text{Actually Pregnant}} + \underbrace{2000(p(G=0)(\underbrace{.0001}_{\text{false +ve}}))}_{\text{falsely identified as pregnant}} \\
&= 1000(.974 + .0001) \\
&= 974.1
\end{aligned}
$$

Group B people conduct test on their own :

$$
\begin{aligned}
&= \underbrace{8000(p(G=1)(1 - \underbrace{.25}_{\text{false -ve}}))}_{\text{Actually Pregnant}} + \underbrace{8000(p(G=0)(\underbrace{.0001}_{\text{false +ve}}))}_{\text{falsely identified as pregnant}} \\
&= 4000(.750 + .0001) \\
&= 3000.4
\end{aligned}
$$

3.Wrong result *Group A*:

$$
\begin{aligned}
W_A &= 2000(p(T=0|G=1)p(G=1) + p(T=1|G=0)p(G=0)) \\
&= 1000(.0001 + .026) \\
&= 26.1
\end{aligned}
$$

Wrong result *Group B*:

$$
\begin{aligned}
W_B &= 8000(p(T=0|G=1)p(G=1) + p(T=1|G=0)p(G=0)) \\
&= 4000(.0001 + .25) \\
&= 1000.4
\end{aligned}
$$

2.3 Third Subtask

1. General expression for a posterior distribution.

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

2.Posterior :

$$p(\mu|x_1, x_2...., x_n) = \frac{p(x_1, x_2..., x_n|\mu)p(\mu)}{p(x_1, x_2, ..., x_n)}$$

$$p(\mu|x_1, x_2...., x_n) = \frac{\mathcal{N}(x_1, x_2, ..., x_n|\mu, \sigma^2)\mathcal{N}(\mu|\mu_0, \sigma_0^2)}{\int_{-\infty}^{\infty}\mathcal{N}(x_1, x_2, ..., x_n|\mu, \sigma^2)\mathcal{N}(\mu|\mu_0, \sigma_0^2)d\mu}$$

Considering data Point as i.i.d's:

$$= \frac{(\prod_{i=1}^{n}\mathcal{N}(x_i|\mu.\sigma^2))\mathcal{N}(\mu|\mu_0, \sigma_0^2)}{\int_{-\infty}^{\infty}\prod_{i=1}^{n}\mathcal{N}(x_i|\mu, \sigma^2)\mathcal{N}(\mu|\mu_0, \sigma_0^2)d\mu}$$