ML1 Home Assignment 1

id: 12179078

Volodymyr Medentsiy

№1.1.

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \qquad b = (1\ \ 2\ \ 5)^T$$

a)

$$AB = \begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2+4 & 2+3 & 3+4 \\ 3+5 & 3+4 & 4+5 \\ 4+6 & 4+5 & 5+6 \end{pmatrix} =$$

$$= \begin{pmatrix} 6 & 5 & 7 \\ 8 & 7 & 9 \\ 10 & 9 & 11 \end{pmatrix}$$

b) To find whether A and B are invertible, we will find det A and det B. X - is invertible $\Leftrightarrow \det X \neq 0$

$$\det A = 2 \cdot \begin{vmatrix} 4 & 5 \\ 5 & 6 \end{vmatrix} - 3 \begin{vmatrix} 3 & 5 \\ 4 & 6 \end{vmatrix} + 4 \cdot \begin{vmatrix} 3 & 4 \\ 4 & 5 \end{vmatrix} =$$

$$= 2 \cdot (24 - 25) - 3(18 - 20) + 4(15 - 16) =$$

$= -2 + 6 - 4 = 0 \Rightarrow$ A is not invertible.

$$\det B = 1\begin{vmatrix}1 & 1\\ 0 & 1\end{vmatrix} - 1\cdot\begin{vmatrix}0 & 1\\ 1 & 1\end{vmatrix} = 1+1=2 \Rightarrow$$

$\Rightarrow$ B is invertible

$$B^{-1} = \frac{1}{2}\begin{pmatrix}1 & 1 & -1\\ -1 & 1 & 1\\ 1 & -1 & 1\end{pmatrix}^T = \begin{pmatrix}1/2 & 1/2 & -1/2\\ -1/2 & 1/2 & 1/2\\ 1/2 & -1/2 & 1/2\end{pmatrix}^T =$$

$$= \begin{pmatrix}1/2 & -1/2 & 1/2\\ 1/2 & 1/2 & -1/2\\ -1/2 & 1/2 & 1/2\end{pmatrix}$$

Answer: A - is not invertible, B - is invertible.

c)

$$AB = \begin{pmatrix}6 & 5 & 7\\ 8 & 7 & 9\\ 10 & 9 & 11\end{pmatrix}$$

$$\det AB = 6\begin{vmatrix}7 & 9\\ 9 & 11\end{vmatrix} - 5\begin{vmatrix}8 & 9\\ 10 & 11\end{vmatrix} + 7\begin{vmatrix}8 & 7\\ 10 & 9\end{vmatrix} =$$

$$= 6\cdot(77-81) - 5(88-90) + 7(72-70) = -6\cdot 4 + 5\cdot 2 + 7\cdot 2 =$$

$$= -24 + 24 = 0 \Rightarrow \nexists (AB)^{-1}.$$

Answer: AB is not invertible.

d) $Bx = b \Rightarrow x = B^{-1}b = \begin{pmatrix} 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} =$

$$= \begin{pmatrix} 1/2 - 1 + 5/2 \\ 1/2 + 1 - 5/2 \\ -1/2 + 1 + 5/2 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$$

$Ax = b$ – because $\nexists A^{-1}$ this system can have no solutions or infinitely many.

Let's find out how many solutions this system have:

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 1 \\ 3 & 4 & 5 & 2 \\ 4 & 5 & 6 & 5 \end{array}\right) \sim \left(\begin{array}{ccc|c} 2 & 3 & 4 & 1 \\ 0 & -1/2 & -1 & 1/2 \\ 0 & -1 & -2 & 3 \end{array}\right) \sim$$

$$\sim \left(\begin{array}{ccc|c} 2 & 3 & 4 & 1 \\ 0 & -1 & -2 & 3 \\ 0 & 0 & 0 & 2 \end{array}\right) \Rightarrow$$ This system has no solutions, because in echelon form exists row $(0\ 0\ 0 \mid 2)$

№1.2.

a) $\frac{d}{dx}\left(2/x^2 + x^{-7} + x^3\right) = 2\cdot(-2)x^{-3} - 7x^{-8} + 3x^2 = \frac{-4}{x^3} - \frac{7}{x^8} + 3x^2$

b) $\frac{d}{dx}\left(xe^{-x^{1/5}}\right) = x\cdot e^{-x^{1/5}}\cdot(-1)\cdot\frac{1}{5}x^{-4/5} + e^{-x^{1/5}}.$

c) $\frac{d}{dx}\left(1/x + \ln(x^2)\right) = -\frac{1}{x^2} + \frac{1}{x^2}\cdot 2x = -\frac{1}{x^2} + \frac{2}{x}.$

d) $\frac{d}{dx}\left((1+e^{-x})^{-1}\right) = -1\cdot(1+e^{-x})^{-2}\cdot(-1)e^{-x} = \frac{e^{-x}}{(1+e^{-x})^2}$

e) $\frac{d}{dx}(\max\{0,x\})$:

if $x<0 \Rightarrow \frac{d}{dx}(\max\{0,x\}) = 0$

if $x>0 \Rightarrow \frac{d}{dx}(\max\{0,x\}) = 1.$

if $x=0 \Rightarrow \nexists\ \frac{d}{dx}(\max\{0,x\})$,

because $\frac{d}{dx}(\max\{0,x\})\Big|_{x\to 0-} \neq \frac{d}{dx}(\max\{0,x\})\Big|_{x\to 0+}.$

g) $\frac{df(x)}{dx} \in \mathbb{R}$

with $f: \mathbb{R} \to \mathbb{R}$, $x \in \mathbb{R}$.

h) $\frac{df(x)}{dx}$, with $f(x): \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$

$$\frac{df(x)}{dx} = \begin{pmatrix} \frac{df(x)}{dx_1} & \cdots & \frac{df(x)}{dx_n} \end{pmatrix} \in \mathbb{R}^{1 \times n}$$

i) $\frac{df(x)}{dx}$, with $f(x): \mathbb{R}^n \to \mathbb{R}^m$, $x \in \mathbb{R}^n$

$$\frac{df(x)}{dx} = \begin{pmatrix} \frac{df(x)_1}{dx_1} & \cdots & \frac{df(x)_1}{dx_n} \\ \vdots & \ddots & \vdots \\ \frac{df(x)_m}{dx_1} & \cdots & \frac{df(x)_m}{dx_n} \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

j) $f(x) = 2\exp(x_2 - \ln(x_1^{-1}) - \sin(x_3 x_1^2))$

$$\frac{df(x)}{dx} = \begin{pmatrix} \frac{df(x)}{dx_1} & \frac{df(x)}{dx_2} & \frac{df(x)}{dx_3} \end{pmatrix} \in \mathbb{R}^{1 \times 3}$$

the coordinates of the $\frac{df(x)}{dx} \longrightarrow$

$$\frac{df(x)}{dx_1} = 2\exp(x_2 - \ln(x_1^{-1}) - \sin(x_3 x_1^2)) \cdot \left(\frac{1}{x_1} - \cos(x_3 x_1^2) \cdot x_3 2x_1 \right)$$

$$\frac{df(x)}{dx_2} = 2\exp(x_2 - \ln(x_1^{-1}) - \sin(x_3 x_1^2))$$

$$\frac{df(x)}{dx_3} = 2\exp(x_2 - \ln(x_1^{-1}) - \sin(x_3 x_1^2)) \cdot (-\cos(x_3 x_1^2) \cdot x_1^2)$$

k) $\nabla_y h \in \mathbb{R}$

$h(y) = (g \circ f)(y)$, where $g(x) = x_1^3 + \exp(x_2)$

$f(y) = (y \sin(y), y \cos y)^T$

$h(y) = (y \sin y)^3 + \exp(y \cos y)$

$$\nabla_y h = \frac{d}{dy} h(y) = 3y^2 \sin^3 y + y^3 3 \sin^2 y \cos y + \exp(y \cos y)(\cos y - y \sin y)$$

Using the Chain Rule:

$$\frac{d(g \circ f)(y)}{dy} = \frac{dg}{df} \frac{df}{dy}(y)$$

$$\frac{dg}{df} = \begin{pmatrix} 3x_1^2 & \exp(x_2) \end{pmatrix} = \begin{pmatrix} 3 \cdot (y \sin y)^2 & \exp(y \cos y) \end{pmatrix}$$

$$\frac{df(y)}{dy} = \begin{pmatrix} \sin y + y \cos y \\ \cos y - y \sin y \end{pmatrix}$$

$$\frac{dh(y)}{dy} = \left(3\cdot(y\sin y)^2 \;\; \exp(y\cos y)\right) \begin{pmatrix} \sin y + y\cos y \\ \cos y - y\sin y \end{pmatrix} =$$

$$= 3(y\sin y)^2(\sin y + y\cos y) + \exp(y\cos y)(\cos y - y\sin y) =$$

$$= 3y^2\sin^3 y + 3y^3\sin^2 y\cos y + \exp(y\cos y)(\cos y - y\sin y)$$

$$\frac{dh(y)}{dy} \in \mathbb{R}^1.$$

e) $x := f(y, z) = (y\sin y + z, y\cos y + z^2)^T$

$g(x) = x_1^3 + \exp(x_2)$, $p = (y, z)^T$

$h(y) = (g \circ f)(y) = (y\sin y + z)^3 + \exp(y\cos y + z^2)$

$\nabla_{y,z} h = \nabla_p h = \left(\frac{dh}{dy} \;\; \frac{dh}{dz}\right) \in \mathbb{R}^{1\times 2}$, where

$$\frac{dh}{dy} = 3(y\sin y + z)^2\cdot(\sin y + y\cos y) + \exp(y\cos y + z^2)\cdot$$

$$\cdot(\cos y - y\sin y)$$

$$\frac{dh}{dz} = 3(y\sin y + z)^2 + 2\exp(y\cos y + z^2)\, z$$

№1.3.

a) Let $\Sigma^{-1} = \begin{pmatrix} c_{11} & \dots & c_{1N} \\ \vdots & \ddots & \vdots \\ c_{N1} & \dots & c_{NN} \end{pmatrix}$. Because Σ^{-1} - symmetric $c_{ij} = c_{ji}$

$$\nabla_\mu X^T \Sigma^{-1} \mu = \nabla_\mu \begin{pmatrix} x_1 & \dots & x_N \end{pmatrix} \begin{pmatrix} c_{11} & \dots & c_{1N} \\ \vdots & \ddots & \vdots \\ c_{N1} & \dots & c_{NN} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} =$$

$$= \nabla_\mu (x_1 \dots x_N) \begin{pmatrix} \sum_{i=1}^N c_{1i}\mu_i \\ \vdots \\ \sum_{i=1}^N c_{Ni}\mu_i \end{pmatrix} = \nabla_\mu \sum_{j=1}^N \left(\sum_{i=1}^N c_{ji}\mu_i \right) x_j =$$

$$= \begin{pmatrix} \sum_{j=1}^N c_{j1} x_j & \dots & \sum_{j=1}^N c_{jN} x_j \end{pmatrix} = X^T \Sigma^{-1}$$

b) $\nabla_\mu \mu^T \Sigma^{-1} \mu = \nabla_\mu (\mu_1 \dots \mu_N) \begin{pmatrix} c_{11} & \dots & c_{1N} \\ \vdots & \ddots & \vdots \\ c_{N1} & \dots & c_{NN} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} =$

$$= \nabla_\mu (\mu_1 \dots \mu_N) \begin{pmatrix} \sum_{j=1}^{N} c_{1j}\mu_j \\ \vdots \\ \sum_{j=1}^{N} c_{Nj}\mu_j \end{pmatrix} = \nabla_\mu \left(\sum_{i=1}^{N} \left(\sum_{j=1}^{N} c_{ij}\mu_j \right) \mu_i \right) =$$

$$= \begin{pmatrix} \sum_{j=1}^{N} c_{1j}\mu_j + \sum_{i=1}^{N} c_{i1}\mu_i \\ \vdots \\ \sum_{j=1}^{N} c_{Nj}\mu_j + \sum_{i=1}^{N} c_{iN}\mu_i \end{pmatrix}^T = \left[\begin{array}{c} \Sigma^{-1} \text{ - is a symmetric} \\ \Downarrow \\ c_{ij} = c_{ji} \end{array} \right] =$$

$$= \begin{pmatrix} 2\sum_{i=1}^{N} c_{1i}\mu_i \\ \vdots \\ 2\sum_{i=1}^{N} c_{Ni}\mu_N \end{pmatrix}^T = 2\mu^T \Sigma^{-1}$$

c) $\nabla_w f$, where $f = Wx$ and $W \in \mathbb{R}^{2\times3}$, $x \in \mathbb{R}^3$

$$f = Wx \iff \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \iff$$

$$\Longleftrightarrow \quad \begin{aligned} f_1 &= w_{11}x_1 + w_{12}x_2 + w_{13}x_3 \\ f_2 &= w_{21}x_1 + w_{22}x_2 + w_{23}x_3 \end{aligned}$$

$$\nabla_w f = \begin{pmatrix} \frac{df_1}{dw} \\ \frac{df_2}{dw} \end{pmatrix}, \text{where } \frac{df_i}{dw} = \begin{pmatrix} \frac{df_i}{dw_{11}} & \frac{df_i}{dw_{21}} \\ \frac{df_i}{dw_{12}} & \frac{df_i}{dw_{22}} \\ \frac{df_i}{dw_{13}} & \frac{df_i}{dw_{23}} \end{pmatrix}$$

$$\nabla_w f = \begin{pmatrix} \begin{pmatrix} x_1 & 0 \\ x_2 & 0 \\ x_3 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & x_1 \\ 0 & x_2 \\ 0 & x_3 \end{pmatrix} \end{pmatrix} \in \mathbb{R}^{2 \times (3 \times 2)}$$

d) $\nabla_W f$, where $f(x) = (\mu - y(W,x))^T \Sigma^{-1} (\mu - y(W,x))$; $y(W,x) = W^T x$

$$f(x) = \mu^T \Sigma^{-1} \mu - \mu^T \Sigma^{-1} y(W,x) - y(W,x)^T \Sigma^{-1} \mu + y(W,x)^T \Sigma^{-1} y(W,x) =$$

$$= \mu^T \Sigma^{-1} \mu - 2\mu^T \Sigma^{-1} y(W,x) + y^T(W,x) \Sigma^{-1} y(W,x).$$

General form: $\frac{df}{dW} = \begin{pmatrix} \frac{df}{dW_{11}} & \cdots & \frac{df}{dW_{M1}} \\ \vdots & \ddots & \vdots \\ \frac{df}{dW_{1K}} & \cdots & \frac{df}{dW_{MK}} \end{pmatrix} \in \mathbb{R}^{K \times M}$

1) Let's find $\frac{df}{dW_{ij}} = \sum_{k=1}^{K} \frac{df}{dy_k} \frac{dy_k}{dW_{ij}}$:

using 1.3.a and 1.3.b :

$$\frac{df}{dy} = \left(\frac{df}{dy_1} \cdots \frac{df}{dy_K} \right) = -2\mu^T \Sigma^{-1} + 2y^T(W,x) \Sigma^{-1}$$

$$\frac{df}{dy_k} = -2\mu^T \Sigma^{-1}_{:,k} + 2y^T(W,x) \Sigma^{-1}_{:,k}.$$

2) $y = W^T x \iff \vec{y} = \begin{pmatrix} \sum_{i=1}^{M} W_{i1} x_i \\ \vdots \\ \sum_{i=1}^{M} W_{iK} x_i \end{pmatrix}$

$$y_K = \sum_{i=1}^{M} w_{iK} x_i = (W_{:,K})^T x. \Rightarrow$$

$$\Rightarrow \frac{dy_K}{dw_{ij}} = \begin{cases} x_i & , j = K \\ 0 & , j \neq K \end{cases}$$

3) $$\frac{df}{dW_{ij}} = 2(y-\mu)^T \Sigma^{-1}_{:,j} x_i .$$

$$\text{So } \frac{df}{dW} = \begin{pmatrix} 2(y-\mu)^T \Sigma^{-1}_{:,1} x_1 & \dots & 2(y-\mu)^T \Sigma^{-1}_{:,1} x_M \\ \vdots & \ddots & \vdots \\ 2(y-\mu)^T \Sigma^{-1}_{:,K} x_1 & \dots & 2(y-\mu)^T \Sigma^{-1}_{:,K} x_M \end{pmatrix}.$$

№ 2.1.

a) In my opinion Bart is a farmer, simply because there exists more farmers than librarians. So it's more probable that an arbitrary person will be a farmer.

b) $P(\text{librarian} \mid \text{introvert}) = 0{,}8.$

$P(\text{librarian} \mid \text{extrovert}) = 0{,}1$

$P(\text{extrovert}) = 0{,}7$

ξ ~ a person is an extrovert.

$\xi = 1$ - a person is an extrovert. $P(\xi = 1) = 0{,}7$

$\xi = 0$ - a person is an introvert $P(\xi = 0) = 0{,}3$

η ~ Bart's occupation

$\eta = 0$ – Bart is a librarian

$\eta = 1$ - Bart is a farmer.

$P(\eta = 0 \mid \xi = 0) = 0{,}8$

$P(\eta = 1 \mid \xi = 0) = 0{,}2$

$P(\eta = 0 \mid \xi = 1) = 0{,}1$

$P(\eta = 1 \mid \xi = 1) = 0{,}9.$

c) $P(\eta = 0) = P(\xi = 0) \cdot P(\eta = 0 \mid \xi = 0) + P(\xi = 1) \cdot P(\eta = 0 \mid \xi = 1) =$

$= 0{,}3 \cdot 0{,}8 + 0{,}7 \cdot 0{,}1 = 0{,}24 + 0{,}07 = 0{,}31.$

d) $\begin{cases} P(\eta = 1) = 1000 \cdot P(\eta = 0) \\ P(\eta = 1) + P(\eta = 0) = 1 \end{cases} \Rightarrow \begin{cases} P(\eta = 0) = 1/1001 \\ P(\eta = 1) = 1000/1001 \end{cases}$

Probability of Bart being librarian if he is an introvert:

$$P(\eta=0|\xi=0) = \frac{P(\xi=0|\eta=0)\,P(\eta=0)}{P(\xi=0)} =$$

$$= \left[\begin{array}{l} P(\xi=1|\eta=0)=0{,}1 \Rightarrow \\ \Rightarrow P(\xi=0|\eta=0)=0{,}9 \end{array}\right] = \frac{0{,}9}{1001\cdot 0{,}3} = \frac{3}{1001}$$

Using the actual statistics we found that $P(\eta=0) = 1/1001$ which is significantly lower than the result obtained in the task c). ($P(\eta=0)=0{,}31$).

N2.2.

$D=\{x_1,\ldots,x_N\}$ – N data points obtained from Bernoulli distribution with parameter p.

a) Likelihood $(D) = p(x_1,\ldots,x_N|p) = \left[\{x_1,\ldots,x_N\} - \text{i.i.d.}\right] =$

$$= \prod_{i=1}^{N} p(x_i|p) = \prod_{i=1}^{N} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{N} x_i}(1-p)^{N-\sum_{i=1}^{N} x_i}.$$

b) I think that $p>0{,}5$, though just 2 observations is a very small data set for a proper analyses.

We can also use maximum likelihood estimation to find p which most likely produced $[1,1]$.

$$p^* = \underset{p}{\arg\max} \ln(\text{Likelihood}(D))$$

$$\frac{d}{dp} \ln(\text{Likelihood}(D)) = \frac{d}{dp} \left(\sum_{i=1}^{N} x_i \ln p + \left(N - \sum_{i=1}^{N} x_i\right) \ln(1-p) \right) =$$

$$= \sum_{i=1}^{N} x_i \cdot \frac{1}{p} - \left(N - \sum_{i=1}^{N} x_i\right) \frac{1}{1-p} = \left[\sum_{i=1}^{N} x_i = S \right] = S \cdot \frac{1}{p} - (N-S)\frac{1}{1-p} = 0$$

$$\Rightarrow S(1-p) - (N-S)p = p(-S-N+S) + S = 0 \Rightarrow p^* = \frac{S}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

So in our case: $p^* = \frac{1+1}{2} = 1.$

c) Posterior over p: $P(p|D) = \frac{P(D|p)\,P(p)}{P(D)}$,

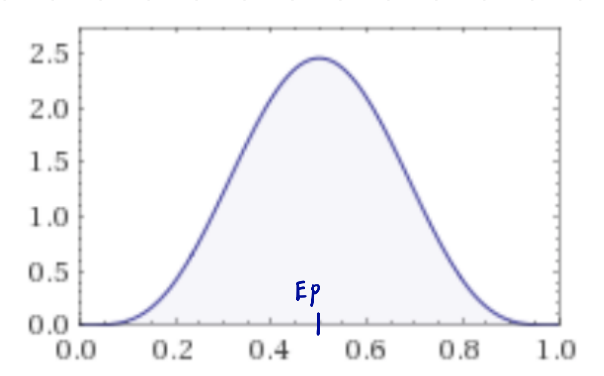
where $P(p)$ - is the prior, $P(D|p)$ - likelihood, $P(D)$ - evidence, $P(p|D)$ - posterior.

d) Our posterior will show that coin most probably is not fair and shift probability density into the region where $p > 0{,}5$.

We can illustrate this by using Beta(5,5) as a prior distribution for p, which will reflect our belief that coin is fair. This prior is convenient for calculations, because Beta distribution is a conjugate prior with Bernoulli likelihood, so the parameter of posterior could be easily updated.

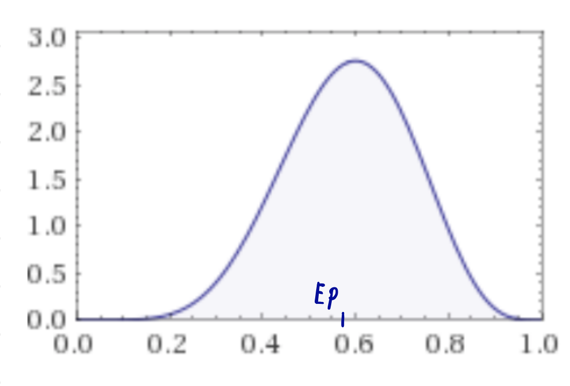
After observing $[1;1]$ new parameters of posterior distribution could be computed as:

$\text{Beta}(5+2;5) = \text{Beta}(7;5)$. New mean is $\frac{7}{12} > 0{,}5$ and we see that there is more probability density in the region $p > 0{,}5$.



PDF of Beta(5;5)

prior distribution of p



PDF of Beta(7;5)

posterior distribution of p.