ML1 Home Assignment 2

id: 12179078

Volodymyr Medentsiy

№1

a) $p(D|\Theta) = p(t_1, \dots t_N | \Omega, w, \Omega) = N(\vec{t} | \Phi w, \Omega) =$

$$= \frac{1}{(2\pi)^{N/2}\sqrt{\det \Omega}} \exp\left(-\frac{1}{2}\left(\Omega^{-1}(t - \Phi w), (t - \Phi w)\right)\right) \text{ - matrix form.}$$

$\left((\Omega^{-1}(t-\Phi w), (t-\Phi w)) = (t-\Phi w)^T \Omega^{-1} (t - \Phi w) \right.$. I mainly use dot product notation in this exercise)

Vector form: Let $\Omega^{-1} = \|c_{ij}\|$ (Ω^{-1} - is also symmetric)

$$p(D|\Theta) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \Omega}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} (t_j - \varphi_{:;j} \cdot w)^2\right),$$

where $\varphi_{:;j}$ - denotes j-th row.

b) $\Omega = A^T D A$, with $D = \mathrm{diag}(d_1, \dots, d_N)$, $A^T = A^{-1}$.

Let's express $p(D|\Theta)$ as a function of A, D, Φ and w.

$$p(D|\Theta) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \Omega}} \exp\left(-\frac{1}{2}(\Omega^{-1}(t - \Phi w), (t - \Phi w))\right) =$$

$$= \frac{1}{(2\pi)^{N/2}\sqrt{\det(A^T D A)}} \exp\left(-\frac{1}{2}\left((A^T D A)^{-1}(t - \Phi w), (t - \Phi w)\right)\right) =$$

$$= \left[\begin{array}{l} \det A^T D A = \det A^T \det D \det A = (\det A)^{-1} \det D \det A = \det D = \prod_{i=1}^{N} d_i \\ (A^T D A)^{-1} = A^{-1} D^{-1} (A^T)^{-1} = A^{-1} D^{-1} A \end{array} \right] =$$

$$= \frac{1}{(2\pi)^{N/2}\sqrt{\det D}} \exp\left(-\frac{1}{2}\left(A^T D^{-1} A (t - \Phi w), (t - \Phi w)\right)\right) = \left[\begin{array}{l} (A^T D^{-1} A u, u) = \\ = (D^{-1} A u, A u) \end{array} \right] =$$

$$= \frac{1}{(2\pi)^{N/2}\sqrt{\det D}} \exp\left(-\frac{1}{2}\left(D^{-1}(At - A\Phi w), (At - A\Phi w)\right)\right) = \left[\begin{array}{l} \tau = At \\ \Psi = A\Phi \end{array} \right] =$$

$$= \frac{1}{(2\pi)^{N/2}\sqrt{\prod_{i=1}^{N} d_i}} \exp\left(-\frac{1}{2}\left(D^{-1}(\tau-\Psi w),(\tau-\Psi w)\right)\right) = N(\tau \mid \Psi w, D)$$

$$\left(N(\tau \mid \Psi w, D) = N(At \mid A\Phi w, D). \right)$$

c) factorize the distribution into a product of univariate Gaussian.

$$\frac{1}{(2\pi)^{N/2}\sqrt{\prod_{i=1}^{N} d_i}} \exp\left(-\frac{1}{2}\left(D^{-1}(\tau-\Psi w),(\tau-\Psi w)\right)\right) =$$

$$= \left[D^{-1} = \begin{pmatrix} 1/d_1 & & 0 \\ & \ddots & \\ 0 & & 1/d_N \end{pmatrix}; \quad \tau - \Psi w = \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_N \end{pmatrix} - \begin{pmatrix} \sum_{i=1}^{N} \psi_{1i} \cdot w_i \\ \vdots \\ \sum_{i=1}^{N} \psi_{Ni} \cdot w_i \end{pmatrix} = \vec{\tau} - \begin{pmatrix} \psi_1 w \\ \vdots \\ \psi_N w \end{pmatrix} \right] =$$

ψ_i - i-th row of matrix Ψ.

$$= \frac{1}{(2\pi)^{N/2}\sqrt{\prod_{i=1}^{N} d_i}} \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{N} \frac{(\tau_i - \psi_i w)^2}{d_i}\right)\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sqrt{d_i}} \exp\left(-\frac{(\tau_i - \psi_i w)^2}{2d_i}\right) =$$

$$= \prod_{i=1}^{N} N(\tau_i \mid \psi_i w, d_i)$$

d) Prior of w:

$$p(w) = N(w \mid 0, \alpha^{-1} I) = \frac{\alpha^{M/2}}{(2\pi)^{M/2}} \exp\left(-\frac{1}{2}\alpha w^T w\right)$$

$$\ln p(w) = \frac{M}{2}\ln\alpha - \frac{M}{2}\ln 2\pi - \frac{\alpha}{2} w^T w$$

e) $p(w|D) = \dfrac{p(D|w)p(w)}{p(D)}$

$p(D|w) = N(t|\Phi w, \Omega)$

$p(w) = N(w|0, \alpha^{-1} I)$

$p(D) = \int p(D|w)p(w)\,dw = \int N(t|\Phi w, \Omega) N(w|0, \alpha^{-1} I)\,dw =$

$= p(t|\Phi, \Omega, \alpha)$

So $p(w|D) = \dfrac{N(t|\Phi w, \Omega) N(w|0, \alpha^{-1} I)}{p(t|\Phi, \Omega, \alpha)}$

f) $\ln p(w|D) = \ln p(D|w) + \ln p(w) - \ln p(D)$

Matrix form:

$\ln p(D|w) = \ln \dfrac{1}{(2\pi)^{N/2}\sqrt{\det D}} \exp\left(-\frac{1}{2}\left(D^{-1}(\tau - \Psi w), (\tau - \Psi w)\right)\right) =$

$= \frac{N}{2} \ln 2\pi - \frac{1}{2} \ln \det D - \frac{1}{2}\left(D^{-1}(\tau - \Psi w), (\tau - \Psi w)\right)$

So $\ln p(w|D) = C - \frac{\alpha}{2} w^T w - \frac{1}{2}\left(D^{-1}(\tau - \Psi w), (\tau - \Psi w)\right)$,

with $C = \frac{N}{2} \ln 2\pi - \frac{1}{2} \ln \det D + \frac{M}{2} \ln \alpha - \frac{M}{2} \ln 2\pi - \ln p(D)$

Factorized component form:

$\ln p(D|w) = \ln \dfrac{1}{(2\pi)^{N/2}\sqrt{\prod_{i=1}^{M} d_i}} \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{N} \frac{\tau_i - \psi_i w}{d_i}\right)\right) =$

$= -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{N} \ln d_i - \frac{1}{2} \sum_{i=1}^{N} \frac{(\tau_i - \psi_i w)^2}{d_i}$

$$\ln p(w) = \frac{M}{2}\ln\alpha - \frac{M}{2}\ln 2\pi - \frac{\alpha}{2}\sum_{i=1}^{M} w_i^2$$

$$\ln p(w|D) = -\frac{1}{2}\sum_{i=1}^{N}\frac{(\tau_i - \psi_i w)^2}{d_i} - \frac{\alpha}{2}\sum_{i=1}^{M} w_i^2 + C$$

with $C = \frac{M}{2}\ln\alpha - \frac{M}{2}\ln 2\pi - \ln p(D) - \frac{N}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln d_i.$

Finding the MAP estimate is easier than finding the full posterior distribution, because we do not need to find the distribution $p(D)$, which is hard to compute.

g) W_{MAP}:

$$w_{MAP} = \arg\max_w p(w|D) = \arg\max_w \left(-\frac{\alpha}{2}w^T w - \frac{1}{2}(D^{-1}(\tau - \Psi w), (\tau - \Psi w))\right)$$

$$\frac{d}{dw}\left(-\frac{\alpha}{2}w^T w - \frac{1}{2}(D^{-1}(\tau - \Psi w), (\tau - \Psi w))\right) = \left[(D^{-1}(\tau - \Psi w), (\tau - \Psi w)) = \right.$$

$$\left. \begin{aligned} &= (D^{-1}\tau - D^{-1}\Psi w, \tau - \Psi w) = (D^{-1}\tau)^T\tau - (D^{-1}\Psi w)^T\tau - (D^{-1}\tau)^T\Psi w + \\ &+ (D^{-1}\Psi w)^T\Psi w = \tau^T D^{-1}\tau - w^T\Psi^T D^{-1}\tau - \tau^T D^{-1}\Psi w + w^T\Psi^T D^{-1}\Psi w \end{aligned}\right] =$$

$$= -\alpha w^T + \frac{1}{2}(\Psi^T D^{-1}\tau)^T + \frac{1}{2}\tau^T D^{-1}\Psi - w^T\Psi^T D^{-1}\Psi =$$

$$= -\alpha w^T + \tau^T D^{-1}\Psi - w^T\Psi^T D^{-1}\Psi = 0 \iff w^T(-\alpha I - \Psi^T D^{-1}\Psi) = -\tau^T D^{-1}\Psi$$

$$\Rightarrow w^T = (\alpha I + \Psi^T D^{-1}\Psi)^{-1}\tau^T D^{-1}\Psi.$$

$$\Rightarrow w_{MAP} = (\alpha I + \Psi^T D^{-1}\Psi)\cdot\Psi^T D^{-1}\tau$$

h) $w_{MAP} = (\alpha I + \Phi^T A^T D^{-1} A \Phi)^{-1} \Phi^T A^T D^{-1} A t =$

$= (\alpha I + \Phi^T \Omega^{-1} \Phi)^{-1} \Phi^T \Omega^{-1} t.$

Note: w_{MAP} is $\underset{w}{\arg\max}$, because $-\frac{\alpha}{2} w^T w - \left(\frac{1}{2}\left(D^{-1}(\tau - \Psi w)(\tau - \Psi w)\right)\right)$ is convex in w.

N3

a) $\lambda_{MLE} = \underset{\lambda}{\arg\max}\, p(x_1, \dots, x_N \mid \lambda)$

$$p(x_1, \dots, x_N \mid \lambda) = \prod_{i=1}^{N} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$\ln p(x_1, \dots, x_N \mid \lambda) = \ln \prod_{i=1}^{N} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = -N\lambda + \sum_{i=1}^{N} (x_i \ln \lambda - \ln x_i!)$$

$$\frac{d}{d\lambda} \ln p(x_1, \dots, x_N \mid \lambda) = -N + \sum_{i=1}^{N} \frac{x_i}{\lambda} = 0 \Rightarrow \lambda^* = \frac{\sum_{i=1}^{N} x_i}{N},$$ λ^* is indeed argmax

because $\frac{d^2}{d\lambda^2} \ln p(x_1, \dots, x_N \mid \lambda) = -\frac{1}{\lambda} \sum_{i=1}^{N} x_i \Big|_{\lambda = \frac{\sum_{i=1}^{N} x_i}{N}} = -N < 0 \Rightarrow \lambda_{MLE} = \frac{\sum_{i=1}^{N} x_i}{N}.$

b) $p(\lambda) \propto \exp(-\lambda/a)$, for $\lambda \geq 0$ and 0 otherwise. $(a > 0) \Rightarrow$

$$\Rightarrow p(\lambda) = \begin{cases} \frac{1}{a} e^{-\lambda a}, & \lambda \geq 0 \\ 0, & \lambda < 0 \end{cases}$$

$$p(\lambda \mid D) = \frac{p(D \mid \lambda)\, p(\lambda)}{p(D)} \Rightarrow \lambda_{MAP} = \underset{\lambda}{\arg\max}\, p(\lambda \mid D) = \underset{\lambda}{\arg\max} \ln p(\lambda \mid D) =$$

$$= \underset{\lambda}{\arg\max} \left(\ln p(D \mid \lambda) + \ln p(\lambda) - \ln p(D)\right) = \underset{\lambda}{\arg\max} \left(\ln p(D \mid \lambda) + \ln p(\lambda)\right)$$

$$\ln p(D \mid \lambda) = \ln \prod_{i=1}^{N} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = -\lambda N + \sum_{i=1}^{N} x_i \ln \lambda$$

$$\ln p(\lambda) = \ln \frac{1}{a} \exp\left(-\frac{\lambda}{a}\right) = -\ln a - \frac{\lambda}{a}$$

$$\frac{d}{d\lambda}\left(\ln p(D|\lambda) + \ln p(\lambda)\right) = -N + \frac{1}{\lambda}\sum_{i=1}^{N} x_i - \frac{1}{a} = 0 \Rightarrow$$

$$\Rightarrow \lambda_{MAP} = \frac{\sum_{i=1}^{N} x_i}{N + \frac{1}{a}}.$$

c)

Effect of the prior: With smaller a the prior will shift more probability density near 0 and the MAP solution for λ will be also closer to 0.

In general $\lambda_{MAP} \leq \lambda_{MLE}$. (the equality holds when $a \to \infty$).

$$\lambda_{MAP} \xrightarrow[a \to 0]{} 0$$

$$\lambda_{MAP} \xrightarrow[a \to \infty]{} \frac{\sum_{i=1}^{N} x_i}{N} = \lambda_{MLE}$$

№2

We observe $\begin{cases} c = \cos\theta + \delta\varepsilon \\ s = \sin\theta + \delta\varepsilon \end{cases}$, $\varepsilon \sim N(0; 1)$

so $c \sim N(\cos\theta, \delta^2)$, $s \sim N(\sin\theta, \delta^2)$.

Since c and s are linearly independent $\Rightarrow$ (s, c) is Gaussian vector and $p(s, c | \theta, \delta^2) = \frac{1}{2\pi\delta} \exp \cdot \left(-\frac{1}{2\delta^2}\left((s - \sin\theta)^2 + (c - \cos\theta)^2\right)\right)$

Let's find θ_{MLE}:

$$\frac{d}{d\theta}\ln p(s,c|\theta,\delta^2) = \frac{d}{d\theta}\left(-\frac{1}{2\delta^2}\left((s-\sin\theta)^2+(c-\cos\theta)^2\right)\right) =$$

$$= \frac{(s-\sin\theta)}{\delta^2}\cdot\cos\theta - \frac{(c-\cos\theta)\sin\theta}{\delta^2} = 0 \Rightarrow$$

$$\Rightarrow s\cos\theta - \sin\theta\cos\theta - c\sin\theta + \sin\theta\cos\theta =$$

$$= s\cos\theta - c\sin\theta = 0 \Rightarrow s - c\,tg\theta = 0 \Rightarrow tg\theta = s/c$$

$$\Rightarrow \theta^* = \left[\begin{array}{l}\arctan(s/c)\\ \pi+\arctan(s/c)\end{array}\right.$$

To find the exact solution let's find $\frac{d^2}{d\theta^2}\ln p(s,c|\theta,\delta^2) =$

$$= -\frac{1}{\delta^2}\cdot(s\sin\theta + c\cos\theta)$$

when $\theta = \arctan(s/c)$: $\frac{d^2}{d\theta^2}\ln p(s,c|\theta,\delta^2) = \left[\begin{array}{l}\sin(\arctan x) = \frac{x}{\sqrt{x^2+1}}\\ \cos(\arctan x) = \frac{1}{\sqrt{x^2+1}}\end{array}\right] =$

$$= \frac{-1}{\delta^2}\left(\frac{s^2}{c\sqrt{(s/c)^2+1}} + \frac{c}{\sqrt{(s/c)^2+1}}\right) = \frac{-1}{\delta^2 c}\left(\frac{s^2+c^2}{\sqrt{(s/c)^2+1}}\right) < 0 \text{ if } c>0$$

if $c=0$ then $\arctan(s/c) = \left[\begin{array}{l}-\pi/2,\ s>0\\ \pi/2,\ s<0\end{array}\right.$

when $\theta = \pi + \arctan(s/c)$ then: $\frac{d^2}{d\theta^2}\ln p(s,c|\theta,\delta^2) =$

$$= \left[\begin{array}{l}\sin(\arctan x+\pi) = -\frac{x}{\sqrt{x^2+1}}\\ \cos(\arctan x+\pi) = -\frac{1}{\sqrt{x^2+1}}\end{array}\right] = -\frac{1}{\delta^2}\left(-\frac{s^2}{c\sqrt{(s/c)^2+1}} - \frac{c}{\sqrt{(s/c)^2+1}}\right) =$$

$$= \frac{1}{\delta^2 c}\cdot\frac{(s^2+c^2)}{\sqrt{(s/c)^2+1}} < 0 \text{ if } c<0.$$

So we can conclude that $\theta_{MLE} = \begin{cases} \arctan\left(\frac{s}{c}\right), & \text{if } c \geq 0 \\ \pi + \arctan\left(\frac{s}{c}\right), & \text{if } c < 0 \end{cases}$

and $\theta_{MLE} \in [-\pi/2; 3\pi/2)$.