TEK5020/9020 – Mønstergjenkjenning
Høsten 2023

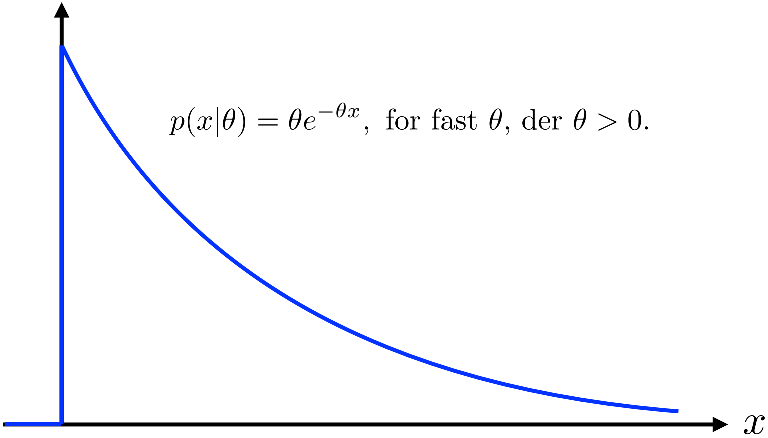
Løsningsforslag – Øvingsoppgaver 2

Idar Dyrdal (idar.dyrdal@its.uio.no)

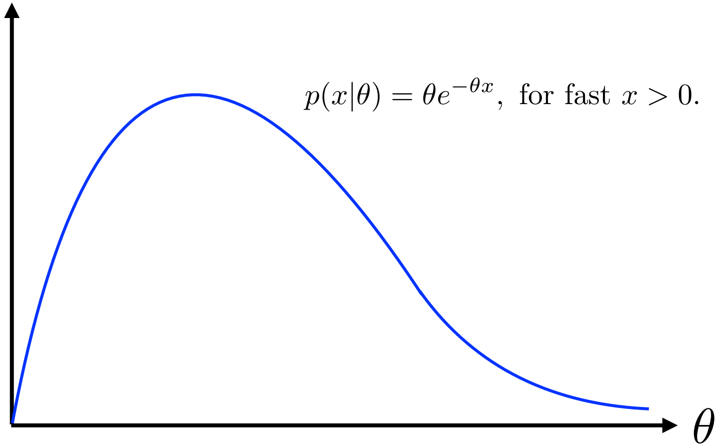
UiO : Institutt for teknologisystemer

23. august 2023

Oppgave 1a



Oppgave 1b



Oppgave 1c

Likelihoodfunksjonen blir her

$$p(\mathscr{X} \mid \theta) = \prod_{i=1}^{n} p(x_i \mid \theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n \prod_{i=1}^{n} e^{-\theta x_i}$$

slik at Log-likelihoodfunksjonen blir

$$\begin{aligned} \mathscr{L}(\theta) &= \ln p(\mathscr{X} \mid \theta) \\ &= n \ln \theta + \sum_{i=1}^{n} (-\theta x_i) \\ &= n \ln \theta - \theta \sum_{i=1}^{n} x_i. \end{aligned}$$

Oppgave 1c (forts.)

Den deriverte med hensyn til θ blir

$$\frac{d\mathscr{L}(\theta)}{d\theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i,$$

som skal settes lik null for å finne maksimum av Log-likelihoodfunksjonen, slik at

$$\frac{n}{\theta} = \sum_{i=1}^{n} x_i.$$

Maksimum likelihood-estimatet av parameteren θ blir da

$$\hat{\theta} = \underline{\underline{\frac{n}{\sum_{i=1}^{n} x_i}}}.$$

Oppgave 2

Sannsynligheten for $\boldsymbol{x}$ kan her skrives som

$$P(\boldsymbol{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i} (1 - \theta_i)^{1-x_i} \quad \text{(Bernoullifordelingen)}$$

Likelihood-funksjonen blir da

$$P(\mathscr{X} \mid \boldsymbol{\theta}) = \prod_{k=1}^{n} P(\boldsymbol{x}_k \mid \boldsymbol{\theta}) = \prod_{k=1}^{n} \prod_{i=1}^{d} \theta_i^{x_{ik}} (1 - \theta_i)^{1-x_{ik}}$$

der treningssettet er

$$\mathscr{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_2\}$$

og x_{ik} er komponent nr. i i treningsvektor $\boldsymbol{x}_k$.

Oppgave 2 (forts.)

Log-likelihoodfunksjonen blir da

$$\mathscr{L}(\boldsymbol{\theta}) = \ln P(\mathscr{X} \mid \boldsymbol{\theta}) = \sum_{k=1}^{n} \sum_{i=1}^{d} \{x_{ik} \ln \theta_i + (1 - x_{ik}) \ln (1 - \theta_i)\}$$

som gir

$$\frac{\partial \mathscr{L}(\boldsymbol{\theta})}{\partial \theta_j} = \sum_{k=1}^{n} \sum_{i=1}^{d} \left\{ x_{ik} \frac{\delta_{ij}}{\theta_i} - (1 - x_{ik}) \frac{\delta_{ij}}{1 - \theta_i} \right\} = 0 \quad \forall\, \theta_j,$$

der $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$ (Kronecker delta)

Oppgave 2 (forts.)

Dette gir likningssystemet

$$\sum_{k=1}^{n} \left\{ \frac{x_{jk}}{\theta_j} - \frac{1 - x_{jk}}{1 - \theta_j} \right\} = 0, \quad j = 1, \cdots, d$$

slik at

$$\hat{\theta}_j = \frac{1}{n} \sum_{k=1}^{n} x_{jk}$$

og derav

$$\underline{\underline{\hat{\boldsymbol{\theta}} = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{x}_k}},$$

hvilket skulle bevises.

Oppgave 3

Denne oppgaven går ut på å finne maksimum-likelihood estimatet av kovariansmatrisen i den multivariate normalfordelingen. Her er altså

$$p(\boldsymbol{x}|\Sigma) \sim N(\boldsymbol{\mu}, \Sigma),$$

der $\boldsymbol{\mu}$ er kjent og Σ er ukjent, og vi skal vise at maksimum-likelihood estimatet for Σ er gitt ved

$$\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^t.$$

Dette gjøres i flere trinn i de følgende deloppgavene.

Oppgave 3a

Første trinn i utledningen er å bevise matriseidentiteten $a^t A\boldsymbol{a} = \mathrm{Tr}\left\{A\boldsymbol{a}\boldsymbol{a}^t\right\}$, der $\boldsymbol{a}$ er en vektor og A er en kvadratisk matrise. Venstre side kan skrives på komponentform som

$$\boldsymbol{a}^t A\boldsymbol{a} = \sum_i a_i \left(\sum_j A_{ij} a_j\right) = \sum_{ij} a_i A_{ij} a_j = \sum_{ij} A_{ij} a_i a_j .$$

Hvert element i matrisen $A\boldsymbol{a}\boldsymbol{a}^t$ på høyre side kan skrives som

$$\left\{A\boldsymbol{a}\boldsymbol{a}^t\right\}_{ij} = \sum_k A_{ik} a_k a_j ,$$

som ved innsetting i høyre side gir

$$\mathrm{Tr}\left\{A\boldsymbol{a}\boldsymbol{a}^t\right\} = \sum_l \left\{A\boldsymbol{a}\boldsymbol{a}^t\right\}_{i=j=l} = \sum_l \sum_k A_{lk} a_k a_l = \sum_{lk} A_{lk} a_l a_k .$$

Dette viser at de to sidene er identiske, slik at $a^t A\boldsymbol{a} = \mathrm{Tr}\left\{A\boldsymbol{a}\boldsymbol{a}^t\right\}$, hvilket skulle bevises.

Oppgave 3b

Her skal likelihoodfunksjonen omskrives ved hjelp av identiteten i første deloppgave, slik at den blir

$$
\begin{aligned}
p(\mathscr{X} \mid \Sigma) &= p(\boldsymbol{x}_1, \ldots, , \boldsymbol{x}_n \mid \Sigma) \\
&= \prod_{k=1}^{n} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}_k - \boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x}_k - \boldsymbol{\mu})} \\
&= \frac{1}{(2\pi)^{\frac{nd}{2}}} |\Sigma^{-1}|^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{k=1}^{n}(\boldsymbol{x}_k - \boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x}_k - \boldsymbol{\mu})} \qquad \left(\text{siden } |\Sigma| = \frac{1}{|\Sigma^{-1}|}\right) \\
&= \frac{1}{(2\pi)^{\frac{nd}{2}}} |\Sigma^{-1}|^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{k=1}^{n} \mathrm{Tr}\left\{\Sigma^{-1}(\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^t\right\}} \qquad \text{(Se deloppgave a)} \\
&= \underline{\underline{\frac{1}{(2\pi)^{\frac{nd}{2}}} |\Sigma^{-1}|^{\frac{n}{2}} e^{-\frac{1}{2}\mathrm{Tr}\left\{\Sigma^{-1}\sum_{k=1}^{n}(\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^t\right\}}}}, \qquad \text{hvilket skulle bevises.}
\end{aligned}
$$

Oppgave 3c

Her kan vi definere en matrise A der

$$A = \Sigma^{-1}\hat{\Sigma} \quad \text{der} \quad \hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^t .$$

Egenverdiene til A er gitt ved $A\mathbf{e}_i = \lambda_i \mathbf{e}_i, \quad i = 1, \ldots, d$, slik at

$$|A| = \prod_{i=1}^{d} \lambda_i \text{ og } \operatorname{Tr} A = \sum_{i=1}^{d} \lambda_i \text{ (kan vises).}$$

Likelihoodfunksjonen kan derved skrives som

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n | \Sigma) = \frac{1}{(2\pi)^{\frac{nd}{2}}} \left(\frac{|A|}{|\hat{\Sigma}|} \right)^{\frac{n}{2}} e^{-\frac{n}{2}\operatorname{Tr} A} = \frac{(\prod_{i=1}^{d} \lambda_i)^{\frac{n}{2}}}{(2\pi)^{\frac{nd}{2}} |\hat{\Sigma}|^{\frac{n}{2}}} e^{-\frac{n}{2}\sum_{i=1}^{d} \lambda_i} .$$

Oppgave 3d

Med $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)^t$ kan log-likelihoodfunksjonen nå skrives som

$$\mathscr{L}(\boldsymbol{\lambda}) = \frac{n}{2}\sum_{i=1}^{d} \ln \lambda_i - \frac{n}{2}\sum_{i=1}^{d} \lambda_i - \frac{nd}{2}\ln(2\pi) - \frac{n}{2}\ln|\hat{\Sigma}|,$$

slik at

$$\frac{\partial \mathscr{L}(\boldsymbol{\lambda})}{\partial \lambda_k} = \frac{n}{2\lambda_k} - \frac{n}{2} = 0 \quad \text{for } k = 1, \cdots, d.$$

Dette gir løsningen

$$\underline{\underline{\lambda_1 = \lambda_2 = \cdots = \lambda_d = 1}}.$$

Oppgave 3d (forts.)

Siden $\hat{\Sigma} = \Sigma A$, gir multiplikasjon med egenvektorene til A følgende

$$\hat{\Sigma}\boldsymbol{e}_k = \Sigma A \boldsymbol{e}_k, \quad k = 1, \ldots, d$$
$$\Downarrow$$
$$\hat{\Sigma}\boldsymbol{e}_k = \Sigma \lambda_k \boldsymbol{e}_k, \quad k = 1, \ldots, d$$
$$\Downarrow$$
$$\hat{\Sigma}\boldsymbol{e}_k = \Sigma \boldsymbol{e}_k, \quad (\text{siden } \lambda_k = 1)$$
$$\Downarrow$$
$$\hat{\Sigma} = \Sigma \quad (\text{siden alle } \boldsymbol{e}_k \neq 0)$$

Dette viser at $\hat{\Sigma}$ er maksimum-likelihoodløsningen for Σ, hvilket skulle bevises.

Oppgave 4

Likningssystemet for maksimum-likelihood (ML) estimatet av parametervektoren $\boldsymbol{\theta}$ i en gitt tetthetsfunksjon $p(\boldsymbol{x}|\boldsymbol{\theta})$ er

$$\sum_{k=1}^{n} \nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{x}_k \mid \boldsymbol{\theta}) = 0,$$

der $\boldsymbol{x}_k$, $k = 1, \ldots, n$ er et sett av treningssampler.

Her skal vi finne ML-estimatet av parameteren μ i log-normalfordelingen

$$p(x|\mu) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0$$

der parameteren σ er kjent.

Oppgave 4 (forts.)

Innsetting i likningssystemet gir

$$\sum_{k=1}^{n} \nabla_{\mu} \ln p(x_k \mid \mu) = \sum_{k=1}^{n} \frac{d}{d\mu} \left\{ -\frac{(\ln x_k - \mu)^2}{2\sigma^2} - \ln\left(x_k \sigma \sqrt{2\pi}\right) \right\} = 0,$$

der gradienten i dette tilfellet reduseres til den deriverte med hensyn på parameteren μ. Dette gir likningen

$$\sum_{k=1}^{n} \left\{ -2\frac{(\ln x_k - \mu)}{2\sigma^2} \right\} (-1) = 0,$$

som gir

$$\sum_{k=1}^{n} (\ln x_k - \mu) = 0, \text{ slik at } \sum_{k=1}^{n} \ln x_k = n\mu.$$

Parameterestimatet blir derved

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} \ln x_k,$$

hvilket skulle vises.