Øvingsoppgaver
TEK5020/9020 – Mønstergjenkjenning
Del 2 – Parametriske metoder

Høsten 2023

Oppgave 1

La egenskapen x være eksponentialfordelt slik at

$$p(x|\theta) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & \text{ellers.} \end{cases}$$

a) Skisser $p(x|\theta)$ som funksjon av x for en fast verdi av parameteren θ, der $\theta > 0$.
b) Skisser $p(x|\theta)$ som funksjon av θ for en fast verdi av egenskapen x.
c) Vis at maksimum likelihood estimatet for θ er gitt ved

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

der $x_1, x_2, \ldots, x_n$ er innbyrdes uavhengige sampler trukket fra fordelingen $p(x|\theta)$.

Oppgave 2

La $\boldsymbol{x}$ være en binær (0,1) egenskapsvektor med en multivariat Bernoullifordeling gitt ved

$$P(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i},$$

der $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^t$ er en parametervektor der θ_i er sannsynligheten for at $x_i = 1$. Vis at maksimum likelihood estimatet for $\boldsymbol{\theta}$ er

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

der $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ er treningssamplene.

Oppgave 3

La $p(\boldsymbol{x}|\Sigma) \sim N(\boldsymbol{\mu},\Sigma)$ der $\boldsymbol{\mu}$ er kjent og Σ er ukjent. Vist at maksimum likelihood estimatet for Σ er gitt ved

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

ved å gjennomføre følgende trinn:

a) Vis at $\boldsymbol{a}^t A\,\boldsymbol{a} = \mathrm{Tr}\{A\,\boldsymbol{a}\boldsymbol{a}^t\}$, der *tracen* Tr til en matrise er summen av diagonalelementene.

b) Vis at likelihoodfunksjonen kan skrives på formen

$$p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n|\Sigma) = \frac{1}{(2\pi)^{nd/2}}|\Sigma^{-1}|^{n/2}\exp\left[-\frac{1}{2}\mathrm{Tr}\{\Sigma^{-1}\sum_{k=1}^{n}(\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^t\}\right].$$

c) La $A = \Sigma^{-1}\hat{\Sigma}$ og $\lambda_1,\ldots,\lambda_d$ være egenverdiene til A, og vis at dette leder til

$$p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d|\Sigma) = \frac{1}{(2\pi)^{nd/2}|\hat{\Sigma}|^{n/2}}(\lambda_1\ldots\lambda_d)^{n/2}\exp\left[-\frac{n}{2}(\lambda_1+\cdots+\lambda_d)\right].$$

d) Gjennomfør beviset ved å vise at maksimum likelihood oppnås ved å velge $\lambda_1 = \lambda_2 = \cdots = \lambda_d = 1$.

Oppgave 4

Maksimum-likelihood (ML) estimatet av parametervektoren $\boldsymbol{\theta}$ i en gitt tetthetsfunksjon $p(\boldsymbol{x}|\boldsymbol{\theta})$ kan finnes ved å løse likningssystemet

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

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

Bruk likningssystemet til å vise at ML-estimatet av parameteren μ i log-normalfordelingen

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

der parameteren σ er kjent, er gitt ved

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

der $x_1,\ldots,x_n$ er et sett av univariate (endimensjonale) treningssampler.