Workshop 1

den 28 oktober 2024 20:49

1.1.1 MLE Categorical

$X_1, \dots X_N$ i.i.d. $\text{Cat}(\theta = [\theta_1, \dots, \theta_K])$

$$\underset{\theta, \sum_k \theta_k = 1}{\text{argmax}} \log p(X|\theta)$$, apply Lagrange multipliers

$$L = \log p(X|\theta) + \lambda\left(\sum_k \theta_k - 1\right)$$

$$\frac{\partial L}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}\left[\sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{1}\{x_n = k\} \log \theta_k - \lambda \sum_{k=1}^{K}(\theta_k - 1)\right] =$$

$= \{$ derivative can be pushed inside sums as it is a linear operator. $\}$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{1}\{x_n = k\} \underbrace{\frac{\partial}{\partial \theta_j}\log \theta_k}_{= \frac{1}{\theta_j}} - \lambda\left(\sum_{k=1}^{K} \frac{\partial}{\partial \theta_j}\theta_k\right) =$$

$= \{$ since $\frac{\partial}{\partial \theta_j} f(\theta_k) = 0$ for all $k \neq j$, the only non-zero term in the $\sum_k$-terms are when $k = j$. E.g. $\sum_{k=1}^{K} \frac{\partial}{\partial \theta_j}\theta_k = \frac{\partial}{\partial \theta_j}\theta_j = 1$ $\}$

$$= \sum_{n=1}^{N} \mathbb{1}\{x_n = j\} \frac{1}{\theta_j} - \lambda \cdot 1 \overset{!}{=} 0$$

Notation for we set it equal to 0

$$\Leftrightarrow \frac{1}{\theta_j} \underbrace{\sum_{n=1}^{N} \mathbb{1}\{x_n = j\}}_{\equiv N_j} = \frac{N_j}{\theta_j} = \lambda$$

$$\frac{\partial L}{\partial \lambda} = \sum_k \theta_k - 1 = \sum_k \frac{N_k}{\lambda} - 1 \overset{!}{=} 0 \Leftrightarrow \sum_k N_k - \lambda = 0$$

$$\Leftrightarrow \lambda = \sum_k N_k = N \Rightarrow \theta_j^{MLE} = \frac{N_j}{N}$$

1.1.2 MAP Dirichlet-Categorical

$$\theta \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$$

$$\underset{\theta, \sum_k \theta_k}{\text{argmax}} \log p(X|\theta) p(\theta)$$

$$L = \sum_{k=1}^{K} \left(\sum_{n=1}^{N} \mathbb{1}\{x_n = k\} + \alpha_k - 1 \right) \log \theta_k + \lambda \left(\sum_k \theta_k - 1 \right)$$

$\sum_k$ -terms only non-zero when $k=j$ as in the MLE case

$$\frac{\partial L}{\partial \theta_j} = \left(\sum_{n=1}^{N} \mathbb{1}\{x_n = j\} + \alpha_j - 1\right) \frac{\partial \log \theta_j}{\partial \theta_j} - \lambda \cdot 1$$

$$= (N_j + \alpha_j - 1)\frac{1}{\theta_j} - \lambda \overset{!}{=} 0 \iff \frac{(N_j + \alpha_k - 1)}{\theta_j} = \lambda$$

$$\frac{\partial L}{\partial \lambda} = \sum_k \theta_k - 1 = \sum_k \frac{(N_k + \alpha_k - 1)}{\lambda} - 1 \overset{!}{=} 0$$

$$\iff \lambda = \sum_k (N_k + \alpha_k - 1) = N + \sum_k \alpha_k - K$$

$$\Rightarrow \theta_j^{MAP} = \frac{N_j + \alpha_j - 1}{N + \sum_k (\alpha_k - 1)}$$

Normal with known precision τ

$$x_n | \mu, \tau \sim \mathcal{N}\left(\mu, \frac{1}{\tau}\right) = \frac{\sqrt{\tau}}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{\tau}{2}(x_n - \mu)^2\right\}$$

MLE

$$\arg\max_{\mu} \left[\log\left(\prod_{n=1}^{N} \frac{\sqrt{\tau}}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{\tau}{2}(x_n - \mu)^2\right\}\right)\right]$$

$$= \underset{\mu}{\arg\max} \underbrace{\left[\sum_{n=1}^{N} \frac{1}{2}\log\tau - \frac{1}{2}\log 2\pi - \frac{\tau}{2}(x_n-\mu)^2 \right]}_{\equiv \ell}$$

$$\frac{\partial \ell}{\partial \mu} = \frac{\partial}{\partial \mu} \left[\frac{N}{2}\log\tau - \frac{N}{2}\log 2\pi - \frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2 \right]$$

$$= -\frac{\tau}{2}\sum_{n=1}^{N}\frac{\partial}{\partial\mu}(x_n-\mu)^2 = \tau\sum_{n=1}^{N}(x_n-\mu) = \tau\left(-\mu N + \sum_{n=1}^{N} x_n\right)$$

$$\stackrel{!}{=} 0 \implies \mu_{MLE} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

1.2.3 Posterior derivation

This time, let's use the log pdf:s. "Identifying the distribution" technique

$$\log p(\mu) = \mathcal{N}\left(\mu_0, \frac{1}{\lambda}\right) = \frac{1}{2}\log\lambda - \frac{1}{2}\log 2\pi - \frac{\lambda}{2}(\mu - \mu_0)^2 \qquad (1)$$

$$\log p(\mu \mid X) = \log p(X \mid \mu) + \log p(\mu) - \log p(X)$$

$$\stackrel{+}{=} \log p(X \mid \mu) + \log p(\mu) = \frac{N}{2}\log\tau - \frac{N}{2}\log 2\pi - \frac{\tau}{2}\sum_{n=1}^{N}(x_n - \mu)^2$$

$$+ \frac{1}{2}\log\lambda - \frac{1}{2}\log 2\pi - \frac{\lambda}{2}(\mu - \mu_0)^2 \qquad (2)$$

We anticipate $p(\mu \mid X)$ same parametric family as $p(\mu)$, so we try to rewrite (*) in same form as (1)

First, note

$$(1) \stackrel{+}{=} -\frac{\lambda}{2}(\mu - \mu_0)^2 = -\frac{\lambda}{2}(\mu^2 - 2\mu\mu_0 + \mu_0^2) \stackrel{+}{=} \underline{-\frac{\lambda}{2}(\mu^2 - 2\mu\mu_0)}$$

we want (*) the final expression on this form!

$$(2) \stackrel{+}{=} -\frac{\tau}{2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{\lambda}{2}(\mu - \mu_0)^2 =$$

$$= -\frac{\tau}{2}\left(\sum_n x_n^2 - 2\underbrace{\sum_n x_n}_{=N\cdot\bar{x}}\mu + N\mu^2\right) - \frac{\lambda}{2}\left(\mu^2 - 2\mu\mu_0 + \mu_0^2\right)$$

$$\stackrel{+}{=} -\frac{\tau}{2}\left(-2\mu N\bar{x} + \mu^2\cdot N\right) - \frac{\lambda}{2}\left(\mu^2 - 2\mu_0\mu\right)$$

{ Gather terms related to μ^2 and μ, to get on same form as (*) }

$$= -\frac{\mu^2}{2}\underbrace{\left(N\cdot\tau + \lambda\right)}_{\equiv \lambda'} + \mu\left(\mu_0\lambda + N\cdot\bar{x}\cdot\tau\right)$$

$$\mu_0' \equiv \frac{1}{\lambda'}\cdot\left(\mu_0\lambda + N\bar{x}\tau\right)$$

$$= -\frac{\mu^2}{2}\cdot\lambda' + \mu\cdot\mu_0'\cdot\lambda' = -\frac{\lambda'}{2}\left(\mu^2 - 2\mu\cdot\mu_0'\right)$$

same form as (*) but with λ' and μ_0'

$$\Rightarrow p(\mu|X) = \mathcal{N}\left(\frac{\mu_0\lambda + \tau N\cdot\bar{x}}{N\cdot\tau+\lambda}, (N\cdot\tau+\lambda)^{-1}\right)$$