Aufgabe 3 (statistical model)

Montag, 4. April 2022 16:13

# ① Biased Coin Flipping

1.1

$$P(D|\theta) = \ell(\theta) = \theta^{k} \cdot (1-\theta)^{n-k}$$

$$= Beta(\theta|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \theta^{a-1} \cdot (1-\theta)^{b-1}$$

mit $a = a+k$, $b = b+N-k$ — $k$: ~~heads~~ , $N$: trials

$$\hat{\theta}_{MAP} = \arg\max_{\theta} Beta(\theta|a+k, b+N-k)$$

$$= \frac{\Gamma(a+k+b+N-k)}{\Gamma(a+k)\Gamma(b+N-k)} \cdot \theta^{a+k-1} \cdot (1-\theta)^{b+N-k-1}$$

mit $\Gamma(x) := \int_0^\infty u^{x-1} e^{-u} du.$

argmin $-\log(Beta)$

$$\hookrightarrow NNL(\theta) = -(a+k-1)\cdot\log(\theta) - (b+N-k-1)\cdot\log(1-\theta)$$

(Konstante ignorieren) $\left(\frac{\Gamma(a+N)}{\Gamma(a)\Gamma(k)}\right)$

$$\frac{\partial NNL(\theta)}{\partial\theta} = \frac{-(a+k-1)}{\theta} + \frac{(b+N-k-1)}{1-\theta} = 0$$

$$-(a+k-1)(1-\theta) + (b+N-k-1)\theta = 0$$

$$1-a-k-\theta+\theta a+\theta k+b\theta+N\theta-k\theta-\theta = 0 \qquad |+a|+k|-1$$

$$-2\theta+\theta a+\theta b+N\theta = a+k-1$$

$$a+b+N-2 = \frac{a+k-1}{\theta}$$

$$\hat{\theta}_{MAP} = \frac{k+a-1}{N+a+b-2}$$

2.1 $D = (x^{(1)}, \dots, x^{(N)})$

$$P(D|\theta) = \prod_{i=1}^{N} P(x^{(i)}|\theta) = \ell(\theta)$$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \ell(\theta) \rightarrow \arg\max_{\theta} NLL(\theta) \qquad \text{mit } NLL(\theta) = -\log(\ell(\theta))$$

## 2 ML for the Gaussian

Given data $\mathcal{D} = (\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)})$ assumed to be independently drawn from a multivariate Gaussian distribution with dimension $p$.

$x^{(i)} \in \mathbb{R}^p$

1. Give the formula of the log likelihood function.
2. Let $p = 1$, derive the maximum likelihood solution.

$$\sqrt{2\pi^p |K|} = \sqrt{2\pi^p} \cdot \sqrt{|K|}$$

**Solutions**

1. $p$-dimensional Gaussian

$$\log\left\{ \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{K}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{K}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \right\}$$

Taking logarithm

$$-\frac{p}{2}\log\pi - \frac{1}{2}\log|\boldsymbol{K}| - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{K}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

2. Let $p = 1$:

$$\hat{\mu}_{ML} = \arg\min_{\mu} \sum_{i=1}^{N} \left\{ \frac{1}{2}\log\pi + \frac{1}{2}\log\sigma^2 + \frac{1}{2\sigma^2}(x^{(i)} - \mu)^2 \right\}$$

then

$$\hat{\mu}_{ML} = \arg\min \frac{N}{2}\log\pi + N\log\sigma + \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x^{(i)} - \mu)^2$$

Set the derivative to zero: $(*)$ $\frac{1}{\partial\mu}$ $\frac{1}{2\sigma^2}$ $\Sigma$

$$\frac{\partial}{\partial\mu}\left(\frac{N}{2}\log\pi + N\log\sigma + \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x^{(i)} - \mu)^2\right)^{(*)} = 0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}(x^{(i)} - \mu) = 0 \iff \sum_{i=1}^{N}(x^{(i)} - \mu) = 0$$

$$\sum_{i=1}^{N}(x^{(i)}) = N\mu \qquad = \sum_{i=1}^{N} x^{(i)} - \underbrace{\sum_{i=1}^{N}\mu}_{N\cdot\mu}$$

So finally

$$\boxed{\hat{\mu}_{ML} = \frac{1}{N}\sum_{i=1}^{N} x^{(i)}}$$

For $\sigma$:

$$\arg\min \frac{N}{2}\log\pi + N\log\sigma + \frac{1}{2\sigma^2} \quad (x^{(i)} - \mu)^2$$

Set the derivative to zero:

$$\frac{N}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^{N}(x^{(i)} - \mu)^2 = 0$$

hence

$$N - \frac{1}{\sigma^2}\sum_{i=1}^{N}(x^{(i)} - \mu)^2 = 0.$$

we obtain

$$\hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)} - \mu)^2.$$

2.2 $p = 1$ (dimension) $\underline{\theta} = \theta_1$

$$l(\theta) = \theta^k \cdot (1-\theta)^{n-k}$$

$$\hat{\theta}_{ML} = \arg\max_{\theta} -\log(l(\theta)) = \arg\max -k\cdot\log(\theta) - (n-k)\cdot\log(1-\theta)$$

$$\hat{\theta}_{ML} = \frac{\partial NNL(\theta)}{\partial\theta} = -\frac{k}{\theta} + \frac{n-k}{1-\theta} = 0 \qquad |\cdot\theta \;|\cdot(1-\theta)$$

$$-k(1-\theta) + (n-k)\cdot\theta = 0$$

$$\cancel{\theta k} - k + \theta n - \cancel{\theta k} = 0 \qquad |+k \;|:n$$

$$\underline{\underline{\hat{\theta}_{ML} = \frac{k}{n}}}$$