CDS: Machine Learning

Dirren van Vlijmen
Gerwin de Kruijf
Olivier Brahma

Mackay 3.12

Options

$W_1 W_2$

$W_1 B_2$

$B_1 W_2$

~~B B~~ ← impossible

$$P(W_2|W_1) = \frac{P(W_1, W_2)}{P(W_1)} = \frac{1/3}{2/3} = \frac{1}{2}$$

Extra opgave 3.1

a $N=2$ options $s = \{hh, ht, th, tt\}$

$H_0: p_h = 0.5$ $H_1: p_h \neq 0,5$

assume equal probabilities

$P(H_0) = 0.5$ $P(H_1) = 0.5$

$N_H = 0$ $s = tt$

$P(s=tt|H_0) = 0.5^2 = 0,25$

$P(s=tt|H_1) = 1^2 = 1$

$$P(s=tt) = P(s=tt|H_0)P(H_0) + P(s=tt|H_1)P(H_1)$$
$$= 0,25 \cdot 0,5 + 1 \cdot 0,5$$
$$= 0,625$$

$$P(H_0 \mid s=tt) = \frac{p(s=tt \mid H_0)\, p(H_0)}{p(s=tt)}$$

$$= \frac{0,25 \cdot 0,5}{0,625}$$

$$= 0,2$$

$$p(H_1 \mid s=tt) = \frac{p(s=tt \mid H_1)\, p(H_1)}{p(s=tt)}$$

$$= \frac{1 \cdot 0,5}{0,625}$$

$$= 0,8$$

$N_H = 2$

$s = hh$

~~exactly~~ equal to $NH=0$

$P(H_0 \mid s=hh) = 0,2$

$P(H_1 \mid s=hh) = 0,8$

$N_H = 1$

$s = \{ht, th\}$

$$P(s=\{ht, th\} \mid H_0) = 0,5^2 + 0,5^2$$

$$= 0.5$$

$P(s=\{ht, th\} \mid H_1) < 0,5$ for any $H_1: p_h \neq 0,5$

$$P(s=\{ht, th\}) = P(s=\{ht, th\} \mid H_0) P(H_0) + P(s=\{ht, th\} \mid H_1) P(H_1)$$

$$= 0{,}5 \cdot 0{,}5 + P(s=\{ht, th\} \mid H_1) \cdot 0{,}5$$

$$< 0{,}25 + 0{,}25$$

$$< 0{,}5$$

$$P(H_0 \mid s=\{ht, th\}) = \frac{P(s=\{th, ht\} \mid H_0) P(H_0)}{P(s=\{ht, th\})}$$

$$> \frac{0{,}5 \cdot 0{,}5}{0{,}5}$$

$$> \frac{0{,}25}{0{,}5}$$

$$> 0{,}5$$

$$P(H_1 \mid s=\{ht, th\}) = \frac{P(s=\{th, ht\} \mid H_0) P(H_0)}{P(s=\{ht, th\})}$$

$$= \frac{P(s=\{th, ht\} \mid H_0) P(H_1)}{(P(s=\{ht, th\} \mid H_0) P(H_0) + P(s=\{ht, th\} \mid H_1) P(H_1))}$$

$$P(H_1 \mid s=\{ht,th\}) = \frac{P(s=\{ht,th\} \mid H_1)\cdot 0{,}5}{(0{,}5\cdot 0{,}5 + P(s=\{ht,th\} \mid H_1)\cdot 0{,}5)}$$

$$= \frac{P(s=\{ht,th\} \mid H_1)}{0{,}5 + P(s=\{ht,th\} \mid H_1)}$$

$$< \frac{0{,}5}{0{,}5+0{,}5}$$

$$< 0{,}5$$

b An unfair coin will prefer one result (side) over the other. Heads-Heads or Tails Tails will more likely be unfair. A fair coin will prefer a split result, such as Heads-tails or reverse. In the end, It is important to realise that $N=2$, and the coin should be tossed more to draw meaningful Conclusions.

28.1 (mackay)

we have to calculate the evidence, which is denoted as: $P(D|H)$

For $H = H_0$:

$$P(D|H_0) = \prod_{i=1}^{5} p(x_i|H_0) \quad \text{since all } x\text{'s are i.d.d.}$$
$$= \prod_{i=1}^{5} 1/2 = (1/2)^5 = 1/32 = 0.031$$

possible range of x. 1/possible

For $H = H_1$:

$$P(D|H_1) = \int_{-1}^{1} dm \; p(D|H_1, m) \, p(m|H_1) = 1/2$$

$$= \int_{-1}^{1} dm \prod_{i=1}^{5} \frac{1}{2}(1 + m x_i) \cdot \frac{1}{2}$$

$$= \frac{1}{64} \int_{-1}^{1} dm \, (1 + m0.3)(1 + 0.5m)(1 + 0.7m)(1 + 0.8m)(1 + 0.9m)$$

" This integral will be done numerically "
↳ we used the file "28.1 numerical integration.py"

$= 4.929/64 = 0{,}077\ldots$

1.3 Chapter 27

1 a)

$$p(\lambda \mid r) \propto p(r \mid \lambda) \cdot p(\lambda)$$

$$p(\lambda \mid r) \propto \exp(-\lambda)\frac{\lambda^r}{r!} \cdot \frac{1}{\lambda}$$

$$\propto \exp(-\lambda)\frac{\lambda^{r-1}}{r!}$$

$$Q = -\ln p(\lambda \mid r)$$

$$= -\left(\ln(e^{-\lambda}) + \ln(\lambda^{r-1}) - \ln(r!)\right)$$

$$\propto \lambda - (r-1)\log(\lambda)$$

$$\frac{\partial Q}{\partial \lambda} = 1 - \frac{(r-1)}{\lambda}$$

$$0 = 1 - \frac{r-1}{\lambda} = r-1 \rightarrow \lambda_{MAP}$$

$$A = \left.\frac{\partial^2 Q}{\partial \lambda^2}\right|_{\lambda = \lambda_{MAP}} = \frac{r-1}{(r-1)^2} = \frac{1}{r-1}$$

$$p(\lambda) = \sqrt{\frac{A}{2\pi}} \exp\left\{-\frac{A}{2}(\lambda - \lambda_{map})^2\right\}$$

$$= \frac{1}{\sqrt{2\pi(r-1)}} \exp\left(-\frac{1}{2(r-1)}(\lambda - (r-1))^2\right)$$

$$= N(\lambda ; r-1, r-1)$$

1.3 Chapter 27

1. b)

$$p(y) = \frac{dy}{d\lambda}^{-1} \cdot p(\lambda)$$

$$p(y) = \frac{1}{\lambda}^{-1} \cdot \frac{1}{\lambda} = \lambda \cdot \frac{1}{\lambda} = 1$$

c) $$p(\log \lambda | r) \propto p(r|\lambda) \cdot 1$$
$$\propto \exp(-\lambda)\frac{\lambda^r}{r!}$$

$$Q = -\ln p(\log \lambda | r)$$
$$= -(\ln(e^{-\lambda}) + \ln(\lambda^r) - \ln(r!))$$

$$\propto \lambda - r\log(\lambda) = e^y - ry$$

$$\frac{\partial Q}{\partial y} = e^y - r$$

$$e^y - r = 0$$
$$y_{map} = \log(r)$$

$$A = \left.\frac{\partial^2 Q}{\partial y^2}\right|_{y=y_{map}} = \left.e^y\right|_{y=y_{map}} = e^{\log(r)} = r$$

$$P_Z(y) = \sqrt{\frac{A}{2\pi}} \exp\left\{-\frac{A}{2}(y-y_{map})^2\right\}$$

$$= \frac{1}{\sqrt{2\pi \cdot 1/r}} \exp\left\{-\frac{1}{2 \cdot 1/r}(y-y_{map})^2\right\}$$

fill in $y = \log(\lambda)$ $\quad N(\log\lambda; \log r; 1/r)$
$y_{map} = \log(r)$

CDS: Machine Learning

~~2014~~ 28.1

a.) $p(D|w_0, w_1) = \prod_{i=1}^{N} \mathcal{N}(t_i | w_0 + w_1 x_i, \sigma^2)$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \prod_{i=1}^{N} \exp\left(\frac{-1}{2\sigma^2}(t_i - w_0 - w_1 x_i)^2\right)$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\frac{-1}{2\sigma^2}\sum_i (t_i^2 + w_0^2 + w_1^2 x_i^2 - 2t_i w_0 + 2w_0 w_1 x_i - 2w_1 t_i x_i)\right)$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\frac{-1}{2\sigma^2}\left(\sum_i t_i^2 + w_0^2 N + w_1^2 \sum_i x_i^2 - 2w_0 \sum_i t_i - 2w_1 \sum_i t_i x_i\right)\right)$$

Since we can use that $\sum_i x_i = 0$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\frac{-1}{2\sigma^2}\sum_i t_i^2\right) \exp\left(\frac{-1}{2\sigma^2}\left(w_0^2 N - 2w_0 \sum_i t_i\right)\right) \exp\left(\frac{-1}{2\sigma^2}\left(w_1^2 \sum_i x_i^2 - 2w_1 \sum_i t_i x_i\right)\right)$$

b.) $$\frac{P(D|H_2)}{P(D|H_1)} = \frac{\iint dw_0\, dw_1\, p(D|w_1, w_0, H_2)\, p(w_0, w_1 | H_2)}{\iint dw_0\, dw_1\, p(D|w_0, w_1, H_2)\, p(w_0, w_1 | H_1)}$$

*). we know that $p(D|w_1, w_0, H_i)$ is the same for both and described by a.)

*) $p(w_0, w_1 | H_2) = \mathcal{N}(w_0 | 0, 1)\, \mathcal{N}(w_1 | 0, 1)$

*) $p(w_0, w_1 | H_1) = \mathcal{N}(w_0 | 0, 1) \big|_{w_1 = 0}$

Since within the expression of a.) the functions factorize for w_0 and w_1 and w_0 dependence is the same for both models, all terms independent of both w_0 and w_1 and the terms depended on w_0 drop out.

Further all terms with w_1 in a.) have w_1 in the exponent → for H_1: $w_1 = 0$ → ~~1~~ denominator stays 1. in the nominator we are left with the following z.o.z.

$$\frac{P(D|H_2)}{P(D|H_1)} = \frac{1}{\sqrt{2\pi}} \int dw_1 \exp\left(-\frac{w_1^2}{2\sigma^2}\sum_i x_i^2 + \frac{w_1}{\sigma^2}\sum_i x_i t_i - \frac{w_1^2}{2}\right)$$

where we filled in $\mathcal{N}(w_1|0,1)$

$$= \frac{1}{\sqrt{2\pi}} \int dw_1 \exp\left(\frac{-w_1^2 N}{2}\left(\frac{\langle x^2\rangle}{\sigma^2} + \frac{1}{N}\right) + \frac{w_1 N}{\sigma^2}\langle xt\rangle\right)$$

which is ~~the~~ a standard integral with:

$a = N\left(\frac{\langle x^2\rangle}{\sigma^2} + \frac{1}{N}\right)$ and $J = \frac{N}{\sigma^2}\langle xt\rangle$

Thus we have:

$$= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{2\pi}}{\left(N\left(\frac{\langle x^2\rangle}{\sigma^2} + \frac{1}{N}\right)\right)^{1/2}} \exp\left(\frac{N^2\langle xt\rangle^2}{2(\sigma^2)^2\left(N\left(\frac{\langle x^2\rangle}{\sigma^2} + \frac{1}{N}\right)\right)}\right)$$

$$= \frac{1}{\left(N\left(\frac{\langle x^2\rangle}{\sigma^2} + \frac{1}{N}\right)\right)^{1/2}} \exp\left(\frac{N\langle xt\rangle^2}{2\left(\langle x^2\rangle\sigma^2 + \frac{(\sigma^2)^2}{N}\right)}\right)$$

c.) $\sigma^2 = 1 \rightarrow \left(N\left(\langle x^2\rangle + \frac{1}{N}\right)\right)^{1/2} \exp\left(\frac{N\langle xt\rangle^2}{2\left(\langle x^2\rangle + \frac{1}{N}\right)}\right)$

if $\langle xt\rangle^2 = \frac{\log(N)}{N}$ we find

$$\frac{\sqrt{N}}{\left(\langle x^2\rangle + \frac{1}{N}\right)^{1/2}} \exp\left(\left(2\left(\langle x^2\rangle + \frac{1}{N}\right)\right)^{-1}\right)$$

~~if $N \to \infty$~~: if N becomes large we can neglect $\frac{1}{N}$

$$\rightarrow \frac{\sqrt{N}}{\sqrt{2\langle x^2\rangle}} \exp\left(\frac{1}{2\langle x^2\rangle}\right)$$

thus $N \rightarrow$ large gives > 1 and thus we prefere model H_2.

if $\langle xt\rangle^2 < \frac{\log(N)}{N}$ we will find a negative dependence of N and this < 1, which shows a preference for H_1

28.2

a.)

$$P(D|H_0) = \prod_{i=1}^{n} P(x_i|H_0) \cdot CF$$

where CF denotes the combinatorial factor we need since we disregarded the order!

$P(x_i|H_0) = \frac{1}{k}$, since we assume a fair dice.

$$\rightarrow P(D|H_0) = \left(\frac{1}{k}\right)^n \cdot CF = \frac{n!}{n_1! \dots n_k! (k)^n}$$

28.2

b.) $P(D|H_1) = \int d\vec{p}\; P(D|\vec{p}, H_1)\, P(\vec{p}|H_1)$

$P(D|\vec{p}, H_1) = C_F \prod_{i=1}^{k} (p_i)^{n_i}$ (C_F similar to a.))

$P(\vec{p}|H_1)$: uniform dirichlet distribution

we use 23.30 from Mackay with $\alpha m_i = 1$

and $\alpha\vec{m} \in \mathbb{R}^k$.

$\rightarrow p(\vec{p}|H_1) = \frac{1}{Z(\alpha\vec{m})} \prod_{i=1}^{k} p_i^{1-1}\; \delta(\sum_{i=1}^{k} p_i - 1)$

$Z(\alpha\vec{m}) = \frac{1}{(k-1)!}$ (see also special case dirichlet distribution from wikipedia)

$\rightarrow p(\vec{p}|H_1) = (k-1)!\; \delta(\sum_{i=1}^{k} p_i - 1)$

$\Rightarrow P(D|H_1) = C_F \int d\vec{p}\; (k-1)! \prod_{i=1}^{k} (p_i)^{n_i}\; \delta(\sum_{i=1}^{k} p_i - 1)$

which is similar to the I given within the question.

$\rightarrow I = C_F (k-1)!\, I_1$

where $I_1 = \int_{\text{simplex}} d\vec{p} \prod_{i=1}^{k} (p_i)^{n_i}\; \delta(\sum_{k=1}^{k} p_i - 1)$

Now we notice that this is just the normalization constant of a different dirichlet distribution, with

$\alpha m_i - 1 = n_i \quad \rightarrow \alpha m_i = n_i + 1$

$\rightarrow m_i = \frac{n_i + 1}{\alpha}$, and we need $\sum_i m_i = 1$

$\rightarrow \alpha = \sum_i (n_i + 1) = \sum_i n_i + k$

$\rightarrow m_i = \frac{n_i + 1}{\sum_i n_i + k}$

we know $I_1 = \prod_i \Gamma(\alpha m_i) / \Gamma(\alpha)$

$= \prod_i \frac{\Gamma(n_i + 1)}{\Gamma(\sum_i n_i + k)}$

$\Rightarrow P(D|H_1) = C_F (k-1)! \prod_i \frac{\Gamma(n_i + 1)}{\Gamma(\sum_i n_i + k)}$

$= \frac{n!}{n_1! \ldots n_k!} (k-1)! \frac{n_1! \ldots n_k!}{(n+k-1)!} = \frac{n!\,(k-1)!}{(n+k-1)!}$

28.2

c.)

Posterior probability given equal priors:

$$p(H_i \mid D) = \frac{p(D \mid H_i)\, p(H_i)}{p(D)} = \frac{p(D \mid H_i)\, p(H_i)}{p(D \mid H_i)\, p(H_i) + p(D \mid H_0)\, p(H_0)}$$

$p(H_i) = p(H_1) = p(H_0)$

$$\rightarrow \quad p(H_i \mid D) = \frac{p(D \mid H_i)}{p(D \mid H_0) + p(D \mid H_1)}$$

for numerical calculations see:

~~"28.2 extra-exercises.py"~~.

"28.2.py"