CDS: Machine Learning

Dirren van Vlijmen
Gerwin de Kruijf
Olivier Brahma

CDS: Machine Learning

2.10 (Mackay)

$A: \{B, W, W\}$

$B: \{B, B, W\}$

$D = B$ (data)

$$p(A \mid D=B) = \frac{p(D=B \mid A)\, p(A)}{p(D=B)}$$

$$= \frac{p(D=B \mid A)\, p(A)}{p(D=B \mid A)\, p(A) + p(D=B \mid B)\, p(B)}$$

$$= \frac{1/3 \cdot 1/2}{1/3 \cdot 1/2 + 2/3 \cdot 1/2} = \frac{1/3}{1/3 + 2/3} = 1/3 \quad \square,$$

1.1 Chapter 2

1. We consider 0 to be even

even = $\{0, 2, 4, 6, 8, 10\}$

$\mu = 0$ 0 black balls so probability even black balls = 1

$\mu = 10$ 10 black balls so probability even = 1

$\mu = 1 \dots 9$ 0, 2, 4, 6, 8, 10 can be picked

$$P(\mu \mid n_{B=\text{even}}, n) =$$

$$\frac{\sum_{x \in \text{even}} f_\mu^{\,x} (1-f_\mu)^{n-x}}{\sum_{\mu'=0}^{10} f_{\mu'}^{\,x} (1-f_{\mu'})^{n-x}}$$

python program to calculate

$P(\mu = 1 \mid n_{B=\text{even}}, n)$ = 0.46

$\mu = 2$ " = 0.51

$\mu = 3$ " = 0.54

$\mu = 4$ " = 0.55

$\mu = 5$ " = 0.55

$\mu = 6$ " = 0.55

$\mu = 7$ " = 0.54

$\mu = 8$ " = 0.51

$\mu = 9$ " = 0.46

Final Versions CDS: Machine learning

Dirren van Vlijmen
Gerwin de Kruif
Olivier Brahma

1.2

$$p(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

a.)

equal prob. $\forall \mu \rightarrow$ constant : C

$$p(\mu|x,\sigma) = \frac{p(x|\mu,\sigma)\, p(\mu|\sigma)}{p(x|\sigma)}$$

$$= \int d\mu \; p(x|\mu,\sigma)\, p(\mu|\sigma)$$

$\sigma = 1$

$$\rightarrow p(\mu|x,\sigma) = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2}\right)}{\frac{1}{\sqrt{2\pi}} \int d\mu \exp\left(\frac{-(x-\mu)^2}{2}\right)}$$

$$= \frac{\exp\left(\frac{-\mu^2}{2}\right)\exp(x\mu)}{\int d\mu \exp\left(\frac{-\mu^2}{2}\right)\exp(x\mu)}$$

We use the standard gaussian integral with an linear expont term, with $a = 1$ and $J = x$

$$\rightarrow = \frac{\exp\left(\frac{-\mu^2}{2}\right)\exp(x\mu)}{\sqrt{2\pi}\; \exp\left(\frac{x^2}{2}\right)} = \frac{\exp\left(\frac{-x^2}{2}\right)\exp\left(\frac{-\mu^2}{2}\right)\exp(x\mu)}{\sqrt{2\pi}}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\mu-x)^2\right) = \mathcal{N}(\mu|x, \sigma=1)$$

So μ is normally distributed about x with $\sigma = 1$.

b.) $p(\mu \mid x_1, \dots, x_N, \sigma = 1)$

$$= \frac{p(x_1, \dots, x_N \mid \mu, \sigma = 1)\, p(\mu \mid \sigma)}{\int d\mu\, p(x_1, \dots, x_N \mid \mu, \sigma = 1)\, p(\mu \mid \sigma)}$$

i.d.d.

$$\overset{\downarrow}{=} \frac{\prod_{i=1}^{N} p(x_i \mid \mu, \sigma = 1)}{\int d\mu \prod_{i=1}^{N} p(x_i \mid \mu, \sigma = 1)}$$

$*)$ $\prod_{i=1}^{N} p(x_i \mid \mu, \sigma = 1) = \left(\frac{1}{\sqrt{2\pi}}\right)^N \exp\left(\sum_{i=1}^{N} -(x_i - \mu)^2/2\right)$

$\rightarrow \int d\mu \prod_{i=1}^{N} p(x_i \mid \mu, \sigma = 1) = \left(\frac{1}{\sqrt{2\pi}}\right)^N \int d\mu \exp\left(\frac{-(\sum x_i)^2}{2}\right)$

$$* \exp\left(-\frac{\mu^2}{2} \underbrace{\sum_{i=1}^{N}}_{=N}\right) \exp\left(\sum_{i=1}^{N} x_i \mu\right)$$

we can, once again, use a standard integral.

with $a = N$ and $J = \sum_{i=1}^{N} x_i$

$\rightarrow * = \left(\frac{2\pi}{N}\right)^{1/2} \exp\left(\frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j\right)$

$\Rightarrow p(\mu \mid x_1, \dots, x_N, \sigma = 1) =$

$$\frac{\exp\left(\frac{-N\mu^2}{2}\right) \exp\left(\sum_{i=1}^{N} x_i \mu\right)}{\left(\frac{2\pi}{N}\right)^{1/2} \exp\left(\frac{-1}{2N}\left(\sum_{i=1}^{N} x_i \cancel{\mu}\right)^2\right)}$$

$$= \frac{\exp\left(\frac{N}{2}\left(-\left(\sum_{i=1}^{N} x_i\right)^2 \frac{1}{N^2} - \mu^2 + \sum_i x_i \mu / N\right)\right)}{\left(\frac{2\pi}{N}\right)^{1/2}}$$

$$= \left(\frac{N}{2\pi}\right)^{1/2} \exp\left(-\frac{N}{2}\left(\frac{1}{N}\sum_{i=1}^{N} x_i - \mu\right)^2\right)$$

$$= \mathcal{N}\left(\vec{x} \mid \frac{1}{N}\sum_{i=1}^{N} x_i = \mu,\ \sigma = \frac{1}{N}\right)$$

□

1.1 Chapter 2

3. a)

exp. family: $P(x \mid \theta) = h(x) \exp\left(\theta^T \cdot T(x) - A(\theta)\right)$

$$P(x \mid \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n \cdot |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) - \frac{1}{2}\log|\Sigma|\right)$$

- $h(x) = \frac{1}{(2\pi)^{n/2}} = (2\pi)^{-n/2}$

$$\exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) - \frac{1}{2}\log|\Sigma|\right)$$

$$= \exp\left(-\frac{1}{2}\left(\Sigma^{-1}xx^T - 2\mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu\right) - \frac{1}{2}\log|\Sigma|\right)$$

$$= \exp\left(-\frac{1}{2}\Sigma^{-1}xx^T + \mu^T\Sigma^{-1}x - \frac{1}{2}\mu^T\Sigma^{-1}\mu - \frac{1}{2}\log|\Sigma|\right)$$

$$\theta = \begin{bmatrix} \Sigma^{-1}\mu \\ -\frac{1}{2}\Sigma^{-1} \end{bmatrix} \qquad T(x) = \begin{bmatrix} x \\ xx^T \end{bmatrix}$$

- $A(\theta) = \frac{1}{2}\mu^T\Sigma^{-1}\mu + \frac{1}{2}\log|\Sigma|$

- $\theta^T \cdot T(x) = -\frac{1}{2}\Sigma^{-1}xx^T + \mu^T\Sigma^{-1}x$

$$-A(\theta) = -\frac{1}{2}\mu^T\Sigma^{-1}\mu - \frac{1}{2}\log|\Sigma|$$

Combine to get exponential family distribution

1.3

b.) The constraints are the "normal" constraints of the parameters, namely that they are equal to the mean definition of the mean and the variance. For the multivariate case this then becomes:

$$\mathbb{E}[\vec{x}] = \vec{\mu} \quad \text{and} \quad \mathbb{E}((\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T) = \Sigma$$

1.4

a.) $\int dz_1\, q_1(z_1) \log(q_1(z_1))$

$$= \int dz_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{\left(\frac{-z_1^2}{2\sigma_1^2}\right)} \Big(\underbrace{\log((2\pi\sigma_1^2)^{-1/2})}_{①} + \underbrace{\left(-\frac{z_1^2}{2\sigma_1^2}\right)}_{②} \Big)$$

*①: $-\dfrac{\log(\sqrt{2\pi\sigma_1^2})}{(2\pi\sigma_1^2)^{1/2}} \int dz_1 \exp\left(\dfrac{-z_1^2}{2\sigma_1^2}\right)$

which ~~is an~~ standard integral (gaussian):

$$= -\sqrt{\frac{2\pi\sigma_1^2}{2\pi\sigma_1^2}} \log(\sqrt{2\pi\sigma_1^2})$$

②: $\dfrac{-1}{\sqrt{2\pi\sigma_1^2}} \int dz_1\, e^{\left(-\frac{z_1^2}{2\sigma_1^2}\right)} \dfrac{z_1^2}{2\sigma_1^2}$

which is also a standard integral:

$$= -\frac{\sqrt{2\pi\sigma_1^2}}{\sqrt{2\pi\sigma_1^2}} \frac{\sigma_1^2}{2\sigma_1^2} = -\frac{1}{2}$$

$\rightarrow$ ① + ② $= -\log(\sqrt{2\pi\sigma_1^2}) - \frac{1}{2}$ □

For $q_2(z_2)$ the expression follows exactly symmetrically!

b.) $KL(q|p) = \int dz_1 \int dz_2\, q(z_1,z_2) \log\left(\dfrac{q(z_1,z_2)}{p(z_1,z_2)}\right)$

$$= \int dz_1 \int dz_2\, q(z_1) q(z_2) \Big(\underbrace{\log(q(z_1))}_{①} + \underbrace{\log(q(z_2))}_{②} - \underbrace{\log(p(z_1,z_2))}_{③} \Big)$$

①: $\underbrace{\int dz_1\, q(z_1) \log(q(z_1))}_{\text{see a.)}} \underbrace{\int dz_2\, q(z_2)}_{1 \text{ due to normalization}}$

②: $\underbrace{\int dz_2\, q_2(z_2) \log(q(z_2))}_{\text{see a.)}} \underbrace{\int dz_1\, q(z_1)}_{1 \text{ due to normalization.}}$

③: $\iint dz_1 dz_2\, q(z_1) q(z_2) \Big(\underbrace{\tfrac{1}{2}(a z_1^2 + a z_2^2}_{Ⓐ} + \underbrace{2 b z_1 z_2)}_{Ⓑ} \Big)$

with $Ⓐ_1$ above $a z_1^2$ and $Ⓐ_2$ above $a z_2^2$.

Ⓑ: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}$ even * even * odd $\rightarrow \int_{-\infty}^{\infty}$ odd $= 0$

~~$\int dz_1 \int dz_2$~~

$Ⓐ_1$: $\int dz_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{\left(\frac{-z_1^2}{2\sigma_1^2}\right)} \frac{1}{2} a z_1^2 \underbrace{\int dz_2 \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{\left(-\frac{z_2^2}{2\sigma_2^2}\right)}}_{1 \text{ via Normalization}}$

$= \frac{1}{2} a \sigma_1^2$ (since $E_q z_1 = 0$)

due to symmetry: $Ⓐ_2 = \frac{1}{2} a \sigma_2^2$

$\rightarrow$ ① + ② + ③ $= -1 - \log(\sqrt{2\pi\sigma_1^2}) - \log(\sqrt{2\pi\sigma_2^2}) + \frac{1}{2} a (\sigma_1^2 + \sigma_2^2)$

c.) $\frac{\partial}{\partial \sigma_1^2}(kL(q|p)) \overset{!}{=} 0$

$= \frac{1}{2}a$ [illegible] $- \frac{1}{\sqrt{2\pi\sigma_1^2}} \frac{1}{2\sqrt{\sigma_1^2}} \sqrt{2\pi}$

$= \frac{1}{2}a$ [illegible] $- \frac{1}{2}\frac{1}{\sigma_1^2} = 0$

$\rightarrow \frac{1}{\sigma_1^2} = a \rightarrow \sigma_1^2 = \frac{1}{a}$ □

d.) we notice 2 things:

1) $q_i(z_i)$ is the variational solution, and both $q_1(z_1)$ and p follow gaussian structure.

$\rightarrow q_i(z_i) \equiv p_i(z_i)$ where which is a gaussian!)

2.) $\int_{-\infty}^{\infty} dz_1\, p(z_1, z_2) \equiv P_2(z_2)$ (← since we integrate out ~~lgnage~~ z_1 or project onto z_2 so to say.

$\Rightarrow E_p z_1^2 = \iint dz_1 dz_2\, p(z_1, z_2)\, z_1^2$

via 2.) $\int dz_1\, p_1(z_1)\, z_1^2$ [illegible]

via 1.) $= \int dz_1\, q_1(z_1)\, z_1^2 = \sigma_1^2$, symmetry gives the same for $E_p z_2^2$

e.) $E_p z_1^2 = \iint dz_1 dz_2\, p(z_1, z_2)\, z_1^2$

complete the square with exp. of $p(z_1, z_2)$:

$az_1^2 + az_2^2 + 2bz_1z_2 = az_1^2 + \frac{b^2}{a}z_1^2 - \frac{b^2}{a}z_1^2 + az_2^2 + 2bz_1z_2$

$= z_1^2\left(a - \frac{b^2}{a}\right) + \left(\frac{b}{\sqrt{a}}z_1 + \sqrt{a}\,z_2\right)^2$

$\rightarrow E_p z_1^2$ ~~[illegible]~~

$= \int dz_1\, z_1^2 \exp\left(\frac{-1}{2}z_1^2\left(a - \frac{b^2}{a}\right)\right) \int dz_2 \exp\left(\left(\frac{b}{\sqrt{a}}z_1 + \sqrt{a}\,z_2\right)^2\right)$

no contribution to $E_p z_1^2 \rightarrow$ const.

→ The left over term is the variance of a gaussian with mean zero. The form of the exponent then gives $\rightarrow = \frac{1}{\left(a - \frac{b^2}{a}\right)} = \frac{1}{\frac{a^2 - b^2}{a}} = \frac{a}{a^2 - b^2}$ □

28.1 (mackay)

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

For $H = H_0$:

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

$$= \prod_{i=1}^{5} 1/2 = (1/2)^5 = 1/32 = 0.031$$

For $H = H_1$:

possible range of x

$$1/\sigma_{possible} = 1/2$$

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

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

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

" 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)

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

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

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

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

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

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

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

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

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

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

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

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

~~P(\lambda|r)~~

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) $p(\log \lambda \mid r) \propto p(r \mid \lambda) \cdot 1$

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

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

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

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

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

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

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

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

fill in $y = \log(\lambda)$
$y_{map} = \log(r)$

$$\mathcal{N}(\log \lambda ; \log r ; 1/r)$$

CDS: Machine Learning

~~2012~~ 28.1

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

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

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

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

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

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

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

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

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: $w_1 = 0$ → 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.) $\sigma^2 = 1 \rightarrow \frac{1}{\left(N\left(\langle x^2\rangle + \frac{1}{N}\right)\right)^{1/2}} \exp\left(\frac{N\langle xt\rangle^2}{2\left(\langle x^2\rangle + \frac{1}{N}\right)}\right)$

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

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

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

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

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

which is similar to the I given within the question.

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

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

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

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

$\rightarrow m_i = \frac{n_i + 1}{\alpha}$, 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 $I_1 = \prod_i \Gamma(\alpha m_i) / \Gamma(\alpha)$

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

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

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

28.2

c.)

Posterior probability given equal priors:

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

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

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

for numerical calculations see:

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

"28.2.py"

2.1 Perceptron

1. a)

$P \leq n$, sum is limited by $P-1$

change $n-1$ to $P-1$

$$C(P,n) = 2\sum_{i=0}^{P-1} \binom{P-1}{i} = 2 \cdot \sum_{i=0}^{P-1} \binom{P-1}{i} 1^i$$

use formula =

$$2 \cdot (1+1)^{P-1} = 2^1 \cdot 2^{P-1} = 2^P$$

1. b) $P = 2n$, use $2n$ in sum

$$C(2n,n) = 2\sum_{i=0}^{n-1} \binom{2n-1}{i}$$

$$= 2 \cdot \frac{1}{2}(1+1)^{2n-1}$$

$$= 2^{2n-1} = 2^{P-1}$$