CDS: Machine Learning

Dirren van Vlijven
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

$\mu = 1 \ldots 9$ 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

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

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

$\sigma = 1$

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

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

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

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

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

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

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

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

we can, once again, use a standard integral.

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

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

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

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

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

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

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

$\square$.

1.1 Chapter 2

3. a)

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

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

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

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

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

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

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

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

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

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

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

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

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

which is an standard integral (gaussian):

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

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

which is also a standard integral:

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

$$\rightsquigarrow ① + ② = -\log(\sqrt{2\pi\sigma_1^2}) - \frac{1}{2} \quad \square$$

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

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

$$= \int dz_1 \int dz_2 \, q(z_1) q(z_2) \left(\underset{①}{\log(q(z_1))} + \underset{②}{\log(q(z_2))} - \underset{③}{\log(p(z_1, z_2))}\right)$$

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

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

$$③: \iint dz_1 dz_2 \; q(z_1) q(z_2) \left(\frac{1}{2}\underbrace{(\overset{Ⓐ_1}{a z_1^2} + \overset{Ⓐ_2}{a z_2^2}}_{Ⓐ} + \underbrace{2 b z_1 z_2}_{Ⓑ})\right)$$

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

~~$\int dz \int dz_2$~~

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

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

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

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

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

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

d.) we notice 2 things:

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

$\rightsquigarrow q_i(z_i) \equiv p_i(z_i)$ 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 z₁ or project onto z₂ 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 $p(z_1, z_2)$:

$a z_1^2 + a z_2^2 + 2b z_1 z_2 = a z_1^2 + \frac{b^2}{a} z_1^2 - \frac{b^2}{a} z_1^2 + a z_2^2 + 2b z_1 z_2$

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

$\rightarrow E_p z_1^2$

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

no contribution to $E_p z_1^2$ → const.

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