

Mathematical notes on Bishop's book on Pattern Recognition and Machine Learning

Lukas Prokop

October 2016 to February 2016

Contents

1	Preliminaries	2
2	Exercise 1.5	2
3	$\mathbb{E}[ab] = \mathbb{E}[a] \cdot \mathbb{E}[b]$ if a and b are independent	2
4	Exercise 1.6	3
5	Gaussian interpretation of curve fitting	3

1 Preliminaries

2 Exercise 1.5

Given $\text{var}[f] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2]$

Find $\text{var}[f] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$

For the expected value \mathbb{E} , random variables a and b and constant c it holds that

$$\mathbb{E}[a + b] = \mathbb{E}[a] + \mathbb{E}[b] \quad (1)$$

$$\mathbb{E}[c \cdot a] = c \cdot \mathbb{E}[a] \quad (2)$$

$$\mathbb{E}[c] = c \quad (3)$$

Followingly, we can prove the desired equation:

$$\begin{aligned} \text{var}[f] &= \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2] \\ &= \mathbb{E}[f(x)^2 - 2f(x)\mathbb{E}[f(x)] + \mathbb{E}[f(x)]^2] \\ &\stackrel{(1)}{=} \mathbb{E}[f(x)^2] - \mathbb{E}[2 \cdot f(x) \cdot \mathbb{E}[f(x)]] + \mathbb{E}[f(x)]^2 \\ &\stackrel{(2)}{=} \mathbb{E}[f(x)^2] - \mathbb{E}[2] \cdot \mathbb{E}[f(x)] \cdot \mathbb{E}[\mathbb{E}[f(x)]] + \mathbb{E}[f(x)]^2 \\ &\stackrel{(3)}{=} \mathbb{E}[f(x)^2] - 2 \cdot \mathbb{E}[f(x)] \cdot \mathbb{E}[f(x)] + \mathbb{E}[f(x)]^2 \\ &= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2 \end{aligned}$$

3 $\mathbb{E}[ab] = \mathbb{E}[a] \cdot \mathbb{E}[b]$ if a and b are independent

Given two independent random variables a and b

Find $\mathbb{E}[a \cdot b] = \mathbb{E}[a] \cdot \mathbb{E}[b]$

Let $\mathbb{E}[x]$ be defined by

$$\mathbb{E}[x] = \int p(x) \cdot x \, dx$$

Furthermore for two independent random variables a and b , it holds that

$$p(ab) = p(a) \cdot p(b)$$

$$\begin{aligned}
\mathbb{E}[x] &= \int p(x) \cdot x \, dx \\
\mathbb{E}[a \cdot b] &= \int (p(ab) \cdot ab) \, d(ab) \\
&= \int \int (p(a) \cdot p(b) \cdot a \cdot b) \, da \, db \\
&= \int \int (p(a) \cdot a) \cdot (p(b) \cdot b) \, da \, db \\
&= \left(\int p(a) \cdot a \, da \right) \left(\int p(b) \cdot b \, db \right) \\
&= \mathbb{E}[a] \cdot \mathbb{E}[b]
\end{aligned}$$

4 Exercise 1.6

“For two random variables x and y , the *covariance* is defined by

$$\begin{aligned}
\text{cov}[x, y] &= \mathbb{E}_{x,y}[\{x - \mathbb{E}_x[x]\}\{y - \mathbb{E}_y[y]\}] \\
&= \mathbb{E}_{x,y}[xy] - \mathbb{E}_x[x]\mathbb{E}_y[y]
\end{aligned}$$

which expresses the extent to which x and y vary together. If x and y are independent, then their covariance vanishes.”

—page 20

Given two independent random variables x and y with covariance defined by $\text{cov}[x, y] = \mathbb{E}_{x,y}[\{x - \mathbb{E}_x[x]\}\{y - \mathbb{E}_y[y]\}]$

Find $\text{cov}[x, y] = 0$

If two random variables a and b are independent, then it holds that

$$\mathbb{E}[a \cdot b] = \mathbb{E}[a] \cdot \mathbb{E}[b] \quad (4)$$

$$\begin{aligned}
\text{cov}[x, y] &= \mathbb{E}_{x,y}[\{x - \mathbb{E}_x[x]\}\{y - \mathbb{E}_y[y]\}] \\
&= \mathbb{E}_{x,y}[xy - y\mathbb{E}_x[x] - x\mathbb{E}_y[y] + \mathbb{E}_x[x] \cdot \mathbb{E}_y[y]] \\
&\stackrel{(1)}{=} \mathbb{E}_{x,y}[xy] + \mathbb{E}_{x,y}[-y \cdot \mathbb{E}_x[x]] + \mathbb{E}_{x,y}[-x \cdot \mathbb{E}_y[y]] + \mathbb{E}_{x,y}[\mathbb{E}_x[x] \cdot \mathbb{E}_y[y]] \\
&\stackrel{(2)}{=} \mathbb{E}_{x,y}[xy] - \mathbb{E}_x[x] \cdot \mathbb{E}_y[y] - \mathbb{E}_y[y] \cdot \mathbb{E}_x[x] + \mathbb{E}_{x,y}[\mathbb{E}_x[x] \cdot \mathbb{E}_y[y]] \\
&\stackrel{(4)}{=} \mathbb{E}_{x,y}[x] \cdot \mathbb{E}_{x,y}[y] - 2 \cdot \mathbb{E}_x[x] \cdot \mathbb{E}_y[y] + \mathbb{E}_{x,y}[\mathbb{E}_x[x] \cdot \mathbb{E}_y[y]] \\
&\stackrel{(3)}{=} \mathbb{E}_{x,y}[x] \cdot \mathbb{E}_{x,y}[y] - 2 \cdot \mathbb{E}_x[x] \cdot \mathbb{E}_y[y] + \mathbb{E}_x[x] \cdot \mathbb{E}_y[y] \\
&= 0
\end{aligned}$$

5 Gaussian interpretation of curve fitting

Given $p(t \mid x, w, \beta) = \mathcal{N}(t \mid y(x, w), \beta^{-1})$ and

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

Find

$$\ln p(t \mid x, w, \beta) = -\frac{\beta}{2} \sum_{n=1}^N (y(x_n, w) - t_n)^2 + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi)$$

Remember the basic laws for logarithms:

$$\ln(a \cdot b) = \ln a + \ln b \quad (5)$$

$$\ln\left(\frac{a}{b}\right) = \ln a - \ln b \quad (6)$$

$$\ln(\exp(a)) = a \quad (7)$$

$$\ln(a^b) = b \cdot \ln(a) \quad (8)$$

$$\ln p(t \mid x, w, \beta) = (2\pi\beta^{-1})^{-\frac{1}{2}} \exp(-2^{-1}\beta(t - y(x, w))^2) \quad (9)$$

$$\stackrel{(5)}{=} \ln(2\pi\beta^{-1})^{-\frac{1}{2}} + \ln \exp(-2^{-1}\beta(t - y(x, w))^2) \quad (10)$$

$$\stackrel{(7)}{=} \ln\left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} - \frac{\beta}{2}(t - y(x, w))^2 \quad (11)$$

$$\stackrel{(7)}{=} \frac{1}{2} \cdot \ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2}(t - y(x, w))^2 \quad (12)$$

$$\stackrel{(7)}{=} \frac{1}{2} \cdot \ln \beta - \frac{1}{2} \cdot \ln(2\pi) - \frac{\beta}{2}(t - y(x, w))^2 \quad (13)$$

$$\sum_{n=1}^N \ln p(t_n \mid x_n, w, \beta) = \sum_{n=1}^N \left(\frac{1}{2} \ln \beta - \frac{1}{2} \ln 2\pi - \frac{\beta}{2}(t_n - y(x_n, w))^2 \right) \quad (14)$$

$$= \frac{N}{2} \ln \beta - \frac{N}{2} \ln 2\pi - \frac{\beta}{2} \sum_{n=1}^N (y(x_n, w) - t_n)^2 \quad (15)$$