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

Lukas Prokop

October 2016 to February 2016

Contents

1 Preliminaries

2 Exercise 1.5

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

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

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

$$\mathbb{E}[a+b] = \mathbb{E}[a] + \mathbb{E}[b] \tag{1}$$
$$\mathbb{E}[c \cdot a] = c \cdot \mathbb{E}[a] \tag{2}$$
$$\mathbb{E}[c] = c \tag{3}$$

Followingly, we can prove the desired equation:

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

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 $p(x)$ be the probability density function on $\mathbb{R}$ and x be the $\mathbb{R}$-valued random variable (for p). The expectation $\mathbb{E}[x]$ is defined by

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

We will prove the formula under this setting. The case that the probability distribution is defined on $\mathbb{R}^d$ can be discussed analogously. The case that the probability distribution is given on a finite set Ω can be discussed by replacing the integral by the sum over the finite set Ω. 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
>
> $$\begin{aligned}\operatorname{cov}[x,y] &= \mathbb{E}_{x,y}[\{x-\mathbb{E}_x[x]\}\{y-\mathbb{E}_y[y]\}] \\ &= \mathbb{E}_{x,y}[xy]-\mathbb{E}_x[x]\mathbb{E}_y[y]\end{aligned}$$
>
> 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 $\operatorname{cov}[x,y] = \mathbb{E}_{x,y}[\{x-\mathbb{E}_x[x]\}\{y-\mathbb{E}_y[y]\}]$

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

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

$$\mathbb{E}[a\cdot b] = \mathbb{E}[a]\cdot\mathbb{E}[b] \tag{4}$$

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

5 Gaussian interpretation of curve fitting

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

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

Find

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

Remember the basic laws for logarithms:

$$\ln(a \cdot b) = \ln a + \ln b \tag{5}$$

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

$$\ln(\exp(a)) = a \tag{7}$$

$$\ln(a^b) = b \cdot \ln(a) \tag{8}$$

$$\begin{aligned}
\ln p(t \mid x, w, \beta) &= (2\pi\beta^{-1})^{-\frac{1}{2}} \exp\left(-2^{-1}\beta(t - y(x, w))^2\right) && (9)\\
&\overset{(5)}{=} \ln(2\pi\beta^{-1})^{-\frac{1}{2}} + \ln\exp\left(-2^{-1}\beta(t - y(x, w))^2\right) && (10)\\
&\overset{(7)}{=} \ln\left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} - \frac{\beta}{2}(t - y(x, w))^2 && (11)\\
&\overset{(7)}{=} \frac{1}{2} \cdot \ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2}(t - y(x, w))^2 && (12)\\
&\overset{(7)}{=} \frac{1}{2} \cdot \ln\beta - \frac{1}{2} \cdot \ln(2\pi) - \frac{\beta}{2}(t - y(x, w))^2 && (13)\\
\sum_{n=1}^{N} \ln p(t_n \mid x_n, w, \beta) &= \sum_{n=1}^{N} \left(\frac{1}{2}\ln\beta - \frac{1}{2}\ln 2\pi - \frac{\beta}{2}(t_n - y(x_n, w))^2\right) && (14)\\
&= \frac{N}{2}\ln\beta - \frac{N}{2}\ln 2\pi - \frac{\beta}{2}\sum_{n=1}^{N}(y(w_n, w) - t_n)^2 && (15)
\end{aligned}$$

6 Exercise 1.11

Given $\ln p(x \mid \mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi)$

Find $\mu_{\text{ML}} = \frac{1}{N} \cdot \sum_{n=1}^{N} x_n$ for maximized μ and
$\sigma_{\text{ML}} = \frac{1}{N} \cdot \sum_{n=1}^{N} (x_n - \mu_{\text{ML}})^2$ for maximized σ^2

So we want to determine the 2 parameters of a Gaussian distribution, namely μ and σ^2, in the maximum likelihood case. We begin with μ:

1. Derive $\ln p(x \mid \mu, \sigma^2)$ for μ

$$\begin{aligned}
\frac{\partial}{\partial \mu} \ln p(x \mid \mu, \sigma^2) &= \frac{\partial}{\partial \mu} \left(-\frac{1}{2\sigma^2} \cdot \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi) \right) \\
&= \frac{\partial}{\partial \mu} \left(-\frac{1}{2\sigma^2} \cdot \sum_{n=1}^{N} (x_n^2 - 2x_n\mu + \mu^2) - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi) \right) \\
&= -\frac{1}{2\sigma^2} \cdot \sum_{n=1}^{N} (-2x_n + 2\mu) \\
&= -\frac{1}{\sigma^2} \cdot \sum_{n=1}^{N} (\mu - x_n)
\end{aligned}$$

2. Set result zero

$$0 = -\frac{1}{\sigma^2} \cdot \sum_{n=1}^{N} (\mu - x_n) = \sum_{n=1}^{N} (\mu - x_n) = N \cdot \mu - \sum_{n=1}^{N} x_n$$

$$\implies \mu_{\text{ML}} = \frac{1}{N} \cdot \sum_{n=1}^{N} x_n \qquad \text{commonly called “sample mean”}$$

We continue with σ^2 and use the same approach:

1. Derive $\ln p(x \mid \mu, \sigma^2)$ for σ^2

$$\begin{aligned}
\frac{\partial}{\partial \sigma^2} \ln p(x \mid \mu, \sigma^2) &= \frac{\partial}{\partial \sigma^2} \left(-\frac{1}{2\sigma^2} \cdot \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi) \right) \\
&= \frac{1}{2\sigma^4} \cdot \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{N}{2} \cdot \frac{1}{\sigma^2} \\
&= \frac{1}{2\sigma^2} \left(\frac{1}{\sigma^2} \cdot \sum_{n=1}^{N} (x_n - \mu)^2 - N \right)
\end{aligned}$$

2. Set result zero

$$\begin{aligned}
0 &= \frac{1}{2\sigma^2} \left(\frac{1}{\sigma^2} \cdot \sum_{n=1}^{N} (x_n - \mu)^2 - N \right) \\
N \cdot \sigma^2 &= \sum_{n=1}^{N} (x_n - \mu)^2 \\
\sigma^2_{\text{ML}} &= \frac{1}{N} \cdot \sum_{n=1}^{N} (x_n - \mu)^2 \qquad \text{commonly called “sample variance”}
\end{aligned}$$

7 Precision parameter β in the maximum likelihood case

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

Find $\frac{1}{\beta_{\mathrm{ML}}} = \frac{1}{N} \cdot \sum_{n=1}^{N}(y(x_n, w_{\mathrm{ML}}) - t_n)^2$ by maximizing β

1. Derive $\ln p(t \mid x, w, \beta)$ with β

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

2. Set result zero

$$0 = -\frac{1}{2} \cdot \sum_{n=1}^{N}\left(y(x_n, w) - t_n\right)^2 + \frac{N}{2\beta}$$
$$\frac{N}{\beta} = \sum_{n=1}^{N}\left(y(x_n, w) - t_n\right)^2$$
$$\frac{1}{\beta_{\mathrm{ML}}} = \frac{1}{N} \cdot \sum_{n=1}^{N}(y(x_n, w) - t_n)^2$$