Deep Learning

Theoretical Exercises – Week 2 – Chapter 3

Exercises on the book ”Deep Learning” written by Ian Goodfellow,
Yoshua Bengio, and Aaron Courville.
Exercises and solutions by T. Méndez and G. Schuster

FS 2024

1 Exercises on Probability and Information Theory

1. Given a Bernoulli distributed random variable x with $P(\mathrm{x} = 1) = \frac{3}{4}$ and $P(\mathrm{x} = 0) = \frac{1}{4}$.

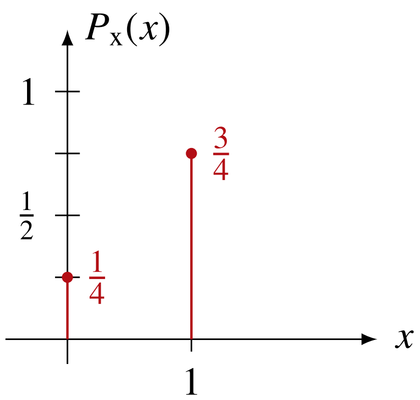
   (a) Draw the probability mass function $P_{\mathrm{x}}(x)$.

   (b) Find the expected value of x, i.e. $\mathbb{E}_{\mathrm{x}}[x]$.

   (c) Find the variance of x, i.e. $\mathrm{Var}_{\mathrm{x}}(x)$.

   Solution:

   (a) Probability mass function:



(b) Expected value:

$$\begin{aligned}
\mathbb{E}_{\mathrm{x}}[x] &= \sum_x x\, P(\mathrm{x} = x) \\
&= 0 \cdot P(\mathrm{x} = 0) + 1 \cdot \mathrm{P}(\mathrm{x} = 1) \\
&= 0 \cdot \frac{1}{4} + 1 \cdot \frac{3}{4} \\
&= \frac{3}{4}
\end{aligned}$$

(c) To calculate the variance, $\mathbb{E}_{\mathrm{x}}\left[x^2\right]$ has to be calculated first:

$$\begin{aligned}
\mathbb{E}_{\mathrm{x}}\left[x^2\right] &= \sum_x x^2\, P(\mathrm{x} = x) \\
&= 0^2 \cdot P(\mathrm{x} = 0) + 1^2 \cdot \mathrm{P}(\mathrm{x} = 1) \\
&= 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{3}{4} \\
&= \frac{3}{4},
\end{aligned}$$

which results in a variance of

$$\begin{aligned}
\mathrm{Var}_{\mathrm{x}}(x) &= \mathbb{E}_{\mathrm{x}}\left[x^2\right] - (\mathbb{E}_{\mathrm{x}}[x])^2 \\
&= \frac{3}{4} - \left(\frac{3}{4}\right)^2 \\
&= \frac{3}{16}.
\end{aligned}$$

Alternatively, the variance can also be calculated as

$$\begin{aligned}
\mathrm{Var}_{\mathrm{x}}(x) &= \mathbb{E}\left[(x - \mathbb{E}_{\mathrm{x}}[x])^2\right] \\
&= \mathbb{E}\left[\left(x - \frac{3}{4}\right)^2\right] \\
&= \left(0 - \frac{3}{4}\right)^2 \cdot P(\mathrm{x} = 0) + \left(1 - \frac{3}{4}\right)^2 \cdot \mathrm{P}(\mathrm{x} = 1) \\
&= \left(0 - \frac{3}{4}\right)^2 \cdot \frac{1}{4} + \left(1 - \frac{3}{4}\right)^2 \cdot \frac{3}{4} \\
&= \frac{3}{16}.
\end{aligned}$$

2. Given a uniformly distributed random variable x between −0.5 and 1.

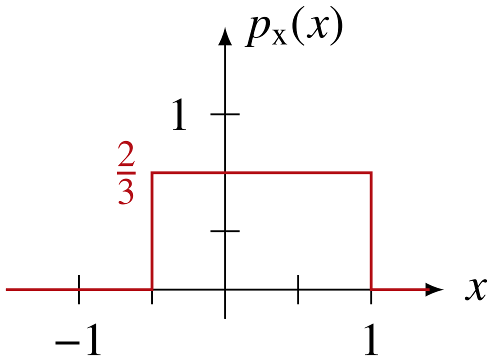
   (a) Draw the probability density function $f_{\mathrm{x}}(x)$.

   (b) Find the expected value of x, i.e. $\mathbb{E}_{\mathrm{x}}[x]$.

   (c) Find the variance of x, i.e. $\mathrm{Var}_{\mathrm{x}}(x)$.

Solution:

(a) Probability density function:



(b) Expected value:

$$
\begin{aligned}
\mathbb{E}_x[x] &= \int x\, p_x(x)\, dx \\
&= \int_{-0.5}^{1} x\, \frac{2}{3}\, dx \\
&= \left[\frac{x^2}{3}\right]_{-0.5}^{1} \\
&= \frac{1}{3} - \frac{1}{12} \\
&= \frac{1}{4}
\end{aligned}
$$

(c) To calculate the variance, $\mathbb{E}_x\left[x^2\right]$ has to be calculated first:

$$
\begin{aligned}
\mathbb{E}_x\left[x^2\right] &= \int x^2\, p_x(x)\, dx \\
&= \int_{-0.5}^{1} x^2\, \frac{2}{3}\, dx \\
&= \left[\frac{2\,x^3}{9}\right]_{-0.5}^{1} \\
&= \frac{2}{9} + \frac{1}{36} \\
&= \frac{1}{4},
\end{aligned}
$$

which results in a variance of

$$
\begin{aligned}
\mathrm{Var}_x(x) &= \mathbb{E}_x\left[x^2\right] - (\mathbb{E}_x[x])^2 \\
&= \frac{1}{4} - \left(\frac{1}{4}\right)^2 \\
&= \frac{3}{16}.
\end{aligned}
$$

3. There are 5000 yellow and 100 red cabs in a city. In a hit-and-run accident, a witness saw a red cab. What is the probability that it was actually a red cab if witnesses in 95% of the cases state the car's colour correctly?

Solution:
The probability $P(c = r \mid w = r)$ is sought. This is the probability that the cab (c) was red, given the witness (w) said red (r). To calculate this probability the Bayes' rule is used

$$P(c = r \mid w = r) = \frac{P(c = r)\, P(w = r \mid c = r)}{P(w = r)} = 27.54\%,$$

where

$$P(c = r) = \frac{100}{5000 + 100} = 0.0196,$$
$$P(w = r \mid c = r) = 0.95,$$

and

$$\begin{aligned} P(w = r) &= P(w = r \mid c = r)P(c = r) + P(w = r \mid c = y)P(c = y) \\ &= 0.95 \cdot \frac{100}{5000 + 100} + 0.05 \cdot \frac{5000}{5000 + 100} \\ &= 0.0676. \end{aligned} \tag{1.1}$$

4. Given are the following sets of samples of two different classes c_1 and c_2:

$$S_1 = \{5.3,\ 5.7,\ 6.1,\ 6.3,\ 6.6\},$$

$$S_2 = \{6.2,\ 6.5,\ 6.9,\ 7.7,\ 8.0,\ 8.3,\ 8.9\}.$$

   (a) Estimate the mean (expected value) and the variance of both classes and sketch their probability density function assuming that the samples of both classes are normally distributed.
   (b) Estimate the prior probabilities $P(y = c_i)$ of both classes.
   (c) Assign the new samples $x_1 = 5$ and $x_2 = 7$ to the class with the highest posterior probability.
   (d) Calculate the entropy $H(x \mid y)$ of x with $y = c_1$ and $y = c_2$ respectively.

Solution:

   (a) The mean and the variance are estimated as

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i$$

and

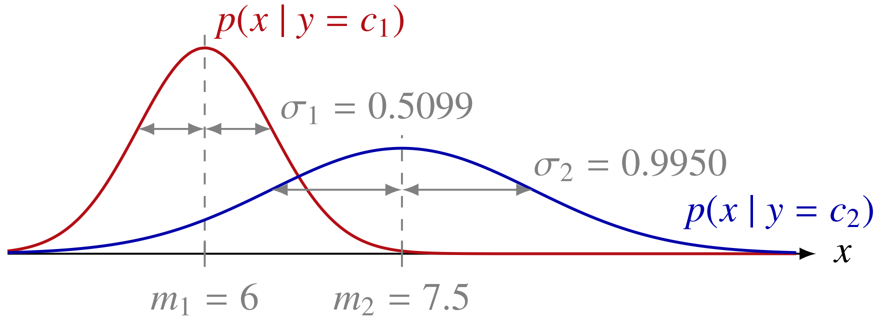
$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - m)^2,$$

thus

$$m_1 = \frac{5.3 + 5.7 + 6.1 + 6.3 + 6.6}{5} = 6,$$
$$m_2 = \frac{6.2 + 6.5 + 6.9 + 7.7 + 8.0 + 8.3 + 8.9}{7} = 7.5,$$
$$\sigma_1^2 = \frac{(5.3 - 6)^2 + (5.7 - 6)^2 + (6.1 - 6)^2 + (6.3 - 6)^2 + (6.6 - 6)^2}{5 - 1} = 0.26,$$

and

$$\sigma_2^2 = \frac{(6.2-7.5)^2+(6.5-7.5)^2+(6.9-7.5)^2+(7.7-7.5)^2+(8.0-7.5)^2+(8.3-7.5)^2+(8.9-7.5)^2}{7-1}$$
$$= 0.99.$$



(b) Prior probabilities:

$$P(y = c_1) = \frac{5}{5+7} = \frac{5}{12}$$
$$P(y = c_2) = \frac{7}{5+7} = \frac{7}{12}$$

(c) If

$$P(y = c_1 \mid x) > P(y = c_2 \mid x),$$

the sample x is assigned to class c_1 otherwise x is assigned to class c_2. Thus, to assign the new samples, $P(y = c_1 \mid x)$ and $P(y = c_2 \mid x)$ have to be determined. Since

$$P(y = c_1 \mid x) = \frac{P(y = c_1)\, p(x \mid y = c_1)}{p(x)}$$

and

$$P(y = c_2 \mid x) = \frac{P(y = c_2)\, p(x \mid y = c_2)}{p(x)},$$

the value of the term $p(x)$ does not change the assignment and can be omitted. The probabilities $P(y = c_1)$ and $P(y = c_2)$ already have been determined in (b) and the values of the probability density functions are calculated as

$$p(x \mid y = c_i) = \sqrt{\frac{1}{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(x - m_i)^2\right).$$

For the first sample $x_1 = 5$ this results in

$$p(x = 5 \mid y = c_1) = 0.1144$$

and

$$p(x = 5 \mid y = c_2) = 0.0171$$

and for the second sample $x_2 = 7$ in

$$p(x = 7 \mid y = c_1) = 0.1144$$

and

$$p(x = 7 \mid y = c_2) = 0.3534.$$

Thus, x_1 is assigned to class c_1, because

$$\begin{aligned}
P(y = c_1)\, p(x = 5 \mid y = c_1) &> P(y = c_2)\, p(x = 5 \mid y = c_2) \\
\frac{5}{12} \cdot 0.1144 &> \frac{7}{12} \cdot 0.0171 \\
0.0476 &> 0.0100,
\end{aligned}$$

and x_2 is assigned to class c_2, because

$$\begin{aligned}
P(y = c_1)\, p(x = 7 \mid y = c_1) &< P(y = c_2)\, p(x = 7 \mid y = c_2) \\
\frac{5}{12} \cdot 0.1144 &< \frac{7}{12} \cdot 0.3534 \\
0.0476 &< 0.2061.
\end{aligned}$$

(d) The entropy of a Gaussian distributed random variable x with mean m and variance σ^2 is

$$\begin{aligned}
H(x) &= -\mathbb{E}[\log p(x)] \\
&= -\int_{-\infty}^{\infty} \log\left(p(x)\right) \cdot p(x)\, dx \\
&= -\int_{-\infty}^{\infty} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\, \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)\right) \cdot p(x)\, dx \\
&= -\int_{-\infty}^{\infty} \left[-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{(x-m)^2}{2\sigma^2}\right] \cdot p(x)\, dx \\
&= \int_{-\infty}^{\infty} \frac{1}{2}\log\left(2\pi\sigma^2\right) \cdot p(x)\, dx + \int_{-\infty}^{\infty} \frac{(x-m)^2}{2\sigma^2} \cdot p(x)\, dx \\
&= \frac{1}{2}\log\left(2\pi\sigma^2\right) \underbrace{\int_{-\infty}^{\infty} p(x)\, dx}_{=1} + \frac{1}{2\sigma^2} \underbrace{\int_{-\infty}^{\infty} (x-m)^2 \cdot p(x)\, dx}_{\sigma^2} \\
&= \frac{1}{2}\log\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\sigma^2 \\
&= \frac{1}{2}\left(\log\left(2\pi\sigma^2\right) + 1\right),
\end{aligned}$$

which indicates that the greater the variance (uncertainty), the greater the entropy. It follows from this that the entropy $H(x \mid y = c_1)$ is

$$\begin{aligned}
H(x \mid y = c_1) &= \frac{1}{2}\left(\log\left(2\pi\sigma_1^2\right) + 1\right) \\
&= \frac{1}{2}\left(\log\left(2\pi\, 0.26\right) + 1\right) \\
&= 0.7454\,\text{nats}
\end{aligned}$$

and the entropy of $H(x \mid y = c_2)$ is

$$\begin{aligned}
H(x \mid y = c_2) &= \frac{1}{2}\left(\log\left(2\pi\sigma_2^2\right) + 1\right) \\
&= \frac{1}{2}\left(\log\left(2\pi\, 0.99\right) + 1\right) \\
&= 1.4139\,\text{nats}.
\end{aligned}$$

Thus, the expected amount of information in a sample of class c_1 and c_2, respectively, is 0.7454 nats[1] and 1.4139 nats, respectively.

[1]The natural unit of information (nat), is a unit of information or entropy, based on natural logarithms and powers of e, rather than the powers of 2 and base 2 logarithms, which define the bit.