ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Unsupervised and reinforcement learning in neural networks
Week 3

Lecturer: Marc-Oliver Gewaltig

Assistants: Dane Corneil
Berat Denizdurduran

Corrections: Whitening, Decorrelation & Independence

Exercise 1

1.1. The definition of the expectation value[1] of a random variable x with distribution probability $p(x)$ is

$$\langle x \rangle := \int dx\, p(x)\, x,$$

so that we can rewrite the correlation term:

$$\begin{aligned}\langle (x_1 - \langle x_1 \rangle)(x_2 - \langle x_2 \rangle) \rangle &= \int dx_1 \int dx_2\, p(x_1, x_2)\, (x_1 - \langle x_1 \rangle)(x_2 - \langle x_2 \rangle) \\ &= \int dx_1 \int dx_2\, p(x_1)\, (x_1 - \langle x_1 \rangle)\, p(x_2)\, (x_2 - \langle x_2 \rangle) \\ &= \left(\int dx_1\, p(x_1)\, (x_1 - \langle x_1 \rangle) \right) \left(\int dx_2\, p(x_2)\, (x_2 - \langle x_2 \rangle) \right) \\ &= (\langle x_1 \rangle - \langle x_1 \rangle)(\langle x_2 \rangle - \langle x_2 \rangle) = 0.\end{aligned}$$

In the last step, we used the fact that $\langle x \rangle$ is independent of x, so that $\int dx\, p(x)\, \langle x \rangle = \langle x \rangle \int dx\, p(x) = \langle x \rangle$ (since $\int dx\, p(x) = 1$).

1.2. If all the x_i and x_j variables are decorrelated, the C_{ij} matrix must vanish outside the diagonal. The diagonal elements correspond to the variances σ_i^2, i.e.:

$$C_{ij} = \begin{cases} \sigma_i^2 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

[1] Here we use the notation $\langle \cdot \rangle$ instead of the more formally correct $E\{\cdot\}$ for the expectation value.

The inverse of such a matrix is a diagonal matrix with elements $1/\sigma_i^2$. Using these two facts, we can rewrite $p(\vec{x})$

$$\begin{aligned}
p(\vec{x}) &= \frac{1}{\sqrt{(2\pi)^N det(C)}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T C^{-1}(\vec{x}-\vec{\mu})\right) \\
&= \frac{1}{\prod_i \sqrt{2\pi\sigma_i^2}} \exp\left(-\sum_i \frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) \\
&= \frac{1}{\prod_i \sqrt{2\pi\sigma_i^2}} \prod_i \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) \\
&= \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) = \prod_i p_i(x_i)
\end{aligned}$$

where $p(x_i)$ is a 1-dim. gaussian distribution with mean μ_i and variance σ_i^2.

Exercise 2

By definition, a whitened data set has: 1) uncorrelated components and 2) variances of all its components equal to 1. This means that its covariance matrix is the identity matrix E:

$$C := \frac{1}{p}\sum_{\mu=1}^{p}(\vec{x}^\mu - \langle\vec{x}\rangle)(\vec{x}^\mu - \langle\vec{x}\rangle)^T = E. \tag{1}$$

We replace $\vec{x}$ by $\vec{y} = R\vec{x}$ and check its covariance matrix C^*:

$$\begin{aligned}
C^* &= \frac{1}{p}\sum_\mu(\vec{y}^\mu - \langle\vec{y}\rangle)(\vec{y}^\mu - \langle\vec{y}\rangle)^T \\
&= \frac{1}{p}\sum_\mu(R\vec{x}^\mu - R\langle\vec{x}\rangle)(R\vec{x}^\mu - R\langle\vec{x}\rangle)^T \\
&= \frac{1}{p}\sum_\mu R(\vec{x}^\mu - \langle\vec{x}\rangle)(\vec{x}^\mu - \langle\vec{x}\rangle)^T R^T \\
&= R \underbrace{\left(\frac{1}{p}\sum_\mu(\vec{x}^\mu - \langle\vec{y}\rangle)(\vec{x}^\mu - \langle\vec{x}\rangle)^T\right)}_{=E} R^T \\
&= RER^T = RR^T = E.
\end{aligned}$$

Exercise 3

This exercise is furiously similar to exercise 1 above. Again using the definition of the expectation value:

$$\langle f(x)\rangle := \int dx\, p(x)\, f(x),$$

we can rewrite the expectation value of the product:

$$\begin{aligned}\langle h_1(x_1)h_2(x_2)\rangle &= \int dx_1 \int dx_2\, p(x_1,x_2)\, h_1(x_1)\, h_2(x_2)\\ &= \int dx_1 \int dx_2\, p(x_1)\, h_1(x_1)\, p(x_2)\, h_2(x_2)\\ &= \int dx_1\, p(x_1)\, h_1(x_1) \int dx_2\, p(x_2)\, h_2(x_2)\\ &= \langle h_1(x_1)\rangle\langle h_2(x_2)\rangle.\end{aligned}$$

Exercise 4

4.1. Let's compute the first, second and fourth moments of the Gaussian distribution. The mean is immediately solved by noticing the symmetry in the distribution:

$$\begin{aligned}\langle x\rangle &= \sqrt{\frac{a}{\pi}}\int_{-\infty}^{\infty} dx\, x\, exp(-ax^2) = \sqrt{\frac{a}{\pi}}\left(\int_{-\infty}^{0} dx\, x\, exp(-ax^2) + \int_{0}^{\infty} dx\, x\, exp(-ax^2)\right)\\ &= \sqrt{\frac{a}{\pi}}\left(\int_{0}^{\infty} dx\, x\, exp(-ax^2) - \int_{0}^{\infty} dx\, x\, exp(-ax^2)\right) = 0\end{aligned}$$

Note that the same applies for all odd moments of any symmetric distribution.
For the second moment, we use the hint:

$$\langle x^2\rangle = \sqrt{\frac{a}{\pi}}\int_{-\infty}^{\infty} dx\, x^2\, exp(-ax^2) = -\sqrt{\frac{a}{\pi}}\frac{d}{da}\underbrace{\int_{-\infty}^{\infty} dx\, exp(-ax^2)}_{=\sqrt{\frac{\pi}{a}}} = -\sqrt{\frac{a}{\pi}}\left(-\frac{1}{2}\sqrt{\frac{\pi}{a^3}}\right) = \frac{1}{2a}$$

For the fourth moment, we just use the same trick twice:

$$\begin{aligned}\langle x^4\rangle &= \sqrt{\frac{a}{\pi}}\int_{-\infty}^{\infty} dx\, x^4\, exp(-ax^2) = -\sqrt{\frac{a}{\pi}}\frac{d}{da}\int_{-\infty}^{\infty} dx\, x^2\, exp(-ax^2)\\ &= \sqrt{\frac{a}{\pi}}\frac{d^2}{da^2}\underbrace{\int_{\infty}^{\infty} dx\, exp(-ax^2)}_{=\sqrt{\frac{\pi}{a}}} = \sqrt{\frac{a}{\pi}}\left(\frac{3}{4}\sqrt{\frac{\pi}{a^5}}\right) = \frac{3}{4a^2}\end{aligned}$$

Using these results one can compute the variance $var(x) = \frac{1}{2a}$, and the kurtosis $\kappa(x) = \frac{3}{4a^2} - 3\left(\frac{1}{2a}\right)^2 = 0$.

4.2. (uniform distribution) Again, we compute the the moments we need to compute the variance and the kurtosis. The first moment is again 0 by symmetry. The second moment is

$$\langle x^2\rangle = \int_{-\sqrt{3}}^{\sqrt{3}} dx\, \frac{1}{2\sqrt{3}}\, x^2 = \frac{1}{2\sqrt{3}}\left[\frac{x^3}{3}\right]_{-\sqrt{3}}^{\sqrt{3}} = 1\,,$$

and the fourth is

$$\langle x^4 \rangle = \int\limits_{-\sqrt{3}}^{\sqrt{3}} dx\, \frac{1}{2\sqrt{3}}\, x^4 = \frac{1}{2\sqrt{3}} \left[\frac{x^5}{5} \right]_{-\sqrt{3}}^{\sqrt{3}} = \frac{9}{5}.$$

Thus the variance is $var(x) = 1$, and the kurtosis is $\kappa(x) = \frac{9}{5} - 3 = -\frac{6}{5}$.

4.3. (Laplace distribution) Here, because the distribution is symmetric, the first moment vanishes again. To compute the second and fourth moments, it is easiest to use the same trick as for the Gaussian, i.e. to notice that

$$-\frac{d}{da} \int dx\, x^{n-1} \exp(-ax) = \int dx\, x^n \exp(-ax)$$

and that

$$\int\limits_0^\infty dx\, x^0 \exp(-ax) = \frac{1}{a}.$$

It is also nice to get rid of the absolute value. For any symmetric functions ($f(x) = f(-x)$) one can write

$$\int\limits_{-\infty}^{\infty} f(x)dx = \int\limits_{-\infty}^{0} f(x)dx + \int\limits_{0}^{\infty} f(x)dx = \int\limits_{0}^{\infty} f(-x)dx + \int\limits_{0}^{\infty} f(x)dx = 2\int\limits_{0}^{\infty} f(x)dx\,.$$

The terms we have to integrate for even moments of this distribution are symmetric. Computing the second moment then reduces to

$$\begin{aligned}\langle x^2 \rangle &= \frac{2}{\sqrt{2}} \int\limits_0^\infty dx\, x^2 \exp(-\sqrt{2}x) = \sqrt{2} \left[\frac{d^2}{da^2} \int\limits_0^\infty dx\, \exp(-ax) \right]_{a=\sqrt{2}} \\ &= \sqrt{2} \left[\frac{2}{a^3} \right]_{a=\sqrt{2}} = 1\,,\end{aligned}$$

and the fourth

$$\begin{aligned}\langle x^4 \rangle &= \frac{2}{\sqrt{2}} \int\limits_0^\infty dx\, x^4 \exp(-\sqrt{2}x) = \sqrt{2} \left[\frac{d^4}{da^4} \int\limits_0^\infty dx\, \exp(-ax) \right]_{a=\sqrt{2}} \\ &= \sqrt{2} \left[\frac{24}{a^5} \right]_{a=\sqrt{2}} = 6\,.\end{aligned}$$

The variance is $var(x) = 1$, and the kurtosis is $\kappa(x) = 6 - 3 = 3$.