Machine Intelligence 1 | **Exercise Sheet 1**
WS 2021/22, Obermayer/Kashef | due: 25.10.2021 at 23:55

Maths primer

Exercises prefixed with **T** are explained in the tutorial. The **H** prefix is for the homework questions.

The maths primer slides (file `maths_primer.pdf` on ISIS) could be helpful for solving the questions in this sheet.

Exercise T1.1: Learning paradigms (tutorial)

(a) Describe the difference between *supervised, unsupervised,* and *reinforcement learning*.

(b) Which of the above learning techniques would be most appropriate in the following cases and what would be the corresponding *observations*, *labels* and/or *rewards*?

- To identify groups of users with the same taste of music.
- To read hand written addresses from letters.
- To teach a robot to walk through a labyrinth.

Exercise T1.2: Additional maths background (optional) (tutorial)

More topics of the maths primer (than what is covered in the homework) can be discussed on-demand.

Exercise H1.1: Distributions and expected values (homework, 2 points)

Let X be a random variable with probability density $p : \mathbb{R} \to \mathbb{R}$ with:

$$p(x) = \begin{cases} c \cdot \sin(x), & x \in [0, \pi] \\ 0, & \text{elsewhere} \end{cases}$$

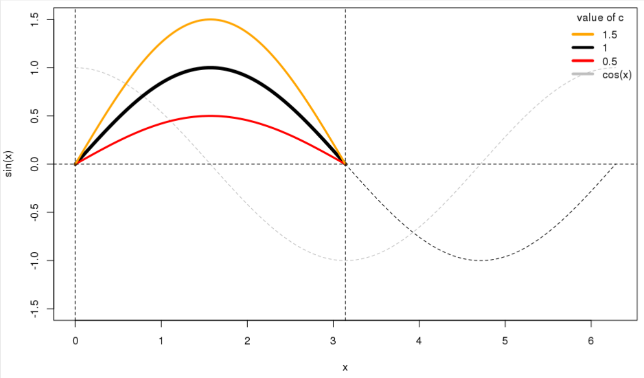
(a) Determine the parameter value $c \in \mathbb{R}$ such that $p(x)$ is indeed a probability density.

(b) Determine the expected value $\langle X \rangle_p$

(c) Determine the variance of X.
Hint: Use the identity $\mathrm{var}(X) = \left\langle X^2 \right\rangle_p - \langle X \rangle_p^2$ for simplicity.

Solution

First let's look at $p(x)$ for different values of c:



(a) For p being a probability density it is required that (i) $p(x) \geq 0 \; \forall x \in \mathbb{R}$ which is fullfilled here. Furthermore, (ii) p must be normalized appropriately:

$$\int_{\mathbb{R}} p(x)dx \quad = \quad 1$$

Therefore, we get for the unknown constant c:

$$c \int_0^\pi \sin(x)dx \quad = \quad c[-\cos(x)]\Big|_0^\pi \quad = \quad 2c \quad \stackrel{!}{=} \quad 1 \quad \rightarrow \quad c = 1/2$$

(b) To calculate the expected value, we use integration by parts, i.e., for any functions f and g:

$$\int_a^b fg'dx \quad = \quad (fg)\Big|_a^b - \int_a^b f'gdx$$

$$\mathbb{E}[X] = \langle X \rangle_p \quad = \quad 1/2 \int_0^\pi x\sin(x)dx \quad = \quad -\frac{1}{2}x\cos(x)\Big|_0^\pi + \frac{1}{2}\underbrace{\int_0^\pi \cos(x)dx}_{=0} \quad = \pi/2$$

(c) To calculate the variance, we proceed in the same way

$$\mathbb{E}[X^2] = \langle X^2 \rangle_p \quad = 1/2 \int_0^\pi x^2 \sin(x)dx \quad = \underbrace{-\frac{1}{2}x^2\cos(x)\Big|_0^\pi}_{=\frac{\pi^2}{2}} \quad + \quad \frac{2}{2}\underbrace{\int_0^\pi x\cos(x)dx}_{=k}$$

with

$$k \quad = \quad x\sin(x)\Big|_0^\pi - \int_0^\pi \sin(x)dx \quad = \quad 0 + \cos(x)\Big|_0^\pi \quad = 0 - 2$$

and therefore

$$\mathbb{E}[X^2] \quad = \quad \frac{\pi^2}{2} - 2$$

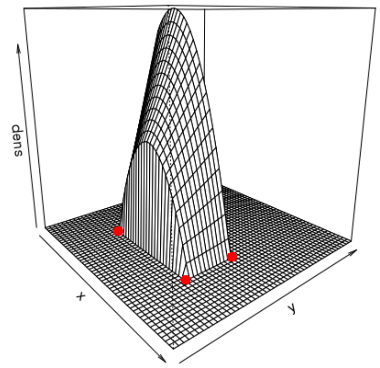
yielding

$$\mathrm{var}(X) \quad = \quad \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad = \quad \frac{\pi^2}{2} - 2 - \frac{\pi^2}{4} \quad = \quad \frac{\pi^2}{4} - 2\,.$$

Exercise H1.2: Marginal densities (homework, 2 points)

Assume the joint probability density of a two-dimensional random vector $\boldsymbol{Z} = (X, Y)^\top$ is

$$p_{\boldsymbol{Z}}(\boldsymbol{z}) = p_{X,Y}(x, y) = \begin{cases} \frac{3}{7}(2-x)(x+y), & x \in [0,2], y \in [0,1] \\ 0, & \text{elsewhere} \end{cases}$$



(a) Write down the marginal densities $p_X(x)$ and $p_Y(y)$ of the variables X and Y.

(b) Determine if the two variables are independent or uncorrelated.

<u>Solution</u>

$$\begin{aligned} p_X(x) &= \int_{-\infty}^{\infty} p_{X,Y}(x,\tilde{y})d\tilde{y} \quad = \frac{3}{7}\int_0^1 (2-x)(x+\tilde{y})d\tilde{y} \\ &= \frac{3}{7}(2x\tilde{y} - x^2\tilde{y} + \tilde{y}^2 - \frac{1}{2}x\tilde{y}^2)\Big|_{y=0}^{1} \\ &= \frac{3}{7}(1 + \frac{3}{2}x - x^2) \qquad \text{for} \quad 0 \leq x \leq 2, \qquad 0 \quad \text{elsewhere} \end{aligned}$$

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{\infty} p_{X,Y}(\tilde{x},y)dx \quad = \frac{3}{7}\int_0^1 (2-\tilde{x})(\tilde{x}+y)d\tilde{x} \\ &= \frac{3}{7}(x^2 - \frac{x^3}{3} + 2xy - \frac{x^2y}{2})\Big|_{x=0}^{2} \\ &= \frac{3}{7}(4 - \frac{8}{3} + 4y - 2y) \\ &= \frac{4}{7} + \frac{6}{7}y \qquad \text{for} \quad 0 \leq y \leq 1, \qquad 0 \quad \text{elsewhere} \end{aligned}$$

independence: not independent as $p_{X,Y}(x,y) \neq p_X(x)p_Y(y)$

comparing coefficients: $p_X(x)p_Y(y)$ has nonzero coefficient for x^2y and $p_{X,Y}(x,y)$ has not; or for a specific value:

$$p_{X,Y}(1,0) = \frac{3}{7} \quad \neq \quad p_X(1)p_Y(0) = \frac{9}{14}\frac{4}{7} = \frac{18}{49}$$

uncorrelatedness: they are not uncorrelated as $\langle XY \rangle_{p_{X,Y}} \neq \langle X \rangle_{p_X} \langle Y \rangle_{p_Y}$ $\left(\frac{10}{21} \neq \frac{6}{7}\frac{4}{7}\right)$

Exercise H1.3: Taylor expansion (homework, 1 point)

For the function $\sqrt{1+x}$, write down the Taylor series around $x_0 = 0$ up to 3rd order.

Solution

Approximating $f(x)$ via Taylor expansion at x_0:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

i.e. for an expansion around $x_0 = 0$

$$f(x) \approx f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + \frac{1}{6}f'''(0)x^3 + \mathcal{O}(x^4)$$

with

$$\begin{aligned} f'(x) &= \frac{1}{2}(x+1)^{-\frac{1}{2}} && \rightarrow f'(0) = 1/2 \\ f''(x) &= -\frac{1}{4}(x+1)^{-\frac{3}{2}} && \rightarrow f''(0) = -1/4 \\ f'''(x) &= \frac{3}{8}(x+1)^{-\frac{5}{2}} && \rightarrow f'''(0) = 3/8 \end{aligned}$$

the Taylor expansion of $\sqrt{1+x} = (1+x)^{\frac{1}{2}}$ around $x_0 = 0$ reads

$$\sqrt{1+x} \approx 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 + ...$$

Exercise H1.4: Determinant of a matrix (homework, 1 point)

Consider the 3×3 matrix

$$\underline{\mathbf{A}} = \begin{pmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{pmatrix}$$

Calculate the determinant and the trace of $\underline{\mathbf{A}}$ (directly, not via eigenvalues).

Solution

Trace: Sum of diagonal elements: $\mathrm{tr}\underline{\mathbf{A}} = 5 + 1 - 11 = -5$

Determinant: For example the rule of Sarrus for 3x3 matrices can be applied: \+ \+ \- / - / - /

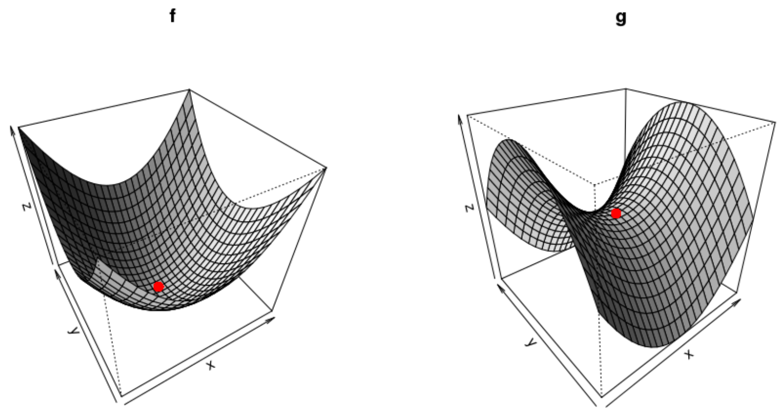
i.e. $\det \underline{\mathbf{A}} = -55 - (64 * 4) - (16 * 16) + (16 * 4) + (160) + (44 * 8) = 9.$

Exercise H1.5: Critical points (homework, 2 points)

Consider the two functions

$$f(x, y) := c + x^2 + y^2$$
$$g(x, y) := c + x^2 - y^2,$$

where $c \in \mathbb{R}$ is a constant.



(a) Show that $\underline{\mathbf{a}} = (0, 0)$ is a critical point of both functions.

(b) Check for f and for g whether $\underline{\mathbf{a}}$ is a minimum, maximum, or no extremum by calculating the Hessian matrix. Make use of the fact that a matrix is positive (negative) definite if and only if all its eigenvalues are positive (negative).

Solution

$$\nabla f(\underline{\mathbf{a}}) = (2x, 2y)\Big|_{(x,y)=\underline{\mathbf{a}}} = (0, 0)$$

and

$$\nabla g(\underline{\mathbf{a}}) = (2x, -2y)\Big|_{(x,y)=\underline{\mathbf{a}}} = (0, 0)$$

$\Longrightarrow$ Necessary condition of extrema (vanishing gradient) is fulfilled at $\underline{\mathbf{a}}$.

Checking for extrema:

Minimum if Hessian matrix $\underline{\mathbf{H}}$ is positive definite (all eigenvalues >0)

Maximum if $\underline{\mathbf{H}}$ is negative definite (all eigenvalues <0)

$\rightarrow$ characteristic polyomial i.e. $\det[\underline{\mathbf{H}} - \lambda \underline{\mathbf{I}}]$:

$$(\underline{\mathbf{H}}_f)(\underline{\mathbf{a}}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \Rightarrow (2 - \lambda)^2 \stackrel{!}{=} 0$$

i.e. all eigenvalues (2&2) are real, positive $\Rightarrow \underline{\mathbf{H}}_f$ is pos. definite. Thus, $\underline{\mathbf{a}}$ is a minimum of f

$$(\underline{\mathbf{H}}_g)(\underline{\mathbf{a}}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \quad \Rightarrow (2 - \lambda)(-2 - \lambda) \stackrel{!}{=} 0$$

positive and negative Eigenvalues (2&-2) $\Rightarrow \underline{\mathbf{H}}_f$ is neither positive nor negative definite. Therefore $\underline{\mathbf{a}}$ is a saddlepoint but no extremum of g.

Exercise H1.6: Bayes rule (homework, 2 points)

Assume it is known that 1% of the population suffer from a certain disease. A company has developed a test for diagnosing the disease, which comes up either positive ("+", disease found) or negative ("-", disease not found). People suffering from the disease (D) are diagnosed positive with probability 0.95, and healthy people ($\bar{D}$) are diagnosed negative with probability 0.999.

Apply Bayes' rule to find

(a) the probabilities that a person for which the test yielded a positive result is indeed suffering from the disease $P(D|+)$, respectively is healthy $P(\bar{D}|+)$.

(b) the probabilities that a person for which the test yielded a negative result is indeed healthy $P(\bar{D}|-)$, respectively is suffering from the disease $P(D|-)$.

Solution

We have

- **Prevalence (prior):** $p(D) = 0.01 \quad \Longrightarrow \; p(\bar{D}) = 0.99$
- **Sensitivity:** $p(+|D) = 0.950 \quad \Longrightarrow \; p(-|D) = 0.05$
- **Specificity:** $p(-|\bar{D}) = 0.999 \quad \Longrightarrow \; p(+|\bar{D}) = 0.001.$

Bayes rule:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\sum_x p(y|x)p(x)}.$$

For example when x can take only two values: $p(x_0|y_1) = \dfrac{p(y_1|x_0)p(x_0)}{p(y_1|x_0)p(x_0) + p(y_1|x_1)p(x_1)}$

Using this we get for our variables:

$$p(D|+) \quad = \frac{p(+|D)p(D)}{p(+)} = \frac{0.95 * 0.01}{0.95 * 0.01 + 0.001 * 0.99} = 0.906$$

and therefore $p(\bar{D}|+) = 1 - p(D|+) = 0.094$.

On the other hand

$$p(\bar{D}|-) \quad = \frac{p(-|\bar{D})p(\bar{D})}{p(-)} = \frac{0.999 * 0.99}{0.999 * 0.99 + 0.05 * 0.01} = 0.9995$$

and therefore $p(D|-) = 1 - p(\bar{D}|-) = 0.0005$

Total 10 points.