Sheet 01

06 November 2020 15:39

Group: MI4

Machine Intelligence 1 **Exercise Sheet 1**
WS 2020/21, Obermayer/Kashef due: 09.11.2019 at 23:55 (bonus sheet)

Maths primer

Exercises prefixed with **T** are be explained by the tutor 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 $\text{var}(X) = \langle X^2 \rangle_p - \langle X \rangle_p^2$ for simplicity.

a) $\int_0^{\pi} c \cdot \sin(x)\, dx = c \cdot -\cos x \Big|_0^{\pi} = 2c = 1$

$\underline{\underline{c = 1/2}}$

value is an integral of a position weighted by distribution, hence

b) $E(x) = \int_0^{\pi} x\, p(x)\, dx = \frac{1}{2}\left[x(-\cos(x)) \Big|_0^{\pi} + \int 1 \cos(x)\, dx \right]$

$= \frac{1}{2}\left[x(-\cos(x)) \Big|_0^{\pi} + \sin x \Big|_0^{\pi} \right] = \frac{\pi}{2}$

Variance is an estimated value of a squared deviation

$$\text{c)}\quad var(x) = E(x^2) - [E(x)]^2$$

$$\Rightarrow \frac{1}{2}\int_0^{\pi} x^2 \sin x \, dx - \left(\frac{\pi}{2}\right)^2$$

$$\Rightarrow \frac{1}{2}\left[x^2(-\cos x)\Big|_0^{\pi} + \int_0^{\pi} x \cos x \, dx \right] - \frac{\pi^2}{4}$$

Calculating by parts (from the derivative of the product rule)

$$\Rightarrow \frac{1}{2}\left[\pi^2(+1) - 0 + \int_0^{\pi} x \, d(\sin x) \right] - \frac{\pi^2}{4}$$

$$\Rightarrow \frac{1}{2}\left[\pi^2 + x \sin x \Big|_0^{\pi} - \int_0^{\pi} \sin x \, dx \right] - \frac{\pi^2}{4}$$

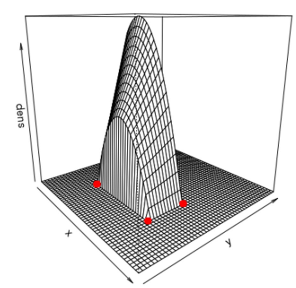
$$\Rightarrow \frac{1}{2}\left[\pi^2 + \cos x \Big|_0^{\pi} \right] - \frac{\pi^2}{4}$$

$$= \frac{\pi^2}{2} + \frac{(-2)}{2} - \frac{\pi^2}{4} = \frac{\pi^2}{4} - 1$$

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.

$$\text{a)}\quad p_X(x) = \int_0^1 p_Z(z) \, dy = \frac{3}{7}\int_0^1 (2-x)(x+y) \, dy$$

$$= \frac{3}{7}(2-x) \int_0^1 (x+y) \, dy$$

$$= \frac{3}{7}(2-x)\left[xy + \frac{y^2}{2} \Big|_0^1 \right]$$

$$= \frac{3}{7}(2-x)\left(x + \frac{1}{2}\right)$$

$$P_y(y) = \int_0^2 P_z(z)\, dx$$

$$= \int_0^2 \frac{3}{7}(2-x)(x+y)\, dx$$

$$= \frac{3}{7}\int_0^2 (2x + 2y - x^2 - xy)\, dx$$

$$= \frac{3}{7}\int_0^2 (2x + 2y - x^2 - xy)\, dx$$

$$= \frac{3}{7}\left[\frac{2x^2}{2} + 2xy - \frac{x^2 y}{2} - \frac{x^3}{3}\right]_0^2 = \frac{3}{7}\left[4 + 4y - 2y - \frac{8}{3}\right]$$

$$= \frac{3}{7}\left[\frac{4}{3} + 2y\right] = \left(\frac{4}{7} + \frac{6y}{7}\right)$$

b) Observing,

$$P_X(x) \cdot P_Y(y)$$

$$= \frac{3}{7}\left[(2-x)\left(x+\frac{1}{2}\right)\right]\frac{1}{7}\left[4+6y\right.$$

$$= \frac{3}{49}(2-x)(4x + 6xy + 2 + 3y)$$

$$\neq P_{X,Y}(x,y) \Rightarrow \underline{\text{correlated}}$$

$$C_v(x,y) := \iint_{x\ y} P_X(x,y) \cdot x \cdot y\, dx\, dy - \left[\int_x P_X(x)\, x\, dx\right]\left[\int_y P_y\, y\, dy\right]$$

$$= \int_{x,y} \frac{3}{7}(2-x)(x+y)\, xy\, dx\, dy - \left[\int \frac{3}{7}(2-x)\left(x+\frac{1}{2}\right)x\, dx\right]\left[\int\left(\frac{4}{7} + \frac{6y}{7}\right) y\, dy\right]$$

$$= \int \frac{3}{7}(2-x)\, x\, dx \int_0^1 (x+y)\, y\, dy - \frac{3}{7}\int_0^2\left(2x^2 + x - x^3 - \frac{x^2}{2}\right) dx \int\left(\frac{4y}{7} + \frac{6y^2}{7}\right) dy$$

$$= \int \frac{3}{7}(2-x)\, x\left(\frac{xy^2}{2} + \frac{y^3}{3}\Big|_0^1\right) dx - \frac{3}{7}\left[\frac{3}{2}\cdot\frac{x^3}{3} + \frac{x^2}{2} - \frac{x^4}{4}\right]_0^2\left[\frac{4y^2}{7.2} + \frac{6y^3}{7.3}\right]$$

$$= \int \frac{3}{7}(2-x)\left(\frac{x^2}{2} + \frac{x}{3}\right) dx - \frac{3}{7}\left[4 + 2 - 4\right]\left[\frac{2}{7} + \frac{2}{7}\right]$$

$$= \frac{3}{7}\int\left(\cancel{x}^2 + \frac{2x}{3} - \frac{x^3}{2} - \frac{x^2}{3}\right) dx - \frac{3}{7}\cdot\frac{8}{7}$$

$$= \frac{3}{7}\left[\frac{x^3}{3} + \frac{x^2}{3} - \frac{x^4}{8} - \frac{x^3}{9}\right]_0^2 - \frac{24}{49} = \frac{3}{7}\left[\frac{8}{3} + \frac{4}{3} - 2 - \frac{8}{9}\right] - \frac{24}{49}$$

$$= \frac{1}{7}\left[\frac{24+12-18-8}{3}\right] - \frac{24}{49}$$

$$= \frac{14}{21} - \frac{24}{49} \neq 0$$

$$\Rightarrow \underline{\text{dependent}}$$

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.

$$T_f^3 = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)(x-x_0)^2}{2} + \frac{f'''}{6}(x-x_0)^3$$

$f(x_0) = \sqrt{1+0} = 1$

$f'(x) = (1+x)^{1/2} = \frac{1}{2}(1+x)^{-1/2}$ $\quad\therefore f'(x_0) = \frac{1}{2}$

$f''(x) = \frac{1}{2}\cdot\frac{-1}{2}(1+x)^{-3/2}$ $\quad\therefore f''(x_0) = \frac{-1}{4}$

$f'''(x) = \left(\frac{-1}{4}\right)\cdot\left(\frac{-3}{2}\right)\cdot(1+x)^{-5/2}$ $\quad\therefore f'''(x_0) = \frac{3}{8}$

$$\therefore T_f^3 = 1 + \frac{1}{2}(x) - \frac{x^2}{8} + \frac{3}{48}x^3$$

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).

$$\begin{aligned}\det(A) &= 5(-11+32) - 4(-88+64) - 4(64-16) \\ &= 5(21) - 4(-24) - 4(48) \\ &= 105 + 96 - 192 = 9\end{aligned}$$

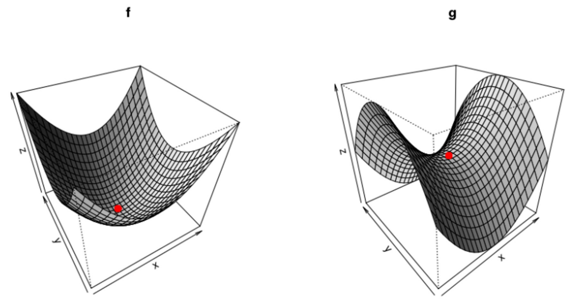
$$\mathrm{Trace}(A) = 5 + 1 + -11 = -5$$

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).

a)

"a" is a critical point, if the derivatives are equal zero.

$$\frac{\partial f}{\partial x} = 2x, \quad \frac{\partial f}{\partial y} = 2y \qquad \vec{\nabla} f\Big|_{0,0} = \begin{pmatrix} 2x \\ 2y \end{pmatrix}\Big|_{0,0} = \vec{0}$$

$$\frac{\partial g}{\partial x} = 2x, \quad \frac{\partial g}{\partial y} = -2y \qquad \vec{\nabla} g\Big|_{0,0} = \begin{pmatrix} 2x \\ -2y \end{pmatrix}_{0,0} = \vec{0}$$

b)

As f is a polynomial, it is infinetely differentiable. It has an extremum at "a" only if it's Hessian matrix is strictly positive definite (local minimum) or strictly negative definite (local maximum). In other cases it has no extremum.

$$H_f = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} : \lambda = 2, \text{ multiplicity} = 2$$

$\lambda > 0 \rightarrow$ stable, minimum

$$H_g = \begin{pmatrix} \frac{\partial^2 g}{\partial x^2} & \frac{\partial^2 g}{\partial x \partial y} \\ \frac{\partial^2 g}{\partial y \partial x^2} & \frac{\partial^2 g}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} : \lambda_1 = 2, \lambda_2 = -2$$

unstable saddle point

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 $P(D|-)$.

$$P(a|b) = \frac{P(a \cap b)}{P(b)}$$

$$P(D) = 1\%, \quad P(\bar{D}) = 99\%$$

$$P(+|D) = 95\% \quad ; \quad P(-|\bar{D}) = 99.9\%$$

$$\downarrow \qquad\qquad\qquad \downarrow$$

$$P(-|D) = 5\% \quad ; \quad P(+|\bar{D}) = 0.1\%$$

a) $$\underline{P(D|+)} = \frac{P(+|D) \cdot P(D)}{P(+)}$$

$$P(D|+) = \frac{95/100}{1049/10^5} \cdot \frac{1}{10^2} = \frac{95 \cdot 10}{1049} \approx \underline{0.906},$$

$$P(+) = P(D)\,P(+|D) + P(\bar{D}) \cdot P(+|\bar{D})$$

$$= 1\% \cdot 95\% + 99\% \cdot 0.1\%$$

$$= \frac{95}{10000} + \frac{99}{100000} = \frac{1049}{100000}$$

$$\underline{P(\bar{D}|+)} = \frac{P(+|\bar{D})\, P(\bar{D})}{P(+)}$$

$$= \frac{0.1/100 \cdot \frac{99}{100}}{1049/10^5}$$

$$= \frac{99}{1049} \approx \underline{0.0944}$$

$P(\bar{D}|+) + P(D|+) = 1.$

b) $P(D|-) = \dfrac{P(-|D) \cdot P(D)}{P(-)}$

$= \dfrac{5\% \quad 1\%}{1 - \frac{1049}{10^5}} = 0.000505$

$P(D|-) + P(\bar{D}|-) = 1$

$P(\bar{D}|-) = \dfrac{P(-|\bar{D}) \cdot P(\bar{D})}{P(-)}$

$= \dfrac{99.9\% \cdot 99\%}{1 - \frac{1049}{10^5}} = 0.99949$