Machine Intelligence 1 - Exercise 1: Math Primer

Liu, Zhiwei 387571 | Moon, Chulhyun 392865 | Wenzel, Daniel 365107 | Ozmen, Cengizhan 388011

Pipo, Aiko 390011

October 25, 2017

---

1 Distributions and expected values (2 Points)

1.1 (a)

As p is obviously piecewise continuous, it is measurable for every c $\in \mathbb{R}$. One has

$$\int_{\mathbb{R}} p(t)\,\mathrm{d}t \;=\; \int_{0}^{\pi} c \cdot sin(t)\,\mathrm{d}t \;=\; c \cdot \int_{0}^{\pi} sin(t)\,\mathrm{d}t \;=\; c(-cos)(t)\big]_{t=0}^{\pi} \;=\; -c(-1-1)) \;=\; 2c.$$

$$\text{Therefore the condition } \int_{\mathbb{R}} p(t)\,\mathrm{d}t \;=\; 1 \text{ is satisfied if and only if c} = \frac{1}{2}\,.$$

As sin(x) ≥ 0 holds for all x $\in [0, \pi]$, $\frac{1}{2}$sin(x) ≥ 0 is satisfied for all x $\in [0, \pi]$, so p is truly a probability density for c $= \frac{1}{2}$.

1.2 (b)

For c $= \frac{1}{2}$ we obtain

$$\langle X \rangle_p \;=\; \int_{\mathbb{R}} t \cdot p(t)\,\mathrm{d}t \;=\; \frac{1}{2}\int_{0}^{\pi} t \cdot sin(t)\,\mathrm{d}t \;=\; \frac{1}{2}\left(-tcos(t)\big]_{t=0}^{\pi} - \int_{0}^{\pi} -cos(t)\,\mathrm{d}t\right) \;=\; \frac{\pi}{2} + \frac{1}{2}sin(t)\big]_{t=0}^{\pi} \;=\; \frac{\pi}{2}.$$

1.3 (c)

We have

$$\langle X^2 \rangle_p \;=\; \int_{\mathbb{R}} t^2 \cdot p(t)\,\mathrm{d}t \;=\; \int_{\nvdash}^{\pi} t^2 \cdot p(t)\,\mathrm{d}t \;=\; \frac{1}{2}\int_{0}^{\pi} t^2 \cdot sin(t)\,\mathrm{d}t \;=$$

$$=\; -t^2cos(t)\big]_{t=0}^{\pi} - 2\int_{0}^{\pi} -t \cdot cos(t)\,\mathrm{d}t = \frac{1}{2}\left(-\pi^2(-1) + 0 + 2\int_{0}^{\pi} t \cdot cos(t)\,\mathrm{d}t\right) \;=$$

$$=\; \frac{\pi^2}{2} + \left(tsin(t)\big]_{t=0}^{\pi} - \int_{0}^{\pi} sin(t)\,\mathrm{d}t\right) \;=\; \frac{\pi^2}{2} + cos(t)\big]_{t=0}^{\pi} \;=\; \frac{\pi^2}{2} - 1 - 1 \;=\; \frac{\pi^2}{2} - 2$$

and consequently V(X) $= \langle X^2 \rangle_p - \langle X \rangle_p^2 = \frac{\pi^2}{2} - 2 - (\frac{\pi}{2})^2 = \frac{\pi^2-8}{4}$.

2 Marginal densities (2 Points)

2.1 (a)

$p_X(x) = \int_{\mathbb{R}} p_{X,Y}(x,y)dy$ for $x \in [0,2]$

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

$$= \frac{3}{7}(2xy + y^2 + -x^2y - \frac{xy^2}{2})|_{y=0}^1 = -\frac{3}{7}x^2 + \frac{9}{14}x + \frac{3}{7}$$

$p_X(x) = -\frac{3}{7}x^2 + \frac{9}{14}x + \frac{3}{7}$ for $x \in [0,2]$

$p_Y(y) = \int_{\mathbb{R}} p_{X,Y}(x,y)dx$ for $y \in [0,1]$

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

$$= \frac{3}{7}(x^2 + 2xy - \frac{x^3}{3} - \frac{x^2y}{2})|_{x=0}^2 = \frac{6}{7}y + \frac{4}{7}$$

$p_Y(y) = \frac{6}{7}y + \frac{4}{7}$ for $y \in [0,1]$

2.2 (b)

If random variables X and Y are independent, $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ yields. Let's check:

$$p_X(x)p_Y(y) = (-\frac{3}{7}x^2 + \frac{9}{14}x + \frac{3}{7})(\frac{6}{7}y + \frac{4}{7}) for x \in [0,2], y \in [0,1]$$

For $x = y = 0$,

$$p_Y(0)p_X(0) = \frac{12}{49} \neq 0 = p_{X,Y}(0,0)$$

Hence, **not independent**.

Let's check uncorrelatedness:
If two random variables X and Y are uncorrelated, $E(XY) = E(X)E(Y)$ yields.

$$E(X) = \int_{\mathbb{R}} x \cdot p_X(x)dx = \int_0^2 x \cdot \frac{3}{7}(-x^2 + \frac{3}{2}x + 1)dx$$

$$= \frac{3}{7}(-\frac{1}{4}x^4 + \frac{1}{2}x^3 + \frac{1}{2}x^2)|_{x=0}^2 = \frac{6}{7}$$

$$E(Y) = \int_{\mathbb{R}} y \cdot p_Y(y)dx = \frac{3}{7}\int_0^1 (2y^2 + \frac{4}{3}y)dy = \frac{3}{7}(\frac{2}{3}y^3 + \frac{2}{3}y^2)|_{y=0}^1 = \frac{4}{7}$$

$$E(XY) = \int_0^2 \int_0^1 x \cdot y \cdot \frac{3}{7}(2-x)(x+y)dydx = \frac{3}{7}\int_0^2 x\{\int_0^1 2xy - x^2y + 2y^2 - xy^2 dy\}dx$$

$$= \frac{3}{7}\int_0^2 x \cdot [xy^2 - \frac{1}{2}x^2y^2 + \frac{2}{3}y^3 - \frac{1}{3}xy^3]_0^1 dx = \frac{3}{7}\int_0^2 \frac{2}{3}x^2 - \frac{1}{2}x^3 + \frac{2}{3}xdx$$

$$= \frac{3}{7}\{\frac{16}{9} - \frac{16}{9} + \frac{4}{3}\} = \frac{10}{21}$$

So, $E(XY) \neq E(X)E(Y)$. Then, **not uncorrelated**.

3 Taylor expansion (1 Point)

Let $t_0(x)$ be the taylor expansion to the third degree at $x_0 = 0$ for $f(x) = \sqrt{1+x}$

Using the definition of the tayler expansion we know that

$$t_0(x) = \sum_{n=1}^{3} \frac{f^n(0) * (x-0)^n}{n\,!} = \frac{1}{0!}f(0)x^0 + \frac{1}{1!}f'(0)x^1 + \frac{1}{2!}f''(0)x^2 + \frac{1}{3!}f'''(0)x^3$$

$$= f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + \frac{1}{6}f'''(0)x^3$$

The derivatives of f(x) can be computed using the chain rule:

$$f'(x) = \sqrt{1+x} \implies f'(0) = 1$$

$$f'(x) = 0.5(x+1)^{-0.5} \implies f'(0) = 0.5$$

$$f''(x) = -0.25(x+1)^{-1.5} \implies f''(0) = -0.25$$

$$f'''(x) = \frac{3}{8}(x+1)^{-2.5} \implies f'''(0) = \frac{3}{8}$$

Therefore:

$$t_0(x) = 1 + 0.5x + \frac{1}{2}(-0.25)x^2 + \frac{1}{6}\frac{3}{8}x^3$$

$$= 1 + 0.5x - \frac{1}{8}x^2 + \frac{1}{16}x^3$$

4 Determinant of a matrix (1 Point)

We define

$$A := \begin{bmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{bmatrix}$$

Then the rule of Sarrus delivers

$$det(A) = 5 \cdot 1 \cdot (-11) + 8 \cdot 8 \cdot (-4) + 16 \cdot 4 \cdot (-4) - (-4) \cdot 1 \cdot 16 - (-4) \cdot 8 \cdot 5 - (-11) \cdot 4 \cdot 8$$

$$= -55 - 256 - 256 + 64 + 160 + 352 = -567 + 576 = 9.$$

Furthermore we have trace(A) = 5 + 1 - 11 = -5

5 Critical points (2 Points)

5.1 (a)

First, we compute the gradients

$$\nabla f(x) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}, \nabla g(x) = \begin{pmatrix} 2x \\ -2y \end{pmatrix}$$

and find their roots:

$$\nabla f(x,y) = 0 \Leftrightarrow (x,y) = (0,0) = \underline{a}$$

$$\nabla g(x,y) = 0 \Leftrightarrow (x,y) = (0,0) = \underline{a}$$

This shows that $\underline{a}$ is indeed a critical point for f and g.

5.2 (b)

Computing the Hessian matrices yields:

$$H_f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, H_g(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

We can immediatly see, that H_f is positive definite (both eigenvalues bigger than zero) and H_g is indefinite (one eigenvalue bigger than zero, one less than zero). We deduce, that $\underline{a}$ is a minimum for f but a saddle point for g.

6 Bayes rule (2 Points)

We know that:

$$P(D) = 0.01 \implies P(\overline{D}) = 0.99$$
$$P(+|D) = 0.95 \implies P(-|D) = 0.05$$
$$P(-|\overline{D}) = 0.999 \implies P(+|\overline{D}) = 0.001$$

Therefore we can calculate

$$P(+) = P(D) \times P(+|D) + P(\overline{D}) \times P(+|\overline{D}) = 0.01 \times 0.95 + 0.99 \times 0.001 = 0.01049$$

Therefore $P(-) = 1 - P(+) = 0.98951$

6.1 (a)

$$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{P(+|D) \times P(D)}{P(+)} = \frac{0.95 \times 0.01}{0.01049} \approx 0.9056$$
$$P(\overline{D}|+) = 1 - P(D|+) \approx 0.09438$$

6.2 (b)

$$P(\overline{D}|-) = \frac{P(-|\overline{D}) \times P(\overline{D})}{P(-)} = \frac{0.99 \times 0.999}{0.98951} \approx 0.9995$$
$$P(D|-) = 1 - P(\overline{D}|-) \approx 0.0005053$$