INF8245E: Machine learning
Assignment #0

Frédéric Marcotte*
Polytechnique Montréal, Département de génie physique, Montréal, H3T 1J4, Canada
(Dated: Sept 14, 10 pm)

- **Honour Code:**

  By enrolling for INF8245E Machine Learning course, I agree that all the work submitted will be mine and original, and will not be plagiarized. Unless otherwise specifically stated by the instructor or TAs, I will not collaborate with anyone on my assignments or tests. I understand that any violation of this honor code will be strictly dealt with.

  Frédéric Marcotte

PROBABILITY THEORY

1. (a) (1 point) Two fair dice are rolled. What is the probability that their sum is greater than 4?
   Let X, Y be two discrete random variables defined on the outcome space of each dice throw. The probability that their sum is greater than 4 is given by:

$$P(X+Y>4) = 1 - P(X+Y \leq 4). \tag{1}$$

   There are 6 combinations that yield a sum lower or equal to 4 with equal probabilities:

$$P(\{1,1\}) + P(\{1,2\}) + P(\{2,1\}) + P(\{2,2\}) + P(\{1,3\}) + P(\{3,1\}) = \frac{6}{36}. \tag{2}$$

   Then, the probability is:

$$P(X+Y>4) = 1 - \frac{6}{36} = \frac{30}{36} = \frac{5}{6}. \tag{3}$$

   (b) (1 point) A probability experiment has four possible outcomes: e_1, e_2, e_3, e_4. The outcome e_1 is four times as likely as each of the three remaining outcomes. Find the probability of e_1.
   The equations to solve are:

$$P(e_1) + P(e_2) + P(e_3) + P(e_4) = 1, \tag{4}$$
$$P(e_1) = 4P(e_2) = 4P(e_3) = 4P(e_4). \tag{5}$$

   It is straightforward to see that:

$$P(e_1) = \frac{4}{7}. \tag{6}$$

   (c) (3 points) Which of the following three events is more likely that a person gets...

   i. exactly 1 six when 6 dice are rolled:

$$P = \binom{6}{1}(1/6)(5/6)^5 = 0.402. \tag{7}$$

   ii. exactly 2 six when 12 dice are rolled:

$$P = \binom{12}{2}(1/6)^2(5/6)^{10} = 0.296. \tag{8}$$

* frederic.marcotte@polymtl.ca

iii. exactly 3 six when 18 dice are rolled:

$$P = \binom{18}{3}(1/6)^3(5/6)^{15} = 0.245. \tag{9}$$

The event that a person gets exactly 1 six when 6 dice are rolled is more likely.

2. (3 points) If $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, and $P(B|A) = \frac{1}{3}$, find the following:

(a) $P(A \text{ and } B)$

$$P(A \text{ and } B) = P(B|A)P(A) = (1/3)(1/2) = \frac{1}{6}. \tag{10}$$

(b) $P(A \text{ or } B)$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = (1/2) + (1/2) - (1/6) = \frac{5}{6}. \tag{11}$$

(c) $P(A|B)$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{(1/3)(1/2)}{(1/2)} = \frac{1}{3}. \tag{12}$$

3. (a) (3 points) Suppose the probability mass function of the discrete random variable is

x	$p(x)$
0	0.2
1	0.1
2	0.4
3	0.3

TABLE I. Probabilities of the discrete random variable

What is the value of $\mathbb{E}[3X + 2X^2]$?

$$\begin{aligned}\mathbb{E}[3X + 2X^2] &= 3\mathbb{E}[X] + 2\mathbb{E}[X^2] \\ &= 3\,(0 \cdot 0.2 + 1 \cdot 0.1 + 2 \cdot 0.4 + 3 \cdot 0.3) + 2\left(0^2 \cdot 0.2 + 1^2 \cdot 0.1 + 2^2 \cdot 0.4 + 3^2 \cdot 0.3\right) \\ &= 14.2. \end{aligned} \tag{13}$$

(b) (3 points) Let a probability density function of a random variable X be $f(x) = 4x^3$ for $0 < x < 1$. Find $\mathbb{E}[X]$ and $Var(X)$.

$$\mathbb{E}[X] = \int_0^1 x f(x)\,\mathrm{d}x = \left[\frac{4}{5}x^5\right]_0^1 = \frac{4}{5}. \tag{14}$$

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_0^1 x^2 f(x)\,\mathrm{d}x - (4/5)^2 = \left[\frac{4}{6}x^6\right]_0^1 - (4/5)^2 = (4/6) - (4/5)^2 = \frac{2}{75}. \tag{15}$$

4. Let X, Y be two continuous random variables with joint PDF:

$$f_{X,Y}(x,y) = \begin{cases} cx + 1 & x, y \geq 0, x + y < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{16}$$

(a) (2 points) Find constant c.

$$\begin{aligned}
\int_0^1 \int_0^{1-x} (cx+1)\,\mathrm{d}y\mathrm{d}x &= \int_0^1 (cx+1)\,[y]_0^{1-x}\,\mathrm{d}x \\
&= \int_0^1 \left(-cx^2 + (c-1)x + 1\right)\,\mathrm{d}x \\
&= \left[-c\frac{x^3}{3} + (c-1)\frac{x^2}{2} + x\right]_0^1 \\
&= \frac{1}{6}c + \frac{1}{2} \\
&= 1.
\end{aligned} \tag{17}$$

The last equality is obtained from the *sum-to-one* property of a probability density function. Thus, we find:

$$c = 3. \tag{18}$$

(b) (2 points) Find the marginal PDFs $f_X(x)$ and $f_Y(y)$. Are X and Y independent?

$$\begin{aligned}
f_X(x) &= \int_0^{1-x} f_{X,Y}(x,y)\,\mathrm{d}y \\
&= (3x+1)\,[y]_0^{1-x} \\
&= (3x+1)(1-x) \\
&= -3x^2 + 2x + 1 \text{ for } 0 < x < 1.
\end{aligned} \tag{19}$$

$$\begin{aligned}
f_Y(y) &= \int_0^{1-y} f_{X,Y}(x,y)\,\mathrm{d}x \\
&= \left[\frac{3}{2}x^2 + x\right]_0^{1-y} \\
&= \frac{3}{2}(1-y)^2 + (1-y) \\
&= \frac{3}{2}y^2 - 4y + \frac{5}{2} \text{ for } 0 < y < 1.
\end{aligned} \tag{20}$$

X and Y are not independent since:

$$f_{X,Y}(x,y) \neq f_X(x)f_Y(y). \tag{21}$$

(c) (2 points) Find $P(Y < 2X^2)$.

$$\begin{aligned}
P(Y < 2x^2 | X = x) &= \int_0^{2x^2} \frac{f_{X,Y}(x,y)}{f_X(x)}\,\mathrm{d}y \\
&= \int_0^{2x^2} \frac{(3x+1)}{(3x+1)(1-x)}\,\mathrm{d}y \\
&= \frac{1}{1-x}[y]_0^{2x^2} \\
&= \frac{2x^2}{1-x} \text{ for } 0 < x < 1/2.
\end{aligned} \tag{22}$$

LINEAR ALGEBRA

5. (3 points) Find the inverse of the matrix:

$$A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ -1 & 1 & 2 & 0 \\ -2 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We use the formula:

$$A^{-1} = \frac{1}{|A|}\mathrm{adj}(A), \tag{23}$$

where

$$|A| = \begin{vmatrix} 1 & 0 & 1 \\ -1 & 1 & 2 \\ -2 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} + \begin{vmatrix} -1 & 1 \\ -2 & 0 \end{vmatrix} = 1 + 2 = 3, \tag{24}$$

and the elements of $\mathrm{adj}(A)$ are

$$(\mathrm{adj}(A))_{ij} = (-1)^{i+j}|A_{\backslash j,\backslash i}|, \tag{25}$$

and $|A_{\backslash j,\backslash i}|$ are computed with the recursive formula for the determinant as in Eq. (24). We find:

$$A^{-1} = \frac{1}{3}\begin{bmatrix} 1 & 0 & -1 & 0 \\ -3 & 3 & -3 & 12 \\ 2 & 0 & 1 & -6 \\ 0 & 0 & 0 & 3 \end{bmatrix}. \tag{26}$$

6. (2 points) Determine the component vector of the vector $V = (1, 7, 7)$ in $\mathbb{R}^3$ relative to the desired basis $B = \{(1, -6, 3), (0, 5, -1), (3, -1, -1)\}$.
We solve the equation $Ax = y$:

$$\begin{bmatrix} 1 & 0 & 3 \\ -6 & 5 & -1 \\ 3 & -1 & -1 \end{bmatrix}\begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ 7 \end{bmatrix}. \tag{27}$$

We find $x_{11} = 4$, $x_{21} = 6$, and $x_{31} = -1$. Thus, the vector V is:

$$4(1, -6, 3) + 6(0, 5, -1) - (3, -1, -1), \tag{28}$$

that we can also express as the component vector $(4, 6, -1)$ relative to the basis B.

7. (5 points) Find eigenvalues and linearly independent eigenvectors for the matrix:

$$A = \begin{bmatrix} 4 & -3 & -3 \\ 3 & -2 & -3 \\ -1 & 1 & 2 \end{bmatrix}. \tag{29}$$

Using the eigenvectors, diagonalize the matrix A.
We solve the equation:

$$\begin{aligned} |(\lambda I - A)| &= (\lambda - 4)\begin{vmatrix} \lambda + 2 & 3 \\ -1 & \lambda - 2 \end{vmatrix} - 3\begin{vmatrix} -3 & 3 \\ 1 & \lambda - 2 \end{vmatrix} + 3\begin{vmatrix} -3 & \lambda + 2 \\ 1 & -1 \end{vmatrix} \\ &= (\lambda - 4)(\lambda^2 - 1) - 3(-3\lambda + 3) + 3(-\lambda + 1) \\ &= \lambda^3 - 4\lambda^2 + 5\lambda - 2 \\ &= (\lambda - 2)(\lambda - 1)^2 \\ &= 0. \end{aligned} \tag{30}$$

The solutions are $\lambda_1 = 1$, $\lambda_2 = 1$, and $\lambda_3 = 2$. To find the eigenvectors x_i, we solve the linear equation $(\lambda_i I - A)x_i = 0$. For λ_3, we solve:

$$\begin{bmatrix} -2 & 3 & 3 \\ -3 & 4 & 3 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_{13} \\ x_{23} \\ x_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{31}$$

We find $x_3 = (-3, -3, 1)$. For $\lambda_1 = \lambda_2 = 1$, we obtain $x_1 = (1, 0, 1)$ and $x_2 = (1, 1, 0)$.
Using the eigenvectors, we diagonalize the matrix A:

$$A = X\Lambda X^{-1} = \begin{bmatrix} 1 & 1 & -3 \\ 0 & 1 & -3 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ -3 & 4 & 3 \\ -1 & 1 & 1 \end{bmatrix}, \tag{32}$$

where the columns of $X \in \mathbb{R}^{3\times 3}$ are the eigenvectors of A and $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)$.

8. (2 points) You are given that the eigenvalues of a matrix A are 3, 2 and 2. Is A invertible?

   Yes. The number of non-zero eigenvalues of $A \in \mathbb{R}^{3\times 3}$ is 3, thus $\operatorname{rank}(A) = 3$. Since A is full rank, then A^{-1} exists.

9. (4 points) For a matrix A of size $n \times n$ you are told that all its n eigenvectors are independent. Let S denote the matrix whose columns are the n eigenvectors of A.

   (a) Is A invertible?
   Depends. If there is one eigenvalue equal to zero, then A is non-invertible. Otherwise, A is invertible.

   (b) Is A diagonalizable?
   Yes. If the eigenvectors of A are linearly independent, then A is diagonalizable.

   (c) Is S invertible?
   Yes. If the columns of $S \in \mathbb{R}^{n\times n}$ are linearly independent eigenvectors, then S is invertible.

   (d) Is S diagonalizable?
   Depends. If the eigenvectors of S are not linearly independent, then S is non-diagonalizable. Otherwise, S is diagonalizable.

REAL ANALYSIS

10. (2 points) Compute $\frac{\mathrm{d}y}{\mathrm{d}x}$ for the following.

   (a) $y = x^4\left(\sin(x^3) - \cos(x^2)\right)$

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= 4x^3\left(\sin(x^3) - \cos(x^2)\right) + x^4\left(3x^2\cos(x^3) + 2x\sin(x^2)\right) \\ &= x^3\left(4\sin(x^3) - 4\cos(x^2) + 3x^3\cos(x^3) + 2x^2\sin(x^2)\right) \end{aligned} \tag{33}$$

   (b) $y = \ln(x^2)$

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{2x}{x^2} = \frac{2}{x} \tag{34}$$

11. (2 points) Compute the critical points of the function $f(x) = x^3 - 6x^2 + 9x + 15$ by using first order derivatives. Use second order derivatives to identify which of the critical points are minima and maxima.

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = 3x^2 - 12x + 9 = 3(x-1)(x-3). \tag{35}$$

The critical points are $x = 1$ and $x = 3$.

$$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} = 6x - 12. \tag{36}$$

$f''(x = 1) < 0$, thus $x = 1$ is a maxima. $f''(x = 3) > 0$, thus $x = 3$ is a minima.