MO435 — Probabilistic Machine Learning
INSTITUTO DE COMPUTAÇÃO — UNICAMP

Introduction, review of probability and Bayes

Lesson No. 1

Patrick de Carvalho Tavares Rezende Ferreira - 175480

1 Abstract

In this class, we approach basic concepts about machine lerning, probability and Bayes theorem.

2 Machine Learning

A brief definition of what machine learning is could be said to be "a set of methods that can automatically find patterns in data and take correct actions or predictions about them".

3 Probability

We define p(A) as the probability for an event "A" to happen, where p(A) ranges from 0 to 1 (0 means no chance to happen and 1 means this event is going to happen anyway). p(X=x) means p(x).

A random variable (r.v.) is a function associating every element from sample space to a real number. If you integrate all values for a random variable distribution, it must sum 1. For example, on a die roll, there is a $p(x) = \frac{1}{6}$ chance that each of the six faces will fall. If we add the total probability of all the possibilities of this random variable, we will have 1 (100% chance of any of the faces coming out).

3.1 Clarifying notations

Something that is important to remember is the precedence of the operators: $p(x|y = c)$ means the probability of x given that y is equal to c, while $p(y = c|x, \Theta)$ represents the probability of y being equal to c, given x and theta.

4 Rules

Here we have some basic rules for two events called "A" and "B":

- $p(A \cup B) = p(A) + p(B) - p(A \cap B)$, but $p(A \cap B) = 0$ if A and B are mutually exclusive.
- $p(A, B) = p(A \cap B) = p(A|B) * p(B)$
- $p(A) = \sum_i p(A, Bi) = \sum_i p(A|Bi) * p(Bi)$
- $p(A|B) = \frac{p(A,B)}{p(B)} = \frac{p(B|A)*p(A)}{p(B)}$ (Bayes theorem)

5 Independence

We can say two variables, are unconditionally independent, conditinally independent or dependent.

- Unconditionally independent: $X \perp Y \Leftrightarrow p(x, y) = p(x) * p(y)$
- Conditionally independent: $X \perp Y|Z \Leftrightarrow p(x, y|Z) = p(x|Z) * p(y|Z)$

6 Continuous random variables

Suppose X is some uncertain continuous quantity. The probability that X lies in any interval $a \le X \le b$ can be computed as follows. Define the events $A = (X \le a)$, $B = (X \le b)$ and W = $(a \le X \le b)$. We have that $B = A \cup W$, and since A and W are mutually exclusive, the sum rules gives:

$$p(B) = p(A) + p(W) \rightarrow p(W) = p(B) - p(A) \tag{1}$$

6.1 Cumulative distributions function (cdf)

$$F(a) = f(x \le a) \tag{2}$$

$$f(a < x \le b) = F(b) - F(a) \tag{3}$$

6.2 Probability density function (pdf)

Given a pdf, we can compute the probability of a continuous variable being in a finite interval as follows:

$$f(a) = \frac{\mathrm{d}}{\mathrm{d}a} F(a) \tag{4}$$

$$P(a \le X \le b) = \int_a^b f(x) dx \tag{5}$$

$$P(x \le X \le x + dx) \approx p(x) dx \, (differential \, of \, probability) \tag{6}$$

7 Expectation

Expectation may be thought as the value you expect for a given distribution. It's usefull, for example, when you want to know the weighted average of a distribution.

$$E[X] = \sum_{x \epsilon X} x * p(x) = \int_x x * p(x) dx \tag{7}$$

The equation above is called "first moment of expectation". The equation below is usefullwhen working with distributionsof distributions.

$$E[f(x)] = \int_x f(x) * p(x) dx \tag{8}$$

We also have "central moments of expectation". The second central moment of expectation is shown bellow and represents the variance for a normal distribution.

$$E[(X-\mu)^2] = \sigma^2 = E[X^2 - 2 * X * \mu + \mu^2] = E[X^2] - 2 * E[X * \mu] + E[\mu^2] = E[X^2] - 2 * \mu^2 + \mu^2 = E[X^2] - \mu^2, \tag{9}$$

where "μ" is the mean for this distribution.

8 Covariance

$$COV(X, Y) = E[(X - E[X]) * (Y - E[Y])] = E[XY] - E[X] * E[Y] \tag{10}$$

9 Monte Carlo

The Monte Carlo approximation consists of a method based on successive samplings to estimate a value.

$$E[f(x)] = \int_x f(x) * p(x)dx = \frac{1}{S} * \sum_S f(x_s) \tag{11}$$

10 Entropy

Entropy can be thought as a measure of uncertainty.

$$H(x) = -\int (p(x) - log(p(x)))dx \tag{12}$$