#### SOLVING CHALLENGES IN BUSINESS STRUCTURES WITH THE HELP OF GENERATIVE MODELS

Marian Sorin IONESCU<sup>a\*</sup>, Olivia NEGOIȚĂ<sup>b</sup>

<sup>a,b</sup>University "Politehnica" Of Bucharest, Romania <sup>a</sup>National Institute For Reserche And Development In Microtechnologies, I.M.T., Bucharest, Romania

## ABSTRACT

For the business models addressed in this study, we propose the implementation of distribution free learning framework concepts and paradigms.

The development of a machine learning process for a predictor identified with a high degree of precision is done with the help of a discriminatory paradigm.

A generative-type approach is developed, using the hypothesis that the underlying distribution used for the sampled and interpreted data has a parametric structure exploiting the so-called parametric density estimation. This choice has the advantage of avoiding learning processes for the distributions underlying the business models, resulting in rigorous predictions.

For the economic models, we consider that the VANIK principle has a relevant degree of efficiency, using a well-defined amount of information.

The originality and solutions proposed in this work come from the idea that in order to manage economic organizations, we must turn to innovative technological concepts and paradigms, such as machine and deep Learning as part of Artificial Intelligence. Therefore, economic activities will have both a controlled degree of uncertainty and a high degree of operational-strategic performance.

**KEYWORDS**: business models, learning through free distribution, generative models

# DOI: 10.24818/IMC/2021/01.14

# **1. INTRODUCTION**

The field of Machine Learning is an integral part of Artificial Intelligence, which has found a wide applicability in the problems faced by economic entities within the modern and globalized ecosystem (Murphy, 2012).

The purpose of our scientific research is to introduce and analyze the concept of "Generative Models", a concept specific to Machine Learning, the paradigms offered, at the same time we offer an extensive and in-depth theoretical presentation of their application for business models, the mathematical foundation that subsequently enables the development of strategic operational algorithms.

A decisive role regarding the economic performance and the generated added value, in the case of business models implemented for modern economic organizations, is held by the concepts of "underlying distribution", and the optimal, Bayes-type category classifier is identified within the paradigms used.

Addressing economic processes with the help of underlying distributions has a much higher degree of complexity than the use of precision predictors. For certain business models which are

<sup>\*</sup>Coresponding author. E-mail address:<u>marian.ionescu@man.ase.ro</u>

apparently difficult to approach, generational mathematical models are an efficient solution with a high degree of accuracy. Using intelligent solutions (specific to digital approaches) for specific economic challenges, mathematical estimation of the used model's parameters offers a simpler path for the decision-maker than the approach with the help of discriminatory predictors.

The current practice of managerial decision, shows us that a specific task, with a high degree of precision, is not always identifiable, thus, the data used in systemic processes are modeled to generate predictions, at a well-defined time, without initializing a specific predictor, or due to distortions generated by the interpretability of the data.

Our study and analysis are based, starting with the classical methods of estimating the parameters used for the data, found in literature under the name of "maximum probabilistic principle", in continuation of the developed reasonings, we introduce two assumptions that are involved in the learning processes.

We consider that a significant degree of relevance, for business models' applications, is also held by "E.M. algorithms" which calculate the maximum probability in the presence of variable sizes, we also consider particularly useful a description of the Bayesian analysis.

This approach is the result of an operational case study, found in the modern business ecosystem, for dynamic economic organizations, looking for appropriate answers to the challenges encountered, (Barber, 2012). We assume that we are studying an economic medical research organization, which manufactures a vaccine to combat an ongoing pandemic, a global challenge, for which concentrated and focused efforts, locally, zonally, globally, must offer solutions with a maximum degree of efficiency, possibly optimal. In this context, it is desirable to estimate the probability of survival for individuals who have used this type of vaccine, a sampling process is developed for this, on a group of m people who have used this type of vaccine.

#### 2. FORECAST AND MAXIMAL PROBABILITY

We consider  $S = (x_1, ..., x_m)$  the "training (learning)" group, where for each*i*,  $x_i = 1$  if is a human individual who has successfully passed through the pandemic and  $x_i = 0$  in other cases. We generate an underlying distribution using a single parameter  $\theta \in [0, 1]$ , which explicitly indicates the probability of survival in the pandemic, this parameter  $\theta$  must be estimated with the help of the "training (learning) group" *S*, it is advisable to use an average representing those "1" in the group *S* as estimators, resulting:

$$\theta^{\wedge} = \frac{1}{m} \sum_{i=1}^{m} x_i \tag{1}$$

It is obvious that  $\mathbf{E}_{\mathbf{s}}[\theta^{\wedge}] = \theta, \theta^{\wedge}$  is a rigorously correct (unbiased) estimator of  $\theta$ , as long as  $\theta$  is the average of random binary variables within the set *m*, using "*Inequality Hoeffding*" with the probability of at least  $1 - \delta$  for the entire choice over *S*, it results:

We can perceive  $\theta^{\wedge}$  as being the "Maximum Likelihood Estimator", (Shalev-Shwartz, 2007) at the onset we define the probability for the "sample" for *S*, under the formalism:

 $\mathbf{P}[S = (x_1, ..., x_m)] = \prod_{i=1}^m \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_i x_i} (1 - \theta)^{\sum_i (1-x_i)}.$ which offers the possibility of defining the "likelihood log" on *S*, using the  $\theta$  parameters, as log of the expression:

$$L(S;\theta) = \log(\mathbf{P}[S = (x_1, \dots, x_m)] = \log(\theta) \sum_i x_i + \log(1-\theta) \sum_i (1-x_i)$$

The maximum likelihood estimator is the parameter that maximizes the induced probability for the business model transposed operationally, strategically:

$$\theta^{\wedge} \in \operatorname{argmax}_{\theta} L(S; \theta) \tag{3}$$

For the proposed case study, that of the economic medical research organization it is obvious that equation (1) is the maximum likelihood estimator, deriving  $L(S; \theta)$ , after  $\theta$  and equaling this mathematical formalism with 0, it results:

 $\frac{\sum_{i} x_{i}}{\theta} - \frac{\sum_{i} (1-x_{i})}{1-\theta} = 0$ , solving this type of equation for  $\theta$ , the estimator formalized in equation (1) results, perfectly adapted to the operationally and strategically implemented business model for our study (Shalev-Shwartz, 2007).

# 3. CONTINUOUS RANDOM VARIABLE AND MAXIMAL LIKELIHOOD ESTIMATION

In the modern business models' strategic implementation and operationalization processes, continuous random variables have a role and a value of highly defined generated economic performance.

Considering *X* a continuous random variable, for most values  $x \in \mathbf{R}$ , we have  $\mathbf{P}[X = x] = 0$  and the way of assessing the likelihood is obvious.

Business models specific to modern economic organizations induce the need to define the estimate as log of the probability density from X to x.

The "training (learning)" group,  $S = (x_1, ..., x_m)$  is staggered in full agreement with the distribution density,  $P_{\theta}$  the probability on S for a given  $\theta$  is defined, expressable by the formalism:

$$L(S;\theta) = \log(\prod_{i=1}^{m} P_{\theta}(x_i)) = \sum_{i=1}^{m} \log(P_{\theta}(x_i)).$$

Positioning ourselves in this thinking paradigm, the maximum likelihood estimator is generated by maximizing the expression  $L(S; \theta)$  after  $\theta$ , perfectly adaptable to almost all the business models addressed, exemplifying, considering a Gauss-type random variable, (Bishop, 2006) for each density function of X, it is set customizable with the help of  $\theta = (\mu, \sigma)$  defined under the formalism:

$$P_{\theta}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

The likelihood may be reformulated as follows:

$$L(S;\theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^{m} (x_i - \mu)^2 - m \log(\sigma \sqrt{2\pi}).$$

A procedure to identify a parameter  $\theta = (\mu, \sigma)$ , optimizing the analyzed process is initiated, the likelihood derivate is sampled,  $\mu$  and  $\sigma$  and compared to 0, from which two mathematical formalisms result, as follows:

$$\frac{d}{d\mu}L(S;\theta) = \frac{1}{\sigma^2} \sum_{i=1}^m (x_i - \mu) = 0$$

$$\frac{d}{d\sigma}L(S;\theta) = \frac{1}{\sigma^3} \sum_{i=1}^m (x_i - \mu)^2 - \frac{m}{\sigma} = 0$$

Solving the previous equations generates the estimated maximum likelihood:

$$\mu^{\hat{}} = \frac{1}{m} \sum_{i=1}^{m} x_i, \qquad \sigma^{\hat{}} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (x_i - \mu^{\hat{}})^2}$$

The estimated maximum likelihood is not always an objective estimator with a high degree of correctness, if for example  $\mu^{\uparrow}$  is not unbiased there is implicitly the possibility of demonstrating that the variant estimator  $\sigma^{\uparrow}$  of is biased.

For a better understanding of the mathematical formalisms used, in the method of estimating the calculations, there is the possibility of "simplifying the notations", as follows.

P[X = x] represents both the likelihood that X = x, in the case of random variables but also the density of distribution of x for all continuous variables.

#### 4. MINIMIZATION OF EMPIRIC RISK AND MAXIMAL PROBABILITY

The probability estimator offers similar characteristics to the "Empirical Risk Minimisation Principle (E.R.M.)", which is the subject of extensive scientific studies, (Devroye et al, 1995).

The E.R.M principle, applied to a group H, for which we select a learning (training) group, under this hypothesis  $h \in H$ , this minimizing empirical risk.

The maximum probabilistic estimator is recognizable in E.R.M. for a particular loss function, for a given parameter  $\theta$  and an observation x, we define the loss of  $\theta$  on x as being represented by the mathematical formalism, (Shalev-Shwartz, 2007).:

 $l(\theta, x) = -\log(P_{\theta}[x]).$ 

(4)

 $l(\theta, x)$ , it in practice the negation of the probability on the observation made x, prca data are distributed in relation (according to)  $P_{\theta}$ .

The loss function (its highlighting) is known in specialized scientific literature as the "log-loss" function.

The maximum probability principle is equivalent to initiating a process of minimizing empirical risk, but the loss function paradigm must be complied with at the same time, and is expressable as follows:

 $argmin_{\theta} \sum_{i=1}^{m} (-\log(P_{\theta}[x_i])) = argmax_{\theta} \sum_{i=1}^{m} \log(P_{\theta}[x_i]).$ 

We assume that the data are distributed in agreement with the distribution P (not necessarily represented in parametric form), the real risk faced by the business model adopted for the economic organization positioned in a modern, dynamic and hypercompetitive ecosystem, according to parameter  $\theta$ , is represented by the mathematical formalism:

$$\boldsymbol{E}_{\boldsymbol{x}}[l(\theta, \boldsymbol{x})] = -\sum_{\boldsymbol{x}} P[\boldsymbol{x}] \log(P_{\theta}[\boldsymbol{x}]) = \left\{ \sum_{\boldsymbol{x}} P[\boldsymbol{x}] \log\left(\frac{P[\boldsymbol{x}]}{P_{\theta}[\boldsymbol{x}]}\right) \right\} (D_{RE}[P||P_{\theta}) + \{\sum_{\boldsymbol{x}} P[\boldsymbol{x}] \log(1/P[\boldsymbol{x}])\} (H(P)$$
(5)

in the symbolism used  $D_{RE}$  is known as being "relative entropy" and H as "entropy function", relative entropy is the quantifiable difference between two probabilities,

specific to the business model adopted, for discrete variables, always non negative this is equal to "0" only if the two distributions are practically the same, the risk assumed by the strategic operational transposition of the organizational business model is minimal when the mathematical relationality  $P_{\theta} = P$  exists.

The previously presented mathematical expression, relationality (5), shows how the operational paradigm of the implemented business model, the generative working hypothesis, directly affects the estimate of density, even in limiting infinite data (Murphy, 2012).

In conclusion, if for the work distribution assumed for the economic organization studied, a certain parametric form, the suboptimality is not identified and possibly the appearance of a business model inferior in terms of mathematical substantiation are perfectly measurable using the relative divergence of entropy evaluated during systemic processes.

# 5. EXTRAPOLATING THE PARADIGM ASSOCIATED TO BUSINESS MODELS

In the case of operational-strategic business models, we have the implicit question of how correct is the choice of the maximum probability estimator in the analysis and modeling of systemic processes encountered in managerial practice, in order to identify a correct response for this type of challenge, it is necessary to define how the quality of an approximation solution for the problem of estimating the organizational economic performance is evaluated.

There are different approaches for losses generated by a particular type of business model implemented organizationally, distinct from the automatic learning systemic processes, generative learning clearly defines conceptually the induced loss.

In some applications, for economic organizations, positioned within modern business ecosystems it is relatively easy to argue that the principle of maximum probability guarantees a low real risk, and for estimating the average of a Gaussian-type variant variable, (Shalev-Shwartz, 2007), we assume that the maximum probabilistic estimator allows the average formalism, as follows:  $\mu^{\wedge} = \frac{1}{m} \sum_{i} x_{i}$ .

If we consider 
$$\mu^*$$
 as the optimal parameter, then,  $E_{x \sim N(\mu^*, 1)}[l(\mu^{\wedge}, x) - l(\mu^*, x)] = E_{x \sim N(\mu^*, 1)} \log\left(\frac{P_{\mu^*}[x]}{P_{\mu^{\wedge}}[x]}\right) = E_{x \sim N(\mu^*, 1)}(-\frac{1}{2}(x-\mu^*)^2 + \frac{1}{2}(x-\mu^{\wedge})^2) = \frac{\mu^{\wedge 2}}{2} - \frac{(\mu^*)^2}{2} + (\mu^* - \mu^{\wedge})E_{x \sim N(\mu^*, 1)} = \frac{\mu^{\wedge 2}}{2} - \frac{(\mu^*)^2}{2} + (\mu^* - \mu^{\wedge})\mu^* = \frac{1}{2}(\mu^{\wedge} - \mu^*)^2.$  (6)

It should be mentioned that in the symbolism used  $\mu^{\wedge}$  is the average of *m* Gauss-type variables, so for strategically and operationally transposed business models, we identify a normal distribution with the most unfavorable mean  $\mu^*$  and variant  $\frac{\sigma^*}{m}$ , we identify the possibility that the limits (margins, boundaries) of the form used are, with the probability of at least  $(1 - \delta)$ , there is  $|\mu^{\wedge} - \mu^*| \leq \epsilon$ , where  $\epsilon$  depends directly on  $\frac{\sigma^*}{m}$  and  $\delta$ .

There are operational situations in which significant exceeding of the maximum probability estimator is found, in this we can exemplify the use of a Bernoulli-type random variable, X, and the mathematical relationality  $P[X = 1] = \theta^*$ .

In probability theory, concentration inequalities offer limits on how a random variable deviates from a certain value (usually its expected value).

The law of large numbers of classical probability theory states that sums of independent random variables are, under very mild conditions, close to their expectations with a high probability. Such amounts are the most basic examples of random variables concentrated around their mean. Recent results show that such behavior is divided by other functions of independent random

variables. Concentration inequalities can be sorted according to how much information about the random variable is needed to use the concentration inequality.

Using the Hoeffding inequality is easy to obtain a guarantee on  $|\theta^* - \theta^{\wedge}|$  which is maintained with a high subsequent probability, so if our goal is to obtain a low expected value for the log loss function defined by formalism (5), assuming that  $\theta^*$  tends to zero, but is greater than it, the probability that no element of the sample size *m* to be 1 is  $(1 - \theta^*)^m$ , this is higher than  $e^{-2\theta^*m}$ .

We notice that whenever  $m \leq \frac{\log(2)}{2\theta^*}$ , the probability that the samples are totally equal to zero is at least 50%, thus the rule in which the maximum probability is set with  $\theta^{\wedge} = 0$  is established, the mathematical formalism that represents the previous observation is:

$$\boldsymbol{E}_{\boldsymbol{x}\sim\theta^*}\left[l(\theta^{\wedge},\boldsymbol{x})\right] = \theta^*l(\theta^{\wedge},1) + (1-\theta^*)l(\theta^{\wedge},0) = \theta^*\log\left(\frac{1}{0}\right) = \infty.$$

The previous examples are those that determine our maximum attention and mathematical prudence (argumentation and justification only by calculation) in applying the paradigm and the maximum probabilistic principle, (Koller & Friedman, 2009).

In particular cases identifiable in the modern business ecosystem, we recommend accessing the totality of software tools, developed with the help of the features offered by Artificial Intelligence vectors, such as Machine & Deep Learning (Murphy, 2012).

## 6. CLASSICAL APPROACH TO BUSINESS MODELS

The automatic learning processes, operationally-strategically transposed for the business models implemented organizationally are addressed in the generative, estimative supposition of the parameters used, with the help of the "naive Bayes-type" analysis classifications.

In strategic forecasts, the prediction is made using a label  $y \in \{0, 1\}$  based on a vector of characteristics, of the form,  $x = (x_1, ..., x_d)$ , in which each  $x_i$  is assumed to be positioned in the group  $\{0, 1\}$ .

The mathematical formalism by which the optimal Bayes-type classifier is evaluated is presented below, as follows:

$$h_{Bayes}(\mathbf{x}) = argmax_{y \in [0,1]} P[Y = y | X = \mathbf{x}]$$

In order to describe the probabilistic function  $P[Y = y|X = \mathbb{Z}x]$  in detail, the parameters  $2^d$  are necessary, each corresponding to P[Y = 1|X = x] for fixed values of  $x \in \{0, 1\}^d$ , this approach that the number of business models (examples) must have an exponential increase, depending on the number of characteristics identified, within the implementation and analysis processes. In the "naive Bayes-type analysis" approach of business models, (Barber, 2012) as for a given

In the "naive Bayes-type analysis" approach of business models, (Barber, 2012) as for a given ticket, the analyzed characteristics are independent from each other, from where it results that:

$$P[X = \mathbf{x}|Y = \mathbf{y}] = \prod_{i=1}^{d} P\{X_i = x_i | Y = y\}.$$

With this hypothesis and using the "Bayes Rule", perhaps the most important rule in "Data Science", mathematically expresses the updating of a supposition, based on certain certainties, evidence, so it provides an exhaustive description of the learning process, so the act of learning, thus the "optimal Bayes classifier" is expressed with the help of the formalism:

 $h_{Bayes}(\mathbf{x}) = argmax_{y \in [0,1]} P[Y = y | X = \mathbf{x}] = argmax_{y \in \{0,1\}} P[Y = y] P[X = \mathbf{x} | Y = y] = argmax_{y \in \{0,1\}} P[Y = y] \prod_{i=1}^{d} P[X_i = x_i | Y = y]$ (7)

For the business model studied, the number of useful parameters that are estimated is the dimensionality 2d + 1, the generative assumption in which we are positioned generates a significant reduction in the number of parameters to be learned, thus using the paradigm of the maximum probabilistic principle in the estimation of these parameters, is generated, which in the scientific literature is known as the "Native Bayes Classification", (Barber, 2012).

#### 7. LINE DISCRIMINATION ANALYSIS

For the study of business models addressing the problem with the help of "Linear Discrimination Analysis (L.D.A.)" is a very good argument about how generative hypotheses simplify specific learning processes, as in the Bayesian classification, the phenomenon of prediction is taken into

account using a label  $y \in \{0, 1\}$  as the basis of the vector with the characteristics  $\mathbf{x} = (x_1, \dots, x_d)$ .

The first generative hypothesis is that  $P[Y = 1] = P[Y = 0 = \frac{1}{2}]$ , the second shows the conditional probability of X for a given Y is a Gauss distribution, and finally, the matrix covariance of the Gauss distribution is the same for two values of the labels used.

Considering  $\mu_0, \mu_1 \in \mathbb{R}^d$  and  $\Sigma$  being the covariance matrix, then the density distribution is expressed by formalism:

$$P[X = \mathbf{x}|Y = y] = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_y)^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_y)).$$

Using the Bayes Rule, the following formalism results:

 $h_{Bayes}(\mathbf{x}) = argmax_{y \in \{0,1\}} P[Y = y] P[X = \mathbf{x}|Y = y]$ , this is the prediction for  $h_{Bayes}(\mathbf{x}) = 1$ , if  $\log \left( \frac{P[Y=1]P[X = \mathbf{x}|Y = 1]}{P[Y=0]P[X = \mathbf{x}|Y = 0]} \right) > 0$ , it is found in the literature under the name "log-probabilistic ratio", in this case, applied on a chosen business model, the log-probabilistic ratio becomes,  $\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_0)^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_0) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_1)$ . It follows, in the mathematical formalism that:

$$\boldsymbol{w} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} \text{ and } \boldsymbol{b} = \frac{1}{2} (\boldsymbol{\mu}_0^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1)$$
 (8)

The optimal Bayes classifier is a linear classifier, the previous derivation generates practical, the generative hypothesis sought, identifying the classification parameters,  $\mu_0$ ,  $\mu_1$ ,  $\Sigma$ , the values are calculable by means of formalism (8).

## 8. E.M ALGORITHM AND LATENT VARIABLES

In the operational-strategic transposition of generative concepts and paradigms within business models, it is hypothesized that all data is generated by sampling from a specific parametric distribution, positioned over the identified space  $\chi$ , sometimes the convention of representation of the distribution used with the help of "latent random variables" is made, an example commonly found in economic systemic applications is that of the mix of k Gauss distributions, (Bishop, 2006).

For  $\chi = \mathbf{R}^d$  we assume that each  $\mathbb{Z}$  is generated in two stages, as follows:

- We choose a random number from the group  $\{1, ..., k\}$ , either Y the random variable (i) corresponding to this choice, we note  $P[Y = y] = c_y$ .
- We choose **x** from the value base *Y* in agreement with the Gauss distribution: (ii)

$$P[X = \mathbf{x}|Y = y] = P[X = \mathbf{x}|Y = y] = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_y)^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_y)).$$
(9)  
density of the business model implemented is:

The density of the business model implemented is:

$$P[X = \mathbf{x}] = \sum_{y=1}^{k} P[Y = y] P[X = \mathbf{x}|Y = y] = \sum_{y=1}^{k} c_y \frac{1}{(2\pi)^{d/2} |\mathcal{L}|^{1/2}} \exp(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_y)^T \mathcal{L}^{-1} (\mathbf{x} - \boldsymbol{\mu}_y)).$$

In the data used for the chosen business model, there is no possibility of identifying variable Y because this is a "hidden" variable, this type of variable is still introduced, so the probability X is well structured and described using a simple parameter.

Generalizing, either  $\theta$  the group of parameters belonging to the common distribution of X and Y, then, the log-likelihood of a 🛛 type observation, admits the following mathematical formalism: 1

$$og(P_{\theta}[X=x]) = log(\sum_{y=1}^{k} P_{\theta}[X=x, Y=y]).$$

Considering a sample, of the type  $S = (x_1, ..., x_m)$  is is desired to identify a value  $\theta$  that maximizes the log-likelihood of S:

$$L(\theta) = \log \prod_{i=1}^{m} P_{\theta}[X = x_i] = \sum_{i=1}^{m} \log(\sum_{y=1}^{k} P_{\theta}[X = x_i, Y = y])$$

The maximum probability estimator is practically the solution to maximizing the problem:

 $argmax_{\theta}L(\theta) = argmax_{\theta}\sum_{i=1}^{m} \log(\sum_{y=1}^{k} P_{\theta}[X = x_i, Y = y]).$ 

In the operationalization of several business models for economic organizations, the sum identifiable in the previous mathematical formulation, under "log" makes it so that the transposition within a dedicated software, in terms of optimization, raises a high level of calculation, (MacKay, 2003).

The most efficient and reliable Expectation-Maximization algorithm for economic entities, positioned in an ultra dynamic and competitive business ecosystem is the one due to "Dempster, Laird, Rubin", this is an iterative procedure to search for a local maximum  $L(\theta)$ , as long as "E.M." does not guarantee finding a global maximum, practical operational solutions, specific and particularly useful to each individual business model approached, are still generated, (MacKay, 2003).

#### 9. ALTERNATIVE TO THE MAXIMIZATION PROCESS E.M. ALGORITHM

Analyzing the E. M. algorithm it is easily observable that it is an alternative to the maximization algorithm, specific to management decision processes is "Objective Function":

$$G(Q,\theta) = F(Q,\theta) - \sum_{i=1}^{m} \sum_{y=1}^{k} Q_{i,y} \log(Q_{i,y}).$$

In the case of the implemented business models, (Shalev-Shwartz, 2007) the second term represents the sum "*entropies of rows*" of Q, defined as follows:  $\mathbf{Q} = \{Q \in [0, 1]^{m,k} : \forall i, \sum_{y=1}^{k} Q_{i,y} = 1\}.$ 

This is the multitude of matrices with rows (lines) defined probabilistically on [k].

As an immediate involvement, we affirm that in the analysis paradigm of the business model in which we position ourselves, the procedure E.M., never decreases the log-likelihood, for the totality t there is the mathematical relationality:

 $L(\theta^{(t+1)}) \ge L(\theta^{(t)}).$ 

## 10. E.M. FOR THE GAUSSIAN MIXED. SOFT K MEANS

A significant case to be analyzed and discussed is that of the mix of k variables *Gauss* in which  $\theta$  is a triplet of the form  $(c, \{\mu_1, \dots, \mu_k\}, \{\sum_1, \dots, \sum_k\})$  where  $P_{\theta}[Y = y] = c_y$  and  $P_{\theta}[X = x|Y = y]$  is presented in relationality (9), (Koller&Friedman, 2009).

By applying the E.M algorithm, we identify two steps to follow, as follows:

- (i) Expectation: for  $\forall i \in [m], y \in [k]$ , results ,  $P_{\theta}(t)[Y = y|X = x_i] = \frac{1}{Z_i} c_y^{(t)} \exp(-\frac{1}{2} ||x_i \mu_y^{(t)}||^2)$  (10)
- (ii) Maximization:  $\theta^{t+1}$  must be chosen to maximize  $\theta^{(t+1)} = argmax_{\theta}F(Q^{(t+1)},\theta)$ , determined in the calculation using the E.M. algorithm, of latent variables, starting from

(iii) 
$$\sum_{i=1}^{m} \sum_{y=1}^{k} P_{\theta}(t) [Y = y | X = x_i] (\log(c_y) - \frac{1}{2} | |x_i - \mu_y||^2)$$
(11)  
Rearranging the terms results in,  $\mu_y = \sum_{i=1}^{m} P_{\theta}(t) [Y = y | X = x_i] x_i$ 

Where  $\mu_y$  is the weighted average of  $x_i$  where the weights are in agreement with the probabilities calculated at the previous Wait Step, in order to identify the optimal c, we continue the developed reasoning, ensuring that c presents the characteristics and therefore is a vector structure, as follows:

$$c_{y} = \frac{\sum_{i=1}^{m} P_{\theta}(t)[Y=y|X=x_{i}]}{\sum_{y'=1}^{k} \sum_{i=1}^{m} P_{\theta}(t)[Y=y'|X=x_{i}]}$$
(12)

It is interesting to make a comparison between algorithms, so in relation to "K – *means algorithm*", within it first is initiated an assignment process for each example according to the cluster, compared to the distance  $||x_i - \mu_y||$ , the probability that each example (case study) belongs to each cluster is determined within the systemic operationalization with the help of E.M.

The centers addressed are updated on the basis of a weighted sum at the level of the entire sample studied, which is why the E.M. algorithm for "k-means" is referred to in the literature as "soft k-means" as well.

# **11. CONCLUSIONS**

The usefulness and efficiency of implementing the solutions offered via Artificial Intelligence, through the sub-domains of systemic activity, Deep Learning and Machine Learning find their confirmation for economic organizations in generating added value and superior managerial performance.

In the case of generative approach through Machine Learning, the distribution of data sampled and used in specific economic processes, in specific cases, in the processes of estimating the parametric density, is intended, (Shalev-Shwartz, 2007), (Bishop, 2006).

The reasonings presented in this work were specific to the actors of the modern economic ecosystem, the underlying distribution of data is presented in a parametric mathematical formalism specific to each chosen business model, the aim pursued in the case of this study is to estimate the value of these parameters.

Some mathematical and economic principles have been presented and used in the processes of estimating parameters, including the maximum probabilistic principle, the Bayesian type estimation.

Relevant importance was given to the specific algorithms used in the operational-strategic systemic implementation under different working hypotheses, underlying to distribution phenomena such as Naïve Bayes, L.D.A., E.M., (Koller & Friedman, 2009)

Finally, this scientific communication, without proposing an exhaustive approach to economic phenomena that have a relevant degree of uncertainty, offered novel solutions, well grounded mathematically, opening the way for further practical and theoretical debates and development.

# REFERENCES

Barber, D.(2012). Bayesian reasoning and machine learning, Cambridge University Press.

Bishop, C.M. (2006). Pattern recognition and machine learning, Vol.1, Springer: New York.

Devroye, L., Gyorfi, L. & Lugosi, G.(1995). A probabilistic theory of pattern recognition, Springer.

- Koller, D. & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*, MIT Press.
- MacKay, D.J. (2003). *Information theory, inference and learning algorithms,* Cambridge University Press.

Murphy, K.P.(2012). *Machine learning: a probabilistic perspective*, The MIT Press.

Shalev-Shwartz, S.(2007). *Online Learning: Theory, Algorithms, and Applications*, PhD thesis, The Hebrew University.