Mathematical and Quantitative Methods

Using SVM for Classification

Dumitru Ciobanu[1]

Abstract: Support Vector Machines (SVMs)have found many applications in various fields. They have been introduced for classification problems and extended to regression. In this paperI review the utilization of SVM for classification problems and exemplify this with application on IRIS datasets. I used the Matlab programming language to implement linear and nonlinear classificators and apply this on the dataset.

Keywords: Support Vector Machines; optimal separating hyperplane; generalized separating hyperplane; nonlinear classifications; Iris dataset

JEL Classification: C02; C38; C45

1 Introduction

Support vector machines have a relatively short history being recently introduced, in the early 1990s. However, they are based on decades of research in computational learning theory done by Russian mathematicians Vladimir Vapnik and Alexey Chervonenkis. This theory, presented in the book of Vapnik from 1982 *Estimation of Dependences Based on Empirical Data*, was called Vapnik-Chervonenkis theory or simply VC theory (Vapnik, 2006). This book describes the implementation of support vector machines for linearly separable data (Cortes & Vapnik, 1995). A number of important extensions were made to the SVM. In 1992, Boser, Guyon and Vapnik proposed the use of kernel trick of Aizerman's to classify data separable using polynomial functions or radial basis functions. In 1995, Cortes and Vapnik extended the theory so that it can be applied for the training data inseparable, using a cost function. Later, in 1996 (Drucker, 1996), was developed a method for regression based on support vector.

It should be noted that there are many different algorithms for SVMs like SVM Lagrangian (LSVM), Lagrangian finite Newton SVM (NLSVM) or finite Newton SVM (NSVM), a comparison between different methods is shown in (Shu-Xia Lu, 2004).

[1] PhD Student, University of Craiova, Romania, Address: A. I. Cuza, no. 13, Craiova, 200585, Romania, Tel.: +40 251/411593, Fax: +40 251/411317, Corresponding author: ciobanubebedumitru@yahoo.com.

A Support Vector Machine (SVM) is a machine learning that can be used in classification problems (Cortes & Vapnik, 1995) and regression problems (Smola, 1996).

In order to perform classification, SVMs seek an optimal hyperplane that separates data into two classes. In Figure 1are presented some possibilities of linear separation of two sets of elements.

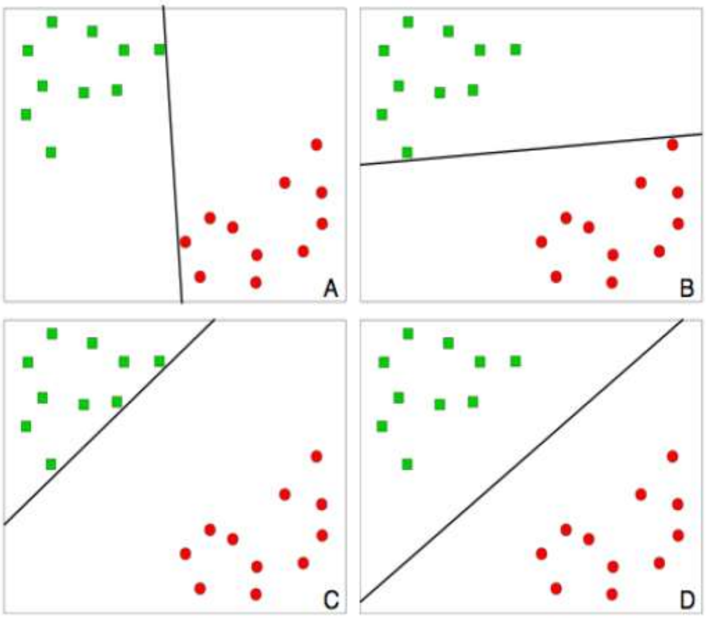

Figure 1. Different variants of linear separation of two sets

(Guggenberger, 2008)

Support vector machine are also called classifiers with maximum edge. This means that the resulted hyperplane maximizes the distance between the closest vectors from different classes taking into account the fact that a greater margin provides increased SVM generalization capability.

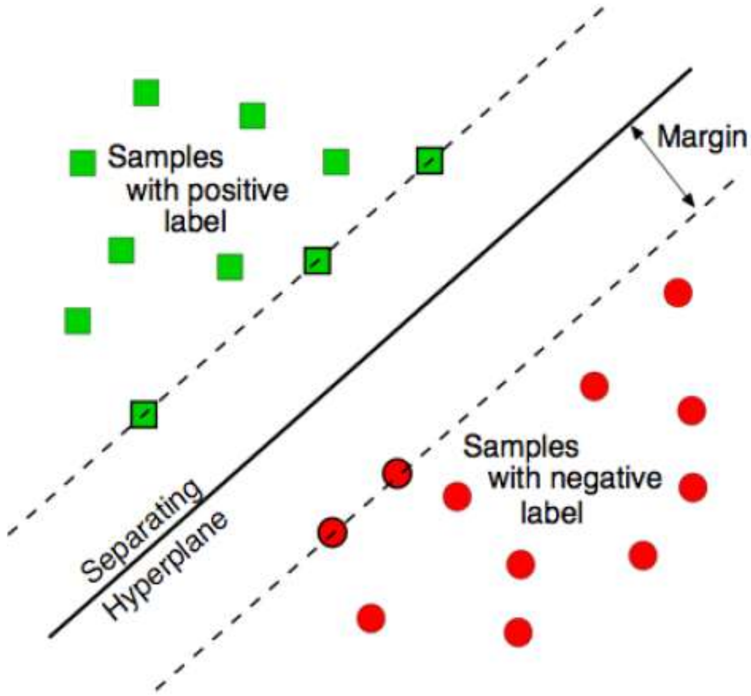

Figure 2. Optimal separating hyperplane. The vectors on dotted lines are support vectors

(Guggenberger, 2008)

The elements closest to the optimal separating hyperplane are called support vectors and only they are considered by the SVMs for the classification task. All other vectors are ignored.

2. Optimal Separating Hyperplane

The basic problem that SVM learns and solves is that of classification in two categories of a data set.

Classification problem implies a set of observations represented as pairs (x_i, y_i), i = 1, …, r, where $x_i \in \mathbb{R}^n$ and $y_i \in \{-1, 1\}$. Each observation contains an n-dimensional vector and an associated class. The aim is to determine the optimal separation hyperplane, that is the hypersurface (n-1)-dimensional, which best separates the two classes Figure 3.

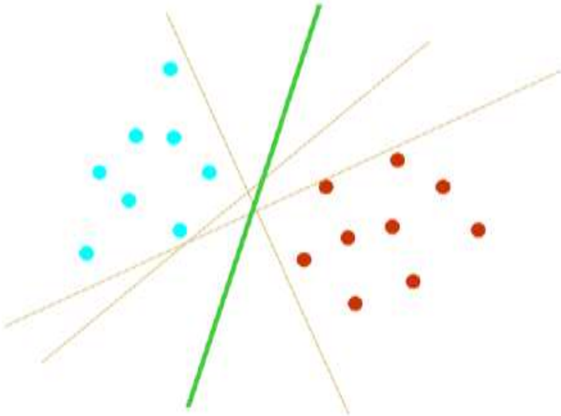

Figure 3. Optimal Separating Hyperplane

(Gunn, 1998)

The simplest situation is that there exist a hyperplane defined by a normal vector *w*, which separates the classes,

$$\langle w, x \rangle + b = 0 \tag{1}$$

Because this hyperplane is invariant to scalar multiplication, we can choose *w* and *b* so as to meet the requirement

$$\min_i \left| \langle w, x_i \rangle + b \right| = 1 \tag{2}$$

Constraint in equation (2) tells us that the norm of weight vector *w* must be equal to the inverse distance from the nearest point of the dataset to hyperplane.

Also, the equation (2) leads to a breakdown of points in two categories.

$$\langle w, x_i \rangle + b \geq 1 \tag{3}$$

$$\langle w, x_i \rangle + b \leq -1 \tag{4}$$

Assuming that the first category corresponds to points labeled 1 and the second category to points labeled -1, the two inequalities are rewritten as

$$y_i \left[\langle w, x_i \rangle + b \right] \geq 1,\ i = 1, \ldots, r\ . \tag{5}$$

$\langle w, x_i \rangle + b = 1$ and $\langle w, x_i \rangle + b = -1$ are two hyperplans parallel with separating hyperplane. This is represented in Fig. 2., where the separating hyperplane is represented by a solid line and those two parallel hyperplans by dotted lines. Dotted lines contain some of the training points. These points are called support vectors and completely determines the solution of classification problem. The distance between the dotted lines is called margin and is to be maximized.

The margin is $\rho(w,b)=\frac{2}{\|w\|}$ and the maximization of margin is equivalent with maximization of the function

$$L(w)=\frac{1}{2}\|w\|^2 \tag{6}$$

with constraints (5).

The solution to optimization problem (6) with constraints (5) is given by the saddle point of Lagrange functional (Minoux, 1986),

$$L(w,b,\alpha)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{r}\alpha_i\left(y_i\left[\left\langle w,x_i\right\rangle+b\right]-1\right) \tag{7}$$

were α is the vector of the Lagrange multipliers.

The Lagrangian must be minimized in rapport with *w*, *b* and maximized in function of $\alpha \geq 0$. Classic theory of Lagrange duality allow us to transform the primal problem (7) in the dual problem, which is easier to solve. The dual problem has the form,

$$\max_{\alpha} W(\alpha)=\max_{\alpha}\left(\min_{w,b} L(w,b,\alpha)\right) \tag{8}$$

That is

$$\max_{\alpha} W(\alpha)=\max_{\alpha}\left(-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}\alpha_i\alpha_j y_i y_j\left\langle x_i,x_j\right\rangle+\sum_{k=1}^{r}\alpha_k\right) \tag{9}$$

and the solution is given by

$$\alpha^*=\arg\min_{\alpha}\left(\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}\alpha_i\alpha_j y_i y_j\left\langle x_i,x_j\right\rangle-\sum_{k=1}^{r}\alpha_k\right) \tag{10}$$

with constraints

$$\alpha_i \geq 0,\ i = 1,\ldots,r$$
$$\sum_{i=1}^{r} \alpha_i y_i = 0 \tag{11}$$

Solving equation (10) with constraints (11) is determined Lagrange multipliers and optimal separating hyperplane, given by

$$w^* = \sum_{i=1}^{r} \alpha_i y_i x_i$$
$$b^* = -\frac{1}{2}\left\langle w^*, x_k + x_s \right\rangle, \tag{12}$$

where x_k and x_s are any of support vectors coming from the two classes, that satisfy relations $\alpha_k, \alpha_s > 0$ and $y_k = -1, y_s = 1$.

Then, the hard classifier (inflexible edges)

$$f(x) = \operatorname{sgn}\left(\left\langle w^*, x \right\rangle + b^*\right). \tag{13}$$

From Kuhn-Tucker conditions,

$$\alpha_i \left(y_i \left[\left\langle w, x_i \right\rangle + b \right] - 1 \right) = 0,\ i = 1,\ldots,r\ , \tag{14}$$

result that only points x_i that satisfy

$$y_i \left[\left\langle w, x_i \right\rangle + b \right] = 1, \tag{15}$$

will have nonzero Lagrange multipliers. These points are called support vectors (SV). If the data is linearly separable all support vectors will be on edge and their number can be very small. Consequently, the hyperplane is determined by a small subset of the training set. Eliminating from the training set points that are not support vectors and recalculate the optimal separating hyperplane will achieve the same result. Thus, support vector machines (SVM) are used to summarize information contained in the training data using support vector.

3. Generalised Optimal Separating Hyperplane

Most times the data provided for classification are not linearly separable. One way to perform classification in such cases is generalized optimal separating hyperplane. It separates linear data supporting classification errors. In Fig. 4. we have an intuitive graphical representation of generalized optimal separating hyperplane.

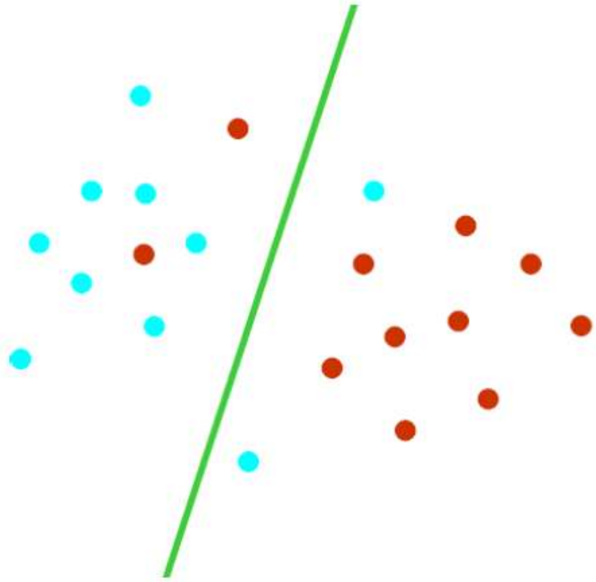

Figure 4. Generalized Optimal Separating Hyperplane

(Gunn, 1998)

Cortes and Vapnik introduced variables $\xi_i \geq 0$ that mesures the classification errors (Cortes & Vapnik, 1995).

In these conditions, the optimization problem will minimize classification errors. Constraints for the inseparable case will be of the form

$$y_i\left[\left\langle w, x_i \right\rangle + b\right] \geq 1 - \xi_i,\ i = 1, \ldots, r\ . \tag{16}$$

where $\xi_i \geq 0$.

Generalized optimal separating hyperplane is determined by the vector w that minimize the functional

$$L(w,\ \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{r}\xi_i \tag{17}$$

with constraints (16), where C is a given constant.

The solution of minimization of the functional (17) with constraints (16) is given by the saddle point of the following Lagrangian (Minoux, 1986),

$$L(w,b,\alpha,\xi,\beta)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{r}\xi_i-\sum_{i=1}^{r}\alpha_i\left(y_i\left[\langle w,x_i\rangle+b\right]-1+\xi_i\right)-\sum_{i=1}^{r}\beta_i\xi_i\,, \quad (18)$$

where α and β are the Lagrange multipliers. The Lagrangian must minimized about *w, b, x* and maximized about α, β. To solve this optimization problem is recalled, as in the classical case at the dual problem

$$\max_{\alpha,\beta} W(\alpha,\beta)=\max_{\alpha,\beta}\left(\min_{w,b,\xi} L(w,b,\alpha,\xi,\beta)\right) \quad (19)$$

Explicitly, the dual problem is written

$$\max_{\alpha} W(\alpha)=\max_{\alpha}\left(-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}\alpha_i\alpha_j y_i y_j\langle x_i,x_j\rangle+\sum_{k=1}^{r}\alpha_k\right) \quad (20)$$

and the solution is

$$\alpha^*=\arg\min_{\alpha}\left(\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}\alpha_i\alpha_j y_i y_j\langle x_i,x_j\rangle-\sum_{k=1}^{r}\alpha_k\right) \quad (21)$$

with constraints

$$\begin{gathered} 0\le\alpha_i\le C,\; i=1,\dots,r \\ \sum_{i=1}^{r}\alpha_i y_i=0 \end{gathered} \quad (22)$$

The solution of minimization problem in the case of linearly inseparable data is identical to those from data linearly separable case except the bounds of Lagrange multipliers. Yet, there was an additional problem, namely determining the coefficient *C*. This parameter offers new possibilities to control over the classifier. Blanz and collaborators have used the value C = 5 (Blanz et al, 1996), other researchers regard *C* as directly related to a regularization parameter (Smola & Scholkopf, 1998), but eventually *C* must be chosen so that to reflect the knowledge of noise from data (Gunn, 1998).

4. Generalization in Multidimensional Feature Space

Another approach to separate two classes is to transfer, using a nonlinear applications, the input space into a feature space with higher dimension in which data can be separated using optimal separating hyperplane Fig. 5.

The idea is based on the method introduced by Aizerman and colleagues (Aizerman, Braverman & Rozonoer, 1964) which eliminates problems arising from increasing the dimension (Bellman, 1961).

Nonlinear functions that can be used must meet certain conditions, known as Mercer conditions. Among the most used functions that satisfy these requirements we mention the polynomial, the base radial and sigmoidal functions.

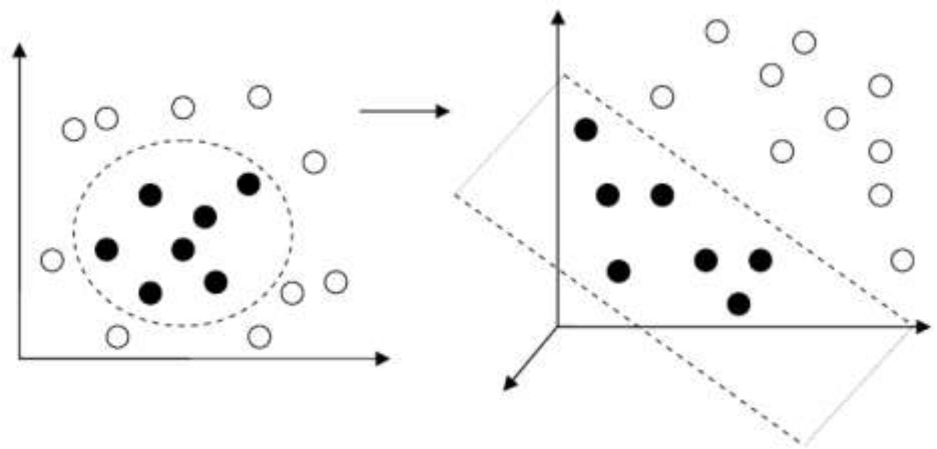

Figure 5. Using a higher dimension space for the linear separation of data

(Lovell & Walder, 2006)

The optimization problem in this case, can be written

$$\alpha^* = \arg\min_{\alpha}\left(\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{k=1}^{r}\alpha_k\right), \tag{23}$$

where $K(\cdot,\cdot)$ is the kernel function that performs nonlinear translation of input space to feature space and the constraints are the same as for generalized linear case

$$0 \le \alpha_i \le C,\ i = 1, ..., r$$
$$\sum_{i=1}^{r} \alpha_i y_i = 0. \tag{24}$$

It solves the optimization problem (23) with the restrictions (24) and determine the Lagrange multipliers. With this is build a hard classifier in feature space

$$f(x) = \operatorname{sgn}\left(\sum_{x_i \in SV} \alpha_i^* y_i K(x_i, x) + b^* \right), \tag{25}$$

where

$$\langle w^*, x \rangle = \sum_{x_i \in SV} \alpha_i^* y_i K(x_i, x)$$
$$b^* = -\frac{1}{2} \sum_{x_i \in SV} \alpha_i^* y_i \left[K(x_i, x_k) + K(x_i, x_s) \right] \tag{26}$$

with x_k and x_s any of the support vectors coming from the two classes.

5. Case Study Iris Dataset

For exemplification of SVM classification we use Iris data set (Fig. 6.). It consists of 150 observations, 50 Iris setosa, 50 Iris versicolor and Iris virginica 50, with 4 characteristics: length of sepals, sepals width, length of petals and petals width.

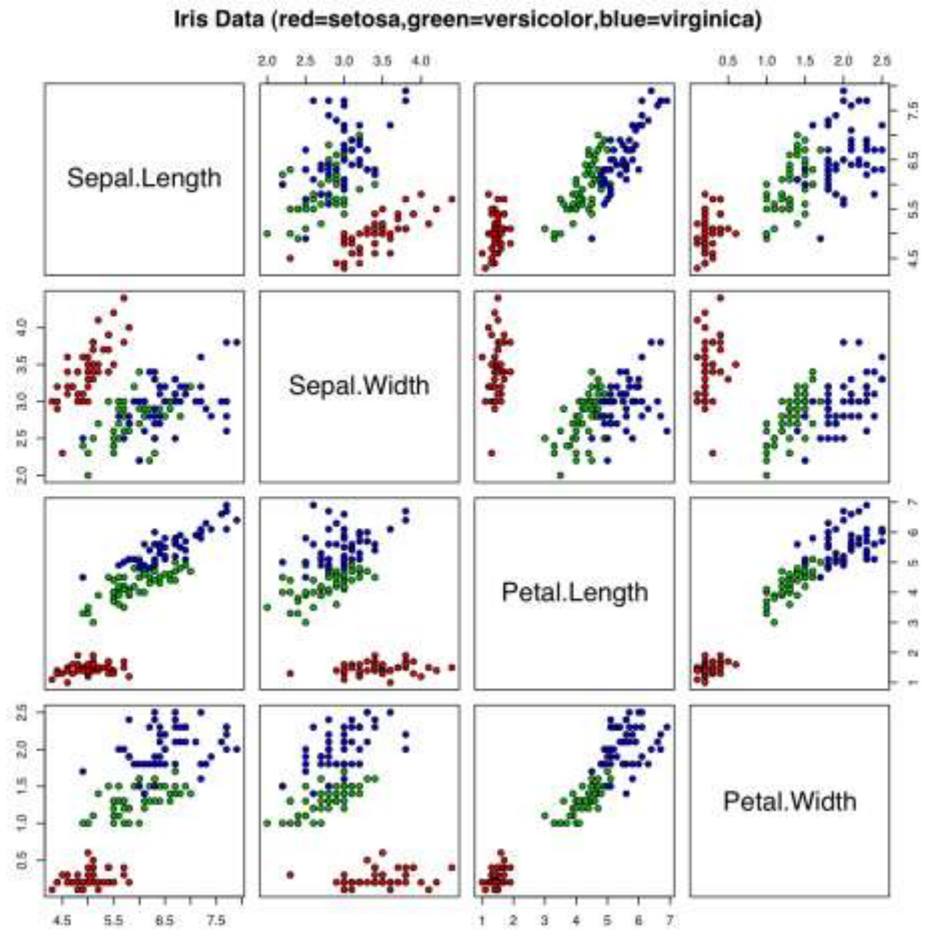

Figure 6. Representations of Iris data set based on pairs of two features
(http://en. wikipedia. org/wiki/File:Anderson%27s_Iris_data_set. png, accessed in 2012)

Iris data set has been extensively used for exemplification of classification and grouping methods because in binary representations have both linearly separable classes (iris setosa - iris versicolor and iris setosa - iris virginica) and classes that are not linearly separable (iris virginica - iris versicolor).

For exemplification of different methods of classification we use the graphic representation sepals length versus petals length.

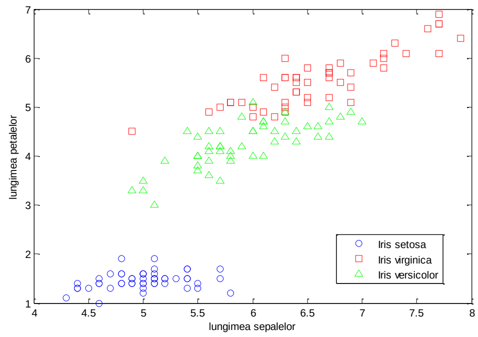

Figure 7. The representation sepals length versus petals length of iris dataset

For linear separation we use classes Iris setosa and Iris versicolor (Fig. 8.) And for nonlinear classification we exemplify using classes iris virginica and iris versicolor (Fig. 10.).

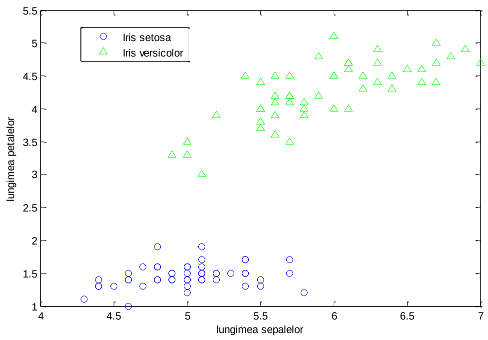

Figure 8. Iris versicolor and iris setosa according to the length of sepals and the length of petals

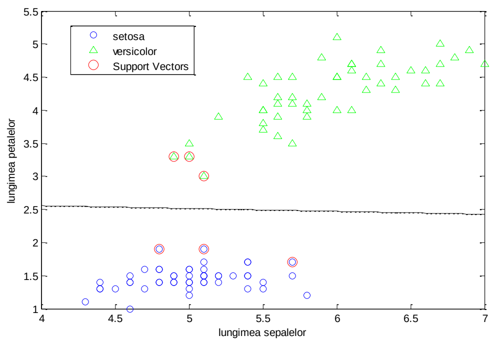

Figure 9. Linear separation of classes iris setosa and iris versicolor with highlighting of support vectors

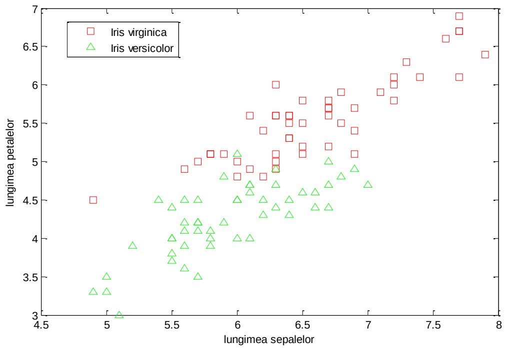

Figure 10. Iris versicolor and iris virginica according to the length of sepals and the length of petals

In this case we note that data are not linearly separable, so we use linear classifier with flexible edges and nonlinear classifiers.

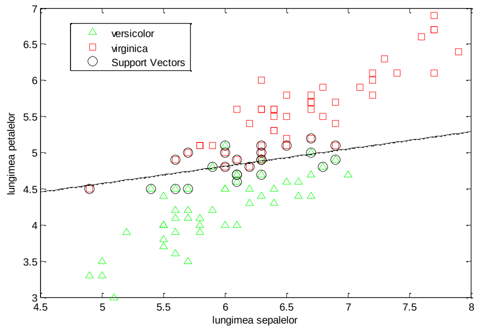

Figure 11. Linear separation with flexible edges of classes iris virginica and iris versicolor with highlighting of support vectors

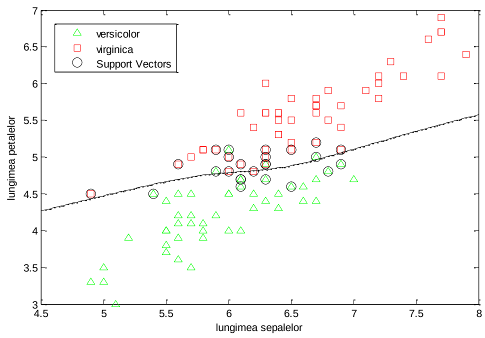

Figure 12. Nonlinear separation using a polynomial kernel of classes iris virginica and iris versicolor with highlighting of support vectors

Figure 12. represents a case of using a polynomial kernel of order 3 and Fig 13. a situation encountered for a kernel of type radial basis function.

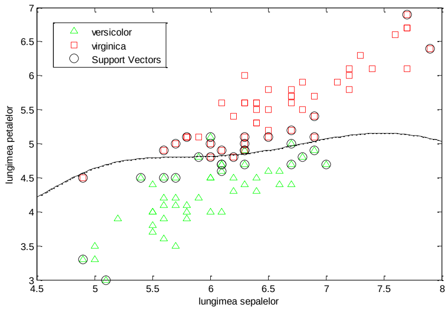

Figure 13. Nonlinear separation using a kernel of type radial basis function of classes iris virginica and iris versicolor with highlighting of support vectors

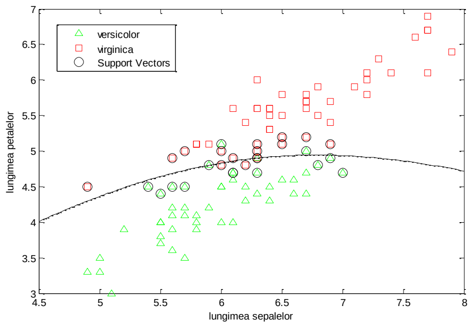

Figure 14. Nonlinear separation using a kernel of type quadratic function of classes iris virginica and iris versicolor with highlighting of support vectors

For a quadratic kernel In the case of a kernel of multilinear perceptron type.

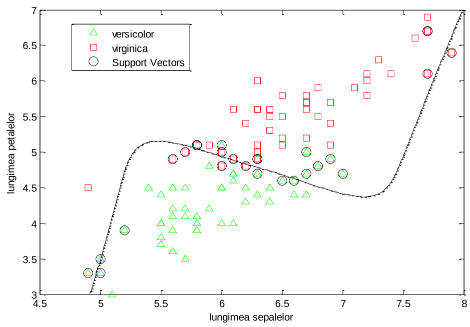

Figure 15. Nonlinear separation using a kernel of type multilinear perceptron of classes iris virginica and iris versicolor with highlighting of support vectors

5. Conclusion

SVM is one of the most promising algorithms in machine learning field and there are many examples in which SVMs are successfully used, for example, text classification, face recognition, character recognition (OCR - Optical Character Recognition), Bioinformatics. On these datasets SVMs apply very well and often exceeds the performance of other traditional techniques. Of course, this is not a magic solution as set forth in (Bennett & Campbell, 2000), there are still some open issues, such as incorporation of domain knowledge, a new model selection and interpretation of results produced by SVMs.

SVMs have been used in several real-world problems:

- classification of text (and hypertext);
- image classification;
- in bioinformatics (protein classification, classification of types of cancer);
- classification of music;
- handwritten character recognition.

In (Chen, Jeong & Hardie, 2008), the authors propose a method GARCH (Generalized AutoRegressive Conditional Heteroscedasticity) based on recurrent SVR whose performance exceeds other approaches such as moving average, recurrent neural networks and parameterized GARCH in terms of their ability to predict the financial market volatility. Important aspect that recommend the use of SVM we mention the absence of local minima, control solution capacity (Christiani & Shawe-Taylor, 2000) and the ability to effectively use multidimensional data (Cortes & Vapnik, 1995).

Strengths of SVM:

- Training is relatively easy to achieve;
- No local optimal, unlike neural networks;
- Suitable for multidimensional data relatively well;
- Non-traditional data such as strings and trees can be used as input to SVM, instead of feature vectors;
- The compromise between complexity and classification error can be controlled explicitly;
- By performing logistic regression (sigmoidal) with SVM on a set of output data, SVM can be interpreted in terms of probability.

Weaknesses of SVM:

- It needs a good choice for kernel function;
- Training takes a long time.

In graphic representations can see the small number of support vectors, basically those who are using the classifier. Due to the small number of support vector classification of new cases require scarce resources of time and computing power.

The best classification for linearly inseparable case, were obtained for polynomial and radial basis kernels which underlines once again the importance of a correct choice for the kernel function used.

6. References

Aizerman, M. A.; Braverman, E. M. & Rozonoer, L.I. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, Vol. 25, pp. 821–837.

Bellman, R. E. (1961). *Adaptive Control Processes*. Princeton, NJ: Princeton University Press.

Bennett, K. P.; Campbell, C. (2000). Support vector machines: hype or hallelujah? *SIGKDD Explorations Newsl*., Vol 2, No. 2, pp. 1–13.

Boser, B.; Guyon, I.; Vapnik, V. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, pp. 144–52.

Blanz, V.; Schölkopf, B.; Bülthoff, H.; Burges, C.; Vapnik, V. & Vetter, T. (1996). *Comparison of view-based object recognition algorithms using realistic 3D models*, In: C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff (eds.): *Artificial Neural Networks - ICANN'96*. Springer Lecture Notes in Computer Science Vol. 1112, Berlin, pp. 251-256.

Chen, S.; Jeong, K.; Härdle, W. (2008). *Support Vector Regression Based GARCH Model with Application to Forecasting Volatility of Financial Returns*, SFB 649 "Economic Risk", Humboldt-Universität zu Berlin, Berlin. Available online at http://edoc.hu-berlin.de/series/sfb-649-papers/2008-14/PDF/14.pdf.

Christiani, N. & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge: Cambridge University Press.

Cortes, C. & Vapnik, V. (1995). Support vector networks. *Machine Learning*, 20(3), pp. 273-297.

Drucker, H.; Burges, C.; Kaufman, L.; Smola, A. & Vapnik, V. (1996). *Support vector regression machine*. Advances in Neural Information Processing Systems, Cambridge: MIT Press 9(9): 155–61.

Guggenberger, A. (2008). *Another Introduction to Support Vector Machines*, Available online at http://mindthegap.googlecode.com/files/AnotherIntroductionSVM.pdf, accessed May 2012.

Gunn, S. R. (1998). *Support Vector Machines for Classification and Regression*, University of Southampton, Available online at http://www.svms.org/tutorials/Gunn1998.pdf, accessed April 2012.

Lovell, B. C.; Walder, C. J. (2006). Support Vector Machines for Business Applications. *Business Applications and Computational Intelligence*. Hershey, U.S.A: Idea Group, pp. 267-290.

Minoux, M. (1986). *Mathematical Programming: Theory and Algorithms*. John Wiley and Sons.

Smola, J. (1996). Regression estimation with support vector learning machine. *Master's thesis*. Munchen: Technische Universitat Munchen.

Smola, A.; Schölkopf, B. (1998). On a Kernel-based Method for Pattern Recognition, Regression, Approximation and Operator Inversion. *GMD Technical Report* No. 1064.

Shu-Xia Lu, X. -Z. W. (2004). A comparison among four SVM classification methods: Lsvm, nlsvm, ssvm and nsvm. *Proceedings of 2004 International Conference on Machine Learning and Cybernetics*, vol. 7, pp. 4277–4282, Shanghai, China.

Vapnik, V. (2006). *Empirical Inference Science*. Afterword in 1982 reprint of Estimation of Dependences Based on Empirical Data.

On the General Theory of Production Functions

Catalin Angelo Ioan[1], Gina Ioan[2]

Abstract: In this paper we will study from an axiomatic point of view the production functions. Also we will define the main indicators of a production function, extending the classical definitions to n inputs and introducing other new. We will modify the notion of global average productivity and replace it with more realistic indicators. On the other hand, the notion of global rate of substitution will be introduced to the analysis of n goods.

Keywords: production function; productivity; marginal

JEL Classification: D01

1. Introduction

In any economic activity, obtaining a result of this means, implicitly, the existence of any number of resources required for a good deployment of the production process. We will assume that resources are indefinitely divisible, which implies the possibility of using specific tools of mathematical analysis to onset specific phenomena.

We then define on $\mathbf{R}^n$ the production space for n fixed resources as $SP=\{(x_1,...,x_n) \mid x_i \geq 0,\ i=\overline{1,n}\}$ where $x \in SP$, $x=(x_1,...,x_n)$ is an ordered set of resources.

Because inside a production process, depending on the nature of applied technology, but also its specificity, not any amount of resources is possible, we will restrict production space to a subset $D_p \subset SP$ – called the domain of production.

In the context of the existence domain of production, we will put the issue of determining its results (output) depending on the resources (inputs) of D_p.

We will call a production function an application:

[1] Associate Professor, PhD, Faculty of Economic Sciences, Danubius University of Galati, Romania, Address: 3 Galati Blvd, Galati, Romania, Tel.: +40372 361 102, fax: +40372 361 290, Corresponding author: catalin_angelo_ioan@univ-danubius.ro.

[2] Assistant Professor, PhD in progress, Faculty of Economic Sciences, Danubius University of Galati, Romania, Address: 3 Galati Blvd, Galati, Romania, Tel.: +40372 361 102, fax: +40372 361 290, e-mail: gina_ioan@univ-danubius.ro.

$$Q:D_p \to \mathbf{R}_+,\ (x_1,...,x_n) \to Q(x_1,...,x_n) \in \mathbf{R}_+ \forall (x_1,...,x_n) \in D_p$$

For an efficient and complex mathematical analysis of production functions, we will require a series of axioms (*but not all essential*) both its definition domain and its scope.

FP1. The production domain is convex.

D_p's convexity only means that if $x=(x_1,..., x_n)$, $y=(y_1,...,y_n) \in D_p$ then $\forall \lambda \in [0, 1]$ follows $\lambda x+(1-\lambda)y=(\lambda x_1+(1-\lambda)y_1,...,\lambda x_n+(1-\lambda)y_n) \in D_p$.

FP2. If all resources are zero then the output is zero.

The FP2 axiom states that $Q(0,..., 0)=0$.

FP3. The production function is continuous.

The continuity, purely mathematical, means that for any fixed point $(\overline{x}_1,...,\overline{x}_n)$ of the domain of production D_p and a range of inputs $(y_k)_{k\geq 1}$, $y_k=\left(y_1^k,...,y_n^k\right)$ that converges to $(\overline{x}_1,...,\overline{x}_n)$, the production $Q\left(y_1^k,...,y_n^k\right)$ converges to $Q(\overline{x}_1,...,\overline{x}_n)$.

An axiom, not necessarily required, but very useful for obtaining significant results is:

FP4. The production function admits partial derivatives of order 2 and they are continuous (the function is of class C^2 on D_p).

Before commenting on this axiom, let note that all elementary functions are of class C^∞ on their domain of definition. Therefore, the class membership C^∞ is no way restrictive. It should also be noted that a function of class C^k, $k\geq 0$ is continuous, therefore the axiom FP4 implies automatically FP3.

FP5. The production function is monotonically increasing in each variable.

The FP5 axiom says that, in caeteris paribus hypothesis for any $i=\overline{1,n}$, if $x_i\geq y_i$ then: $Q(\overline{x}_1,...\overline{x}_{i-1},x_i,\overline{x}_{i+1},\overline{x}_n) \geq Q(\overline{x}_1,...\overline{x}_{i-1},y_i,\overline{x}_{i+1},\overline{x}_n) \forall\, \overline{x}_k \geq 0$, $k=\overline{1,n}$, $k\neq i$ such that $(\overline{x}_1,...\overline{x}_{i-1},x_i,\overline{x}_{i+1},\overline{x}_n)$, $(\overline{x}_1,...\overline{x}_{i-1},y_i,\overline{x}_{i+1},\overline{x}_n) \in D_p$.

If the function Q is at least class C^1, the increasing monotony is equivalent to: $\frac{\partial Q}{\partial x_i} \geq 0$, $i=\overline{1,n}$.

From the axiom FP5 follows the global increasing related to the inequality relationship of $\mathbf{R}^n$:

FP5'. The production function is monotonically increasing with respect to the relationship of inequality of $\mathbf{R}^n$.

Indeed, if $x_1 \geq y_1, ..., x_n \geq y_n$ then:

$$Q(x_1, x_2,...,x_n) \geq Q(y_1, x_2,...,x_n) \geq Q(y_1, y_2,...,x_n) \geq ... \geq Q(y_1, y_2,...,y_n)$$

A condition often mentioned in the definition of production function is:

FP6. The production function is quasi-concave.

The quasi-concavity of a function means that:

$$Q(\lambda x+(1-\lambda)y) \geq \min(Q(x), Q(y)) \ \forall \lambda \in [0, 1] \ \forall x, y \in D_p$$

Geometrically speaking, a quasi-concave function has the property to be above the lowest values recorded at the ends of some segment. This property is equivalent with the convexity of the set $Q^{-1}[a, \infty) \ \forall a \in \mathbf{R}$, where $Q^{-1}[a, \infty) = \{x \in R_p \mid Q(x) \geq a\}$.

On the other hand, let note that any monotone function defined on a convex set is quasi-concave, so the condition can be eliminated.

To simplify further considerations, however, we require an additional condition, namely:

FP6'. The production function is concave.

The concavity of a function means that:

$$Q(\lambda x+(1-\lambda)y) \geq \lambda Q(x)+(1-\lambda)Q(y) \ \forall \lambda \in [0, 1] \ \forall x, y \in D_p$$

or, in other words, its graph is above all straight line determined by any points of it.

Let note also, that a concave function defined on a convex domain is automatically quasi-concave (*but not each other*).

Following the concavity, we have that the production increases more slowly with amplification of production factors.

Considering a production function $Q:D_p \to \mathbf{R}_+$ and $\overline{Q} \in \mathbf{R}_+$ - fixed, the set of inputs which generate the production $\overline{Q}$ called isoquant. An isoquant is therefore characterized by: $\{(x_1,...,x_n) \in D_p \mid Q(x_1,...,x_n) = \overline{Q}\}$ or, in other words, it is the inverse image $Q^{-1}(\overline{Q})$.

We will say that a production function $Q:D_p \to \mathbf{R}_+$ is constant return to scale if $Q(\lambda x_1,...,\lambda x_n) = \lambda Q(x_1,...,x_n)$, with increasing return to scale if $Q(\lambda x_1,...,\lambda x_n) > \lambda Q(x_1,...,x_n)$ and decreasing return to scale if $Q(\lambda x_1,...,\lambda x_n) < \lambda Q(x_1,...,x_n) \ \forall \lambda \in (1, \infty) \ \forall (x_1,...,x_n) \in D_p$.

2. The Main Indicators of Production Functions

Let a production function:

$$Q:D_p\to\mathbf{R}_+,\ (x_1,...,x_n)\to Q(x_1,...,x_n)\in\mathbf{R}_+\forall(x_1,...,x_n)\in D_p$$

We will call the marginal physical production (*marginal productivity*) relative to a production factor x_i: $\eta_{x_i}=\frac{\partial Q}{\partial x_i}$ and represents the trend of variation of production at the variation of the factor x_i.

In particular, for a production function of the form: Q=Q(K, L) we have $\eta_K=\frac{\partial Q}{\partial K}$ - called the marginal efficiency of capital and $\eta_L=\frac{\partial Q}{\partial L}$ - called the marginal efficiency of labor.

We call the average physical production (*productivity*) relative to a production factor x_i: $w_{x_i}=\frac{Q}{x_i}$ and represents the value of production at the consumption of a unit of factor x_i.

In particular, for a production function of the form: Q=Q(K, L) we have: $w_K=\frac{Q}{K}$ - called the productivity (*efficiency*) of capital, and $w_L=\frac{Q}{L}$ - the productivity of labor.

In the general case of the variation of all inputs, for k_1 units of input 1,...,k_n units of input n, we will consider first the simple way $\gamma:[0, 1]\to\mathbf{R}^n$, $\gamma(t)=(tk_1,..., tk_n)$. This is nothing more than the large diagonal of the n-dimensional parallelepiped: $[0, k_1]\times...\times[0, k_n]$. Let also the differential form:

$$dQ=\frac{\partial Q}{\partial x_1}dx_1+...+\frac{\partial Q}{\partial x_n}dx_n$$

that is continuous everywhere after the C^2 character of Q. Along the path γ, the integral of dQ is defined by:

$$\int_\gamma dQ=\int_0^1\left(\frac{\partial Q}{\partial x_1}(\gamma_1(t),...,\gamma_n(t))\gamma'_1(t)+...+\frac{\partial Q}{\partial x_n}(\gamma_1(t),...,\gamma_n(t))\gamma'_n(t)\right)dt$$

Where $\gamma_1,..., \gamma_n$ are the components of γ. The Leibniz-Newton's theorem for exact differential forms (forms with property $\exists Q$ such that $\omega=dQ$) states that: $\int_\gamma dQ =Q(\gamma(1))-Q(\gamma(0))$.

In the present case:

$$Q(k_1,..., k_n)-Q(0,..., 0)=\int_0^1\left(\frac{\partial Q}{\partial x_1}(k_1t,...,k_nt)k_1+...+\frac{\partial Q}{\partial x_n}(k_1t,...,k_nt)k_n\right)dt=$$

$$k_1\int_0^1\eta_{x_1}(k_1t,...,k_nt)dt+...+k_n\int_0^1\eta_{x_n}(k_1t,...,k_nt)dt$$

Because $Q(0)=0$, resulting the final formula:

$$Q(k_1,..., k_n)=k_1\int_0^1\eta_{x_1}(k_1t,...,k_nt)dt+...+k_n\int_0^1\eta_{x_n}(k_1t,...,k_nt)dt$$

The marginal coefficient of a factor x_i is $\gamma_{x_i}=\frac{\partial x_i}{\partial Q}$ and represents the trend of variation of x_i (caeteris paribus) relative to Q or, otherwise, the change production needs for an additional input at an infinitesimal variation of production.

In particular, for a production function of the form: $Q=Q(K, L)$ we have: $\gamma_K=\frac{\partial K}{\partial Q}$ - the marginal capital coefficient and $\gamma_L=\frac{\partial L}{\partial Q}$ - the marginal coefficient of labor.

We will call also, the average coefficient of a production factor x_i: $\nu_{x_i}=\frac{x_i}{Q}$ and it is the necessary of factor (caeteris paribus) to achieve a given level of production.

In particular, for a production function of the form: $Q=Q(K, L)$ we have: $\nu_K=\frac{K}{Q}$ - the average coefficient of capital, and $\nu_L=\frac{L}{Q}$ - the average coefficient of labor.

It is obvious that:

$$\nu_{x_i}=\frac{1}{w_{x_i}},\ w_{x_i}=\frac{1}{\nu_{x_i}}$$

and, if caeteris paribus hypothesis:

$$\eta_{x_i} = \frac{1}{\gamma_{x_i}},\ \gamma_{x_i} = \frac{1}{\eta_{x_i}}$$

In particular, for Q=Q(K, L) we have:

$$\nu_K = \frac{1}{w_K},\ w_K = \frac{1}{\nu_K},\ \nu_L = \frac{1}{w_L},\ w_L = \frac{1}{\nu_L},\ \eta_K = \frac{1}{\gamma_K},\ \gamma_K = \frac{1}{\eta_K},\ \eta_L = \frac{1}{\gamma_L},\ \gamma_L = \frac{1}{\eta_L}$$

It is called global average productivity the ratio of output produced and the sum of all factors of production used:

$$w_{av,\,g} = \frac{Q}{\sum_{i=1}^{n} x_i}$$

With the concept of the notion of average coefficient of a factor of production, we can write:

$$w_{av,\,g} = \frac{Q}{\sum_{i=1}^{n} x_i} = \frac{1}{\sum_{i=1}^{n} \frac{x_i}{Q}} = \frac{1}{\sum_{i=1}^{n} \nu_{x_i}}$$

By analogy with this notion, we will call global marginal productivity:

$$w_{marg,\,g} = \frac{1}{\sum_{i=1}^{n} \gamma_{x_i}}$$

In discrete terms, we have:

$$w_{marg,\,g} = \frac{1}{\sum_{i=1}^{n} \gamma_{x_i}} = \frac{1}{\sum_{i=1}^{n} \frac{\Delta x_i}{\Delta Q}} = \frac{\Delta Q}{\sum_{i=1}^{n} \Delta x_i}$$

therefore the marginal productivity represents the global changes obtained from the additional production from each factor.

In connection with these last two indicators ought to make some clarifications. On the one hand, the global average productivity and the overall marginal have the disadvantage of dividing the production to an amount of heterogeneous factors of production. On the other hand, geometrically speaking, the two types of productivity are not clear and have not an unambiguous representation as in the case of average productivity or marginal corresponding to a single factor.

For this reason, we will define another indicator of global average productivity, even if not appropriately respond to the objection above, will satisfactorily answer to the second requirement.

We will call global average productivity in the meaning of the Euclidean norm, the ratio of the production and the norm of the vector inputs:

$$w_{av,\,gn}=\frac{Q}{\sqrt{\sum_{i=1}^{n}x_i^2}}$$

We therefore have:

$$w_{av,\,gn}=\frac{Q}{\sqrt{\sum_{i=1}^{n}x_i^2}}=\frac{1}{\sqrt{\sum_{i=1}^{n}\left(\frac{x_i}{Q}\right)^2}}=\frac{1}{\sqrt{\sum_{i=1}^{n}\nu_{x_i}^2}}$$

Another useful formula can be obtained considering the angles that determine the input vector with the coordinate axes:

$$\cos\alpha_i=\frac{x_j}{\sqrt{\sum_{i=1}^{n}x_i^2}},\ i=\overline{1,n}$$

It follows:

$$w_{av,\,gn}=\frac{Q}{\sqrt{\sum_{i=1}^{n}x_i^2}}=\frac{Q}{x_j}\frac{x_j}{\sqrt{\sum_{i=1}^{n}x_i^2}}=w_{x_j}\frac{x_j}{\sqrt{\sum_{i=1}^{n}x_i^2}}=w_{x_j}\cos\alpha_j,\ j=\overline{1,n}$$

We will call now the overall marginal productivity in the meaning of the norm:

$$w_{marg,\,gn}=\frac{1}{\sqrt{\sum_{i=1}^{n}\gamma_{x_i}^2}}$$

In discrete terms, we have:

$$w_{marg,\,gn}=\frac{1}{\sqrt{\sum_{i=1}^{n}\gamma_{x_i}^2}}=\frac{1}{\sqrt{\sum_{i=1}^{n}\frac{(\Delta x_i)^2}{(\Delta Q)^2}}}=\frac{\Delta Q}{\sqrt{\sum_{i=1}^{n}(\Delta x_i)^2}}$$

Considering the factors i and j with $i \neq j$, we define the restriction of production area: $P_{ij}=\{(x_1,...,x_n) \mid x_k=a_k=\text{const},\ k=\overline{1,n},\ k\neq i,\ j,\ x_i,\ x_j \in D_p\}$ relative to the two factors when the others have fixed values. Also, let: $D_{ij}=\{(x_i,\ x_j) \mid (x_1,...,x_n)\in P_{ij}\}$ - the domain of production relative to factors i and j.

We define: $Q_{ij}:D_{ij}\to\mathbf{R}_+$ - the restriction of the production function to the factors i and j, i.e.: $Q_{ij}(x_i,\ x_j)=Q(a_1,...,\ a_{i-1},\ x_i,\ a_{i+1},...,\ a_{j-1},\ x_j,\ a_{j+1},...,a_n)$. The functions Q_{ij} define a surface in $\mathbf{R}^3$ for every pair of factors (i, j).

We will call partial marginal rate of technical substitution of the factors i and j, relative to D_{ij} (caeteris paribus), the opposite change in the amount of factor j to substitute a variation of the quantity of factor i in the situation of conservation production level.

We will note below:

$$\text{RMS}(i,\ j,\ D_{ij})=-\frac{dx_j}{dx_i}$$

Since $Q_{ij}(x_i,\ x_j)=Q_0=\text{constant}$, we obtain by differentiation: $dQ_{ij}(x_i,\ x_j)=0$ that is: $\frac{\partial Q_{ij}}{\partial x_i}dx_i+\frac{\partial Q_{ij}}{\partial x_j}dx_j=0$ therefore:

$$-\frac{dx_j}{dx_i}=\frac{\frac{\partial Q_{ij}}{\partial x_i}}{\frac{\partial Q_{ij}}{\partial x_j}}=\frac{\left.\frac{\partial Q}{\partial x_i}\right|_{D_{ij}}}{\left.\frac{\partial Q}{\partial x_j}\right|_{D_{ij}}}=\frac{\left.\eta_{x_i}\right|_{D_{ij}}}{\left.\eta_{x_j}\right|_{D_{ij}}}$$

We can write: $\text{RMS}(i,\ j,\ D_{ij})=\frac{\left.\eta_{x_i}\right|_{D_{ij}}}{\left.\eta_{x_j}\right|_{D_{ij}}}$ which is a function of x_i and x_j. In an arbitrary point $\overline{x}=\left(\overline{x}_1,...,\overline{x}_n\right)$:

$$\text{RMS}(i,\ j,\ \overline{x})=\frac{\eta_{x_i}(\overline{x})}{\eta_{x_j}(\overline{x})}$$

Now consider the case in which all factors consumption varies. Let therefore be an arbitrary point $\overline{x}\in D_p$ such that $Q(\overline{x})=Q_0=\text{constant}$ and $\eta_{x_k}(\overline{x})\neq 0,\ k=\overline{1,n}$. Differentiating with respect to $\overline{x}$ we have: $0=dQ=\sum_{j=1}^{n}\frac{\partial Q}{\partial x_j}dx_j$ from where:

$\frac{\partial Q}{\partial x_i}+\sum_{\substack{j=1\\j\neq i}}^{n}\frac{\partial Q}{\partial x_j}\frac{dx_j}{dx_i}$=0. In terms of marginal production, we can write: $\eta_{x_i}+\sum_{\substack{j=1\\j\neq i}}^{n}\eta_{x_j}\frac{dx_j}{dx_i}$=0. Noting $\frac{dx_j}{dx_i}$=y_j, j=$\overline{1,n}$, j≠i, follows: $\eta_{x_i}+\sum_{\substack{j=1\\j\neq i}}^{n}\eta_{x_j}y_j$ = 0.

With the partial substitution marginal rate introduced above, we get:

$$\sum_{\substack{j=1\\j\neq i}}^{n}\frac{y_j}{RMS(i,j,\overline{x})}=-1$$

The above relationship is nothing but the equation of a hyperplane in $\mathbf{R}^{n-1}$ of coordinates (y_1,..., $\hat{y}_i$,...,y_n) (*the sign ^ meaning that that term is missing*) that intersects the coordinate axes in RMS(i, j, $\overline{x}$). This hyperplane is the locus of consumption factors variations relative to a change in the i-th factor consumption such that the production remains constant and is called the marginal hyperplane of technical substitution between factor i and the other factors (noted below $H_{mi,\,j}$).

In particular, for two factors, the marginal hyperplane of technical substitution between the factor i and the factor j from $\mathbf{R}_+$, is reduced to: $\frac{y_j}{RMS(i,j,\overline{x})}=-1$ where y_j=$\frac{dx_j}{dx_i}$. Therefore, $\frac{dx_j}{dx_i}$=-y_j=$-RMS(i,j,\overline{x})$ which is consistent with the definition of the partial marginal rate of technical substitution.

We will define now the global marginal rate of substitution between the i-th factor and the others as the distance from the origin to the marginal hyperplane of technical substitution, namely:

$$RMS(i,\overline{x})=\frac{1}{\sqrt{\sum_{\substack{j=1\\j\neq i}}^{n}\frac{1}{RMS^2(i,j,\overline{x})}}}=\frac{1}{\sqrt{\sum_{\substack{j=1\\j\neq i}}^{n}\frac{\eta_{x_j}^2(\overline{x})}{\eta_{x_i}^2(\overline{x})}}}=\frac{\eta_{x_i}(\overline{x})}{\sqrt{\sum_{\substack{j=1\\j\neq i}}^{n}\eta_{x_j}^2(\overline{x})}}$$

We note that for the particular case of two factors, is obtained, as above:

$$RMS(i,\overline{x})=\frac{\eta_{x_i}(\overline{x})}{\eta_{x_j}(\overline{x})}$$

Considering now v=(y_1,..., $\hat{y}_i$,..., y_n)$\in H_{mi,\ j}$ we have: $\|v\| = \sqrt{\sum_{\substack{j=1\\ j\neq i}}^{n} y_j^2}$ and from the Cauchy-Schwarz inequality:

$$\frac{\|v\|}{\left|RMS(i,\bar{x})\right|} = \sqrt{\sum_{\substack{j=1\\ j\neq i}}^{n} y_j^2}\ \sqrt{\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{RMS^2(i,j,\bar{x})}} \geq \left|\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{y_j}{RMS(i,j,\bar{x})}\right| = 1$$

therefore $\|v\| \geq \left|RMS(i,\bar{x})\right|$.

Like a conclusion, the global marginal rate of technical substitution is the minimum (in the meaning of norm) of changes in consumption of factors so that the total production remain unchanged.

Considering now the marginal hyperplane of technical substitution: $\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{y_j}{RMS(i,j,\bar{x})} = -1$ the equation of the normal from origin to it, is:

$$\frac{y_1}{\frac{1}{RMS(i,1,\bar{x})}} = \ldots = \frac{y_{i-1}}{\frac{1}{RMS(i,i-1,\bar{x})}} = \frac{y_{i+1}}{\frac{1}{RMS(i,i+1,\bar{x})}} = \ldots = \frac{y_n}{\frac{1}{RMS(i,n,\bar{x})}}$$

from where:

$$\begin{cases} y_1 = \dfrac{\lambda}{RMS(i,1,\bar{x})} \\ \ldots \\ y_{i-1} = \dfrac{\lambda}{RMS(i,i-1,\bar{x})} \\ y_{i+1} = \dfrac{\lambda}{RMS(i,i+1,\bar{x})} \\ \ldots \\ y_n = \dfrac{\lambda}{RMS(i,n,\bar{x})} \end{cases},\ \lambda \in \mathbf{R}$$

The intersection of the normal with the hyperplane, represents the coordinates of the point of minimal norm. We therefore have: $\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{\lambda}{RMS^2(i,j,\overline{x})}=-1$ from where:

$\lambda=-\frac{1}{\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{RMS^2(i,j,\overline{x})}}$ and the point of minimal norm has the coordinates:

$$-RMS^2(i,\overline{x})\left(\frac{1}{RMS(i,1,\overline{x})},\ldots,\frac{\hat{1}}{RMS(i,i,\overline{x})},\ldots,\frac{1}{RMS(i,n,\overline{x})}\right)=$$

$$-\frac{\eta_{x_i}(\overline{x})}{\sum_{\substack{j=1\\ j\neq i}}^{n}\eta_{x_j}^2(\overline{x})}\left(\eta_{x_1}(\overline{x}),\ldots,\hat{\eta}_{x_i}(\overline{x}),\ldots,\eta_{x_n}(\overline{x})\right)$$

which norm is nothing else that $|RMS(i,\ \overline{x})|$.

The coordinates of the above point is no more than minimal vector (in the meaning of norm) of changes in consumption so that total output remains unchanged. We will say briefly that this is the minimal vector of technical substitution of the factor i.

In particular, for a production function of the form: Q=Q(K, L) we have:

$$RMS(K,L)=\frac{\eta_K}{\eta_L},\ RMS(L,K)=\frac{\eta_L}{\eta_K}$$

It is called elasticity of production in relation to a production factor x_i: $\varepsilon_{x_i}=\frac{\frac{\partial Q}{\partial x_i}}{\frac{Q}{x_i}}=$

$\frac{\eta_{x_i}}{w_{x_i}}$ - the relative variation of production at the relative variation of factor x_i.

In particular, for a production function of the form: Q=Q(K, L) we have $\varepsilon_K = \dfrac{\frac{\partial Q}{\partial K}}{\frac{Q}{K}} = \dfrac{\eta_K}{w_K}$ - called the elasticity of production in relation to the capital and $\varepsilon_L = \dfrac{\frac{\partial Q}{\partial L}}{\frac{Q}{L}} = \dfrac{\eta_L}{w_L}$ - the elasticity factor of production in relation to the labor.

3. Application

Considering now a production function $Q:D_p \to \mathbf{R}_+$, $(x_1,...,x_n) \to Q(x_1,..., x_n) \in \mathbf{R}_+ \forall (x_1,..., x_n) \in D_p$ with constant return to scale, let note for an arbitrary factor (for example x_n):

$$\chi_i = \frac{x_i}{x_n},\ i = \overline{1, n-1}$$

We will have:

$$Q(x_1,..., x_n) = x_n Q\left(\frac{x_1}{x_n},...,\frac{x_{n-1}}{x_n},\frac{x_n}{x_n}\right) = x_n Q(\chi_1,...,\chi_{n-1},1)$$

Considering the restriction of the production function at $D_p \cap \mathbf{R}_+^{n-1} \times \{1\}$: $q(\chi_1,...,\chi_{n-1}) = Q(\chi_1,...,\chi_{n-1},1)$ we can write:

$$Q(x_1,..., x_n) = x_n q(\chi_1,...,\chi_{n-1})$$

With the new function introduced, the above indicators are:

- $\eta_{x_i} = \dfrac{\partial q}{\partial \chi_i}$, $i = \overline{1, n-1}$
- $\eta_{x_n} = q - \sum_{i=1}^{n-1} \dfrac{\partial q}{\partial \chi_i} \chi_i$
- $w_{x_i} = \dfrac{q}{\chi_i}$, $i = \overline{1, n-1}$

- $w_{x_n} = q$

- RMS(i, j)= $\dfrac{\dfrac{\partial q}{\partial \chi_i}}{\dfrac{\partial q}{\partial \chi_j}}$, i, j=$\overline{1, n-1}$

- RMS(i, n)= $\dfrac{\dfrac{\partial q}{\partial \chi_i}}{q - \sum_{i=1}^{n-1} \dfrac{\partial q}{\partial \chi_i} \chi_i}$, i=$\overline{1, n-1}$

- RMS(n, j)= $\dfrac{q - \sum_{i=1}^{n-1} \dfrac{\partial q}{\partial \chi_i} \chi_i}{\dfrac{\partial q}{\partial \chi_j}}$, j=$\overline{1, n-1}$

- RMS(i)= $\dfrac{\dfrac{\partial q}{\partial \chi_i}}{\sqrt{\left(q - \sum_{j=1}^{n-1} \dfrac{\partial q}{\partial \chi_j} \chi_j\right)^2 + \sum_{\substack{j=1 \\ j \neq i}}^{n-1} \left(\dfrac{\partial q}{\partial \chi_j}\right)^2}}$, i=$\overline{1, n-1}$

- RMS(n)= $\dfrac{q - \sum_{j=1}^{n-1} \dfrac{\partial q}{\partial \chi_j} \chi_j}{\sqrt{\sum_{j=1}^{n-1} \left(\dfrac{\partial q}{\partial \chi_j}\right)^2}}$

- $\varepsilon_{x_i} = \dfrac{\dfrac{\partial q}{\partial \chi_i}}{\dfrac{q}{\chi_i}}$, i=$\overline{1, n-1}$

- $\varepsilon_{x_n} = \dfrac{q - \sum_{i=1}^{n-1} \dfrac{\partial q}{\partial \chi_i} \chi_i}{q}$

4. Conclusion

After the above analysis, we have seen that the analysis of production functions, on the one hand from the axiomatic point of view and, on the other hand, on the general case of n variables, reveals very interesting aspects.

First, even a very restrictive axiomatization removes some of the common functions (Leontief case), it adds a more austerity to the notion, eliminating the use of, often negligent, of the production function.

On the other hand, the extension of the main indicators in the case of n inputs, allows the removal, on the one hand, of some absurd concepts from our point of view, such as the global average productivity and replacing them with more realistic indicators. On the other hand, the notion of global rate of substitution removes the usual drawback of partial substitutions that restrict the scope sometimes dramatically.

5. References

Chiang, A.C. (1984). *Fundamental Methods of Mathematical Economics.* McGraw-Hill Inc.

Ioan, C.A. & Ioan, G. (2011). A generalisation of a class of production functions. *Applied economics letters.* Coventry, UK: Warwick University, Volume 18, Issue 18, December 2011, pp. 1777-1784.

Ioan, C. A. & Ioan, G. (2011). The Extreme of a Function Subject to Restraint Conditions. *Acta Universitatis Danubius. Œconomica*, Vol. 7, no. 3, pp. 203-207.

Ioan, C. A. & Ioan, G. (2012). *A new approach to utility function* (to appear).

Ioan, C. A. & Ioan, G. (2012). The consumer's behavior after the preferences nature. *Acta Universitatis Danubius. Œconomica* (to appear).

Stancu, S. (2006). *Microeconomics*. Bucharest: Economica.

Varian, H. R. (2006). *Intermediate Microeconomics*. W.W.Norton & Co.