Colloquium Biometricum 41
2011, 73–86

LARS REGRESSION IN DIAGNOSING MAMMOGRAMS: SELECTION OF VARIABLES FOR ANALYSIS

Anna Bartkowiak[1,2], Adam Szustalewicz[1]

[1]Institute of Computer Science, University of Wrocław
Joliot–Curie 15, 50–383 Wrocław, Poland
[2]Wrocław High School of Applied Informatics
Wejherowska 28, 54–239 Wrocław, Poland
e–mail: aba@ii.uni.wroc.pl; asz@ii.uni.wroc.pl

Dedicated to the memory of Professor Wiktor Oktaba

Summary

Diagnosing cancer in a mammogram is a difficult task. Our aim is to explore the usefulness of so called fractal signatures for this purpose. A fractal signature is given by a vector of p real numbers characterizing the roughness of a mammogram considered as a texture file. Fractal signatures of length 48 are considered. Are all of them relevant to make the 2–group diagnosis: non–cancer or cancer? To answer this question, we used the Least–Angle Regression (LARS) which is believed more stable than the traditional forward search. By 5–fold cross–validation we found that only a small subset of variables is relevant for the diagnosis. The considerations are illustrated using data from the MIAS data base.

Key words and phrases: fractal signatures, mammogram diagnostics, linear regression, reduction of predictors, forward search, least angle regression, lars, correct classification, cross–validation

Classification AMS 2010: 62H25

1. Introduction

Breast cancer is one of the most frequent mortal diseases in women. When early diagnosed, there is a great chance that it will be cured. Therefore Medical Care organizes frequent mass screening of adult female population. During the screening, breast radiological images are taken. By inspecting a radiogram frame (called mammogram), specialists may notice a distortion of the breast mass architecture indicating a developing cancer structure. A sample of one mammogram frame is shown in Figure 1, left exhibit. Finding the cancerous structure is a difficult task, see Woodward et al. (2007). There is a need for an automated diagnosis. However, the proposed automated methods use sophisticated algorithms and the result is not obvious (Verma 2008, Sankar and Tomas 2009, 2010).

Our idea is the following one: The growth of a cancerous tumor is fractal–like, and therefore the cell agglomeration should be different from that which resulted from the normal expansion. A radiological frame (mammogram) is in fact a bit–mapped image based on pixels. As such, it is a graphic file, memorized as a matrix of pixels, where each pixel has color attributes expressed numerically as so called unsigned integers or real numbers from the interval [0,1]. Moreover, such graphical file may be viewed as representing a texture (see e.g. Figure 1, right exhibit). The texture from a tumor image should show more roughness as a normal tissue image. To compare the roughness of mammograms we propose to use the method of fractal signatures (Peleg et al. 1984). The method is conceptually simple and intuitively appealing: it permits to translate the roughness of a texture (a 3–dimensional object) to a one–dimensional vector of real numbers. Thus, for each frame of fixed size (one mammogram) one obtains for further analysis one data vector.

We will illustrate the procedure by calculating fractal signature for 60 non–cancer and 60 cancer mammograms. Each mammogram got its group label: $y = -1$ for 'no–cancer' and $y = +1$ for 'cancer'. Is it possible to built a linear discriminant function in the form of a linear regression permitting to classify the available sample into the two groups of data? The prediction may be done either using the full set of recorded variables (components of signatures), or a subset of them. How to find a relevant subset of the recorded variables? We are interested in finding regression equations which not only permit for a high classification accuracy in the training sample, but are able to do it also for test samples, not used for training. After reducing the data to fewer (hopefully relevant) variables, it is easier to perform a more sophisticated analysis, using, e.g. generalized discriminant analysis or/and kernel methods, see e.g. Hastie et al. (2010), Atkinson et al. (2004), Deręgowski and Krzyśko (2010).

Next section (2) describes shortly the *mammo120 data* used for analysis. Section 3 recalls the concept of ordinary least squares regression and shows the

results of classification of the mammo120 data when using the full set of variables. Search for relevant variables using the LARS algorithm (Efron et al. 2004, Hastie et al. 2010, Sjőstrand et al. 2006) is described in Section 4. The efficacy of the models in classification the mammo120 data into two classes (normal and cancer) is evaluated using the full regression model and some selected submodels. The method of 5–fold cross–validation was used for obtaining test samples.

2. The mammogram data 'mammo120' and their fractal signatures

In the following we will analyse a set of 120 mammograms taken from the data base MIAS (Mammographic Image Analysis Society) available at http://peipa.essex.ac.uk/ipa/pix/mias. There are all together 322 mammogram images, each of size 1024×1024 memorized in pmg (Paint Magic) format. The images contain 'normal' breasts, i.e. without malformations, and 'ab–normal', that is with distorted structure, like calcification, benign or malicious tumors. The centers of the distortions and their radiuses may be found in the description of the data base.

We have taken from this source 12 normal mammograms (denoted in the following as: ncncr) and 12 with malicious tumor (denoted in the following as cncr mammograms). This sample was augmented by varying the center of each mammogram by 5 pixels up, down, left and right. In such a way, we got five replicates of each sample, together 120 mammograms for further analysis. The obtained set of mammograms will be in the following called *mammo120*. It contains 60 normal and 60 cancer images. For further analysis, a square frame of 81×81 pixels was cut off from each mammogram according to the following principle: (a) for 'ab–normal' mammograms the indicated center of distortions was taken as the central pixel of the square; (b) for 'normal' mammograms the center was chosen somehow arbitrarily, with the attempt to locate it in similar region as those used in the case (a) above. Exemplary (square) frame from the mammo120 set is shown in left exhibit of Figure 1.

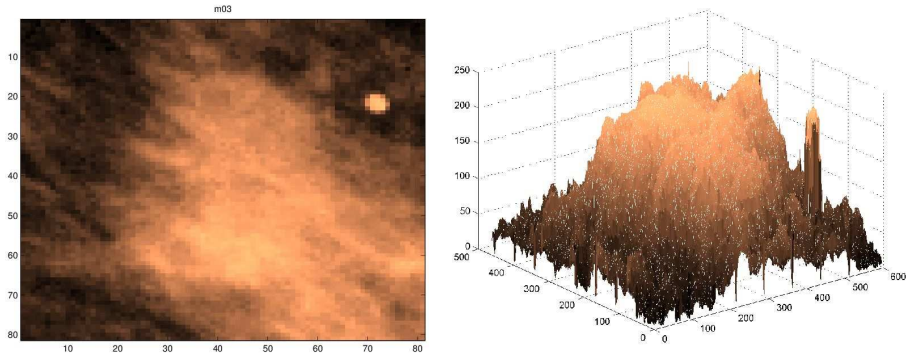

Fig. 1. A sample mammogram (m03). Left: as graphical frame containing a square image of size 81×81 pixels. Right: The same image, viewed as texture

For each square frame, a fractal signature with p=*48* components was calculated using the blanket algorithm (Peleg et al. 1984, Białek 2010). The necessary software (in Matlab) was taken from Białek (2010), where also a preliminary analysis of 10 normal and 10 cancer mammograms may be found. Applying Principal Component Analysis (PCA), he found that the first two Principal Components (PCs) explain more than 95% of total variance. When taking only these two PCs, he found a linear discriminant function with the effect of misclassifying one normal sample (small distortion of the cell architecture, not yet visible to the eye of the expert?) and one cancer sample (wrongly recognized by the expert?). The correct classification percentage – based on the investigated sample – was 90% (Białek 2010).

Now, before starting the analysis, each signature was standardized to have zero mean and unit standard deviation. This has the consequence that we have only $p = 47$ linearly independent variables. Twenty–four exemplary fractal signatures (12 ncnr and 12 cncr samples) are shown in Figure 2.

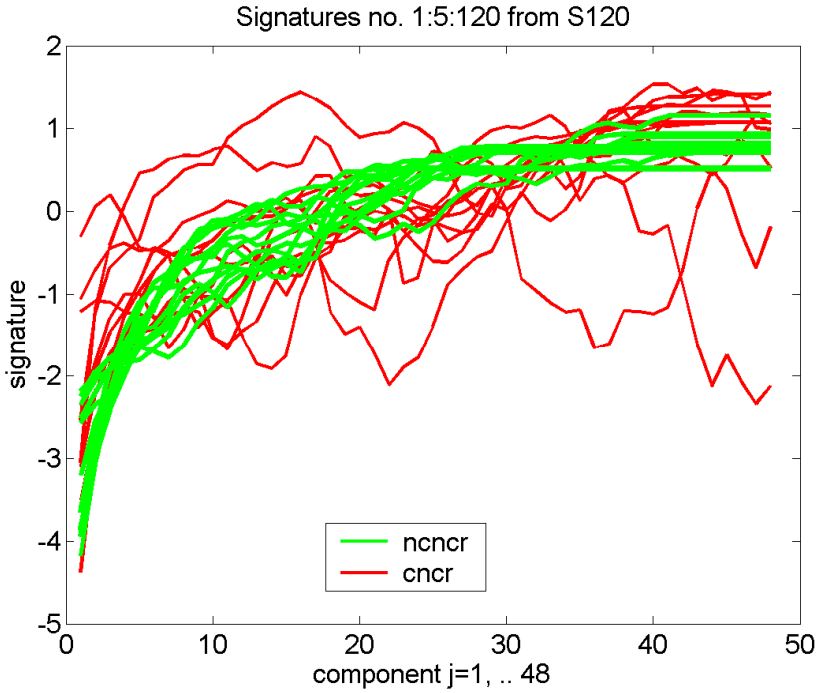

Fig. 2. Curves exhibiting fractal signatures for 12 ncncr and 12 cncr images. Notice the rather regular shape of the normal curves and the much–dispersed shape of the cncr curves. Each curve was row–wise standardized to have zero mean and unit standard deviation

All the recorded signatures were put together into a data matrix $\mathbf{X}$ of size 120×48. Because of the linear dependency of elements in each row, we have dropped the last column to obtain the full rank matrix $\mathbf{X}$ of size 120×47. This matrix was supplemented with a label vector $\boldsymbol{y}$ of size 120×1 with values –1 or +1 indicating the ncncr or cncr status of the respective sample. The pair $(\mathbf{X},\mathbf{y})$ will serve as the basis for analysis in next sections. Summarizing, the method of fractal signatures allows for a representation of a data object (mammogram characterized by $81*81=6561$ pixel values) by a numerical vector with $p=48$ (or 47) components. It may be depicted as a time plot having the shape of a curve, with time meaning subsequent steps of the blanket algorithm. Exemplary curves for 24 mammograms are shown in Figure 2.

3. Full least squares regression

Theoretical framework. Say, we have a set of training data in the form: $(\boldsymbol{x}_1, y_1)$, $(\boldsymbol{x}_2, y_2)$, … , $(\boldsymbol{x}_N, y_N)$. Each $\boldsymbol{x}_i = (x_{i1}, x_{i2,} \ldots, x_{ip})$ is a vector of feature measurements for the i^{th} case; y_i is the observed output, called also target or the dependent variable. The label variable y_i (*i=1,…, N*) takes values $y_i = -1$ for ncncr, and $y_i = 1$ for cncr cases. The whole data may be put together as the pair $(\mathbf{X}, \mathbf{y})$, where $\mathbf{X} = (x_{ij})$, $\mathbf{y} = (y_i)$, $i = 1, \ldots, N$, $j = 1, \ldots, p$.. The classical theory of linear models assumes that

$$\mathbf{y} = \mathbf{X}\mathbf{b} + b_0 + \boldsymbol{\varepsilon} \tag{3.1}$$

where $\mathbf{X}$, the data matrix, is assumed to be composed of fixed real values, $\mathbf{b} = (b_1, \ldots, b_p)$ and b_0 are parameters of the model, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)$ is a vector of independent random errors with expected value equal to zero and variance σ^2 for each *i*.

The classical least squares (LS) minimization criterion is defined as the Residual Sum of Squares (RSS) computed as the quadratic form

$$RSS(\mathbf{b}, b_0) = (\mathbf{y} - \mathbf{X}\mathbf{b} - b_0)^T (\mathbf{y} - \mathbf{X}\mathbf{b} - b_0) \tag{3.2}$$

The LS minimization problem is to find the vector $\hat{\mathbf{b}}$ and the constant $\hat{b}_0$ minimizing the quadratic form $RSS(\mathbf{b}, b_0)$ over all real values of $\mathbf{b}, b_0$.

It can be shown that the regression coefficients $\hat{\mathbf{b}}$ and the constant $\hat{b}_0$ may be obtained as

$$\hat{\mathbf{b}} = \left(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}\ , \qquad \hat{b}_0 = \bar{\mathbf{y}} - \hat{\mathbf{b}}^T \bar{\mathbf{x}} \tag{3.3}$$

where the '~' symbol means 'mean–centered' $\mathbf{X}$ and $\mathbf{y}$; $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_p)^T$ denotes the vector of mean values of consecutive columns of the observed matrix $\mathbf{X}$. The Gramm matrix $\left(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}\right)$, called also the adjusted cross–product matrix of the variables, should be of full rank.

In the case when $\mathbf{X}$ and $\mathbf{y}$ are columnwise centered to have means equal zero, formula (3.3) simplifies to

$$\hat{\mathbf{b}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}, \quad \hat{b}_0 = 0 \tag{3.4}$$

and the predicted values of the target variable $\boldsymbol{y}$ are then obtained as:

$$\mathbf{y} = \mathbf{X}\hat{\mathbf{b}}. \tag{3.5}$$

The computations. Taking the *mammo120 data* given as the pair $(\mathbf{X},\mathbf{y})$ described in Section 2, firstly the full LS regression model formulated in eq. (3.1) was computed. Before starting the calculations, the matrix $\mathbf{X}$ and the vector $\mathbf{y}$ were centered to have zero means. The Matlab function regres was used for the regression calculations. Taking $p = 48$, Matlab has issued the warning that the cross–product matrix $\left(\mathbf{X}^T\mathbf{X}\right)$ is rank–deficient and is only of rank 47. Taking $p = 47$, there were no warnings. We got the estimates of the intercept b_0 and of the regression coefficients $\mathbf{b}$ of size 47×1 together with their 95% confidence intervals. They are shown in Figure 3, top exhibit, in the sequence $b_0, b_1, \ldots, b_{47}$. Notice that the estimated value of b_0 equals *0* and has a confidence interval of width *0*. Notice also the remarkable oscillations of values of succeeding regression coefficients for lower and higher No.s of the variables. Sixteen confidence intervals do not enclose zero – it is said in such a case that these coefficients are statistically significant at the 95% level, which means that these coefficients may be considered as different from zero.

Investigating the global dependency of the target values $\mathbf{y}$ with its explanatory variables recorded in succeeding columns of the data matrix $\mathbf{X}$ we got the multiple squared correlation coefficient *RR* equal to $RR = 0.8619$, with $F = 9.5569$ and $P < 0.0001$. The tested hypothesis H_0 was: The population squared multiple correlation coefficient of the considered variables equals 0. The performed test rejected decidedly that hypothesis. This means that generally, the investigated predictors (the components of the recorded fractal signatures) have some power to predict the ncncr or cncr class indicated in $\mathbf{y}$, so we are entitled to investigate further this topic, for example to explore which considered variables contribute mostly to the stated correlation.

Class assignements. How exactly are the class assignments obtained by the estimated regression coefficients? The predicted values $\hat{\mathbf{y}}$, obtained from eq.

(3.5) as $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$ are shown in Figure 3, bottom exhibit. The first 60 items from the mammo120 data belonged to the ncncr class and had $Y = -1$ as its target value. Looking at the plot one may notice that all values $\hat{y}_i$, $i = 1,\ldots,60$ are negative. So the assignment by the constructed regression function was correct for all the ncncr cases. The remaining 60 items belonged to the cncr class and except one item (no. 84) got positive $\hat{y}_i$ values. Again, the assignments to this class were correct for all but one item. This confirms our previous observations that fractal signatures have some predictive power for cancer diagnosis.

The high prediction accuracy shown in Figure 3, bottom exhibit, sounds very optimistic. However the golden principle in experimental design is: firstly derive some rule (prediction formula) using a 'learning sample', and then test the derived rule (formula) using a different, so called 'test sample'. We have used for this purpose cross–validation samples (five–fold or two–fold cross–validation) and stated that in the test samples the prediction accuracy was about 80% (see Table 1), or lower. This was obtained, when applying linear regression playing at the same time the role of a discriminant function, which is the simplest offer of pattern recognition methods for this purpose. There are other methods, more sophisticated, reported to be more effective, however they are computationally more complex and expensive. So we ask: Is it possible to reduce the set of predictors in our data, to make the diagnosis by more sophisticated algorithms easier? The problem of selecting a smaller set of predictors will be considered in next section.

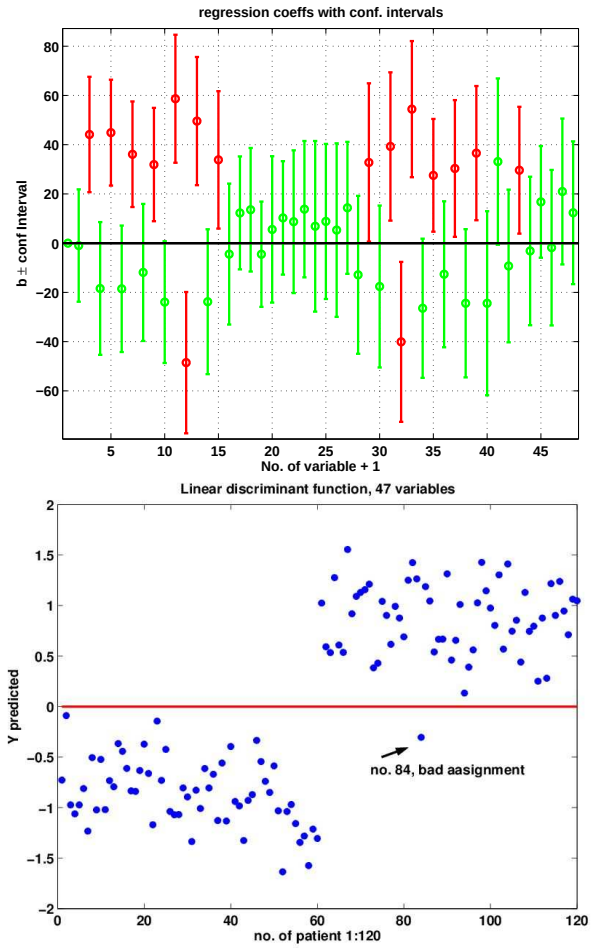

Fig. 3. Full least squares regression in p=47 variables fitted from n=120 data points.
Top: regression coefficients with 95% confidence intervals. Sixteen regression coefficients are statistically significant, which means that they are different from zero.
Bottom: assignment of mammograms to the non–cancer or cancer class by the full linear regression function. Only one item (from the cancer class) is wrongly assigned to the non–cancer class

4. Search for relevant variables using the LARS algorithm

Introduction to the LARS algorithm. Our goal is to find the variables (components of signatures) that matter in establishing the prediction of the ncncr or cncr class. The predicting formula should have the form of a sparse linear re-

gression function, which means that not all predictors will enter as arguments of the sought function. Traditionally this is done by a subset search. With larger number of variables this is done usually using a forward search. There are many variants of such algorithms (Atkinson et al. 2004; Hastie et al. 2010). We decided to use for this purpose the LAR (Least Angle Regression) algorithm, as proposed originally by Efron et al. 2004), see also Hastie et al. (2010). LAR is intimately connected with the LASSO algorithm, which permits to control the L1 norm of the derived regression coefficients. The combination LAR + LASSO got the name LARS. In the following we will use only the LAR version of the algorithm.

The LAR algorithm is a relative newcomer and represents a soft computing approach. In traditional algorithms, the selected variables enter hardly the active variables set. On the opposite, the LAR algorithm takes from a predictor "as much as it deserves". Description of LARS taken from Hastie et al. (2010): "At the first step it identifies the variable most correlated with the response. Rather than fit this variable completely, LAR moves the coefficient of this variables continuously toward its least–square value (causing its correlation with the evolving residual to decrease in absolute value). As soon, as another variable 'catches up' in terms of correlation with the residual, the process is paused. The second variable then joins the active set, and their coefficients are moved together in a way that keeps their correlations tied and decreasing The process is continued until all variables are in the model and ends at the full least–squares fit." The result of the procedure appears as a curve showing the evolving path of the values of the regression coefficients.

The LARS algorithm works in following steps (see Hastie et al. 2010):

1. Standardize the predictors so to have mean zero and unit norm. Define the residual vector $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$, with $\hat{\mathbf{y}}$ denoting the vector $\mathbf{y}$ evaluated under the presently assumed regression model, i.e. using the regression coefficients $\{b_j\}$ being in the regression set. Assume at the beginning that $b_1 = b_2 = \cdots = b_p = 0$, which means that the regression set is empty. Then $\hat{\mathbf{y}} = 0$.
2. Find the predictor $\mathbf{x}_j$ most correlated with $\mathbf{r}$.
3. Move b_j from zero towards its LS coefficient computed from the scalar product $\langle \mathbf{x}_j, \mathbf{r} \rangle$ until some other competitor $\mathbf{x}_k$ has as much correlation with the current residual as does $\mathbf{x}_j$. Notice, this is done by inspecting the scalar products of the respective variables.
4. Move b_j and b_k in the direction defined by the LS regression of the current residual $\mathbf{r}$ on $(\mathbf{x}_j, \mathbf{x}_k)$, until some competitor $\mathbf{x}_l$ has as

much correlation with the current residual – as the current active variables $\mathbf{x}_j$, $\mathbf{x}_k$. Retain the values b_j and b_k at this moment and use them as the LS estimates computed from eq. (3.3) taking as the matrix $\mathbf{X}$ the columns of active variables at this step.

5. Continue in this way until all p predictors have been entered. After $\min(N-1, p)$ steps, we arrive at the full least square regression.

The LAR algorithm is believed to be more stable and more 'democratic', as compared to the traditional forward search. The regression coefficients and fitted values are computed in a more cautious way than using LS algorithm which is described as 'greedy'. The authors (Efron et al. 2004) elaborated an efficient algorithm for computing the full path of the development of the regression coefficients and implemented it in S and R. The complexity is within the range of the ordinary LS algorithm. A Matlab implementation by K. Skoglund is available at http://www2.imm.dtu.dk/~ksjo/kas/software/index.html (Sjőstrand, 2006). We gratefully acknowledge the use of that software for our computing of the LAR regression using the lars function from that package.

Results of computations. The LARS algorithm yielded the same full regression coefficients as the LS method, – which was to be expected. Travelling towards the final full solution from $k = 0$ (no variables in the active set) to $k = 47$ (all variables in the active set), the applied algorithms has produced (and retained) in each step a sparse set of regression coefficients. The sets were not identical with those obtained by Matlab stepwise or stepwisefit procedures working upwards. Moreover, not the same variables were selected by Skoglund's lars and Matlab stepwise functions. For example, among the active variables obtained for $k = 7$, only *three* variables were the same.

The conclusion from this part of analysis might be as follows: The data allow for a good classification of its items into the ncncr and cncr classes. However this happens when testing the regression equation 'by re–substitution', that is, using the same data. To obtain a reasonable estimate of the prediction error, one should test it using an independent sample of data. The following questions were asked:

1. What are the generalization abilities of the obtained function, when tested on an independent sample? Does it depend from the number of the predictors included into the predictive equation?
2. Not all predictors are necessary for the prediction, which means that with a smaller amount of them we may achieve (nearly) the same goal. Which variables are the most relevant and should be included into the predictive set of variables?

To answer these questions, a cross–validation experiment was performed. We have used in the experiment the five–fold cross–validation method (5 CV), see e.g. Hastie et al. (2010) and obtained results shown in Figure 4.

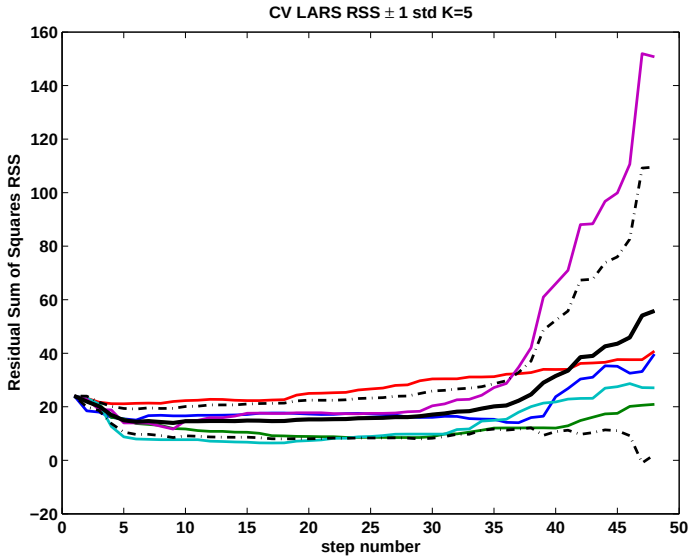

Fig. 4. Residual Sum of Squares (RSS) from test parts of the data obtained by 5–fold crossvalidation. Mean of the RSS with one standard deviation bounds is superimposed. Notice that at the beginning, the RSS is declining – until the step number k=5, then it remains flat, until k=30, and next it rises quite sharply

The applied method proceeded as follows: The entire data was subdivided into five non–overlapping parts; each of them serving in turn as a test sample for the regression calculated from the remainder of the data. For each test sample the RSS (Residual Sum of Squares) was evaluated at each step of the LARS algorithm. This resulted in five curves exhibiting the dependency of the RSS statistics from the step number. The curves are shown in Figure 4.

The mammo120 data contained 60 ncncr and 60 cncr cases. This yielded at each stage of the CV procedure 96 cases as training sample and 24 cases as test sample. Both samples were individually centered – before entering the LAR algorithm. Each of the five curves in Figure 4 shows the RSS evaluated from 24 independent cases. The exhibited CV RSS curves put against the step number (equal to the number of active variables of the actual regression equation + 1) have a characteristic shape: firstly they are decaying, which means, that the variables entering the active set of the regression, are really effective and they reduce the error. Next, there is a period of stability: the variables do not improve the separability of the ncncr and cncr sets, neither deteriorate it. This

happens for a range of approximately 25 additional variables. Finally, when introducing more variables, the error – in average – starts to rise.

Each CV learning sample yielded at step *k* the sparse linear equation

$$y^k(\mathbf{x}) = b_1^k x_1 + b_2^k x_2 + ... + b_p^k x_p \tag{4.1}$$

which, at step *k*, contained at appropriate places *k* non–zero regression coefficients b_j^k designated by the LARS algorithm, the remaining ones being by definition equal to zero. Using the above regression equation (4.1), the expected values $\hat{\boldsymbol{y}}$ were calculated from (3.5). They served for class–assignments: cases $\boldsymbol{x}$ producing *positive* values of $y(x)$ were classified as 'cncr'; cases producing *non–positive* values $y(x)$ were classified as 'ncncr'. Obviously, the number of correct classifications depends on the number of variables *k* in the regression set. Taking $k = [6, 20, 30, 35, 47]$, we obtained the percentages of correct classification shown in Table 1.

Table 1. Percentage of correct classification, when taking k variables. Symbols CV1, CV2, CV3, CV4, CV5 denote results from the five test samples obtained by 5–fold cross–validation

k	CV1	CV2	CV3	CV4	CV5	average
6	83.3333	83.3333	66.6667	91.6667	91.6667	83.3333
20	79.1667	91.6667	54.1667	91.6667	87.5000	80.8333
30	87.5000	83.3333	45.8333	91.6667	83.3333	78.3333
35	87.5000	83.3333	41.6667	95.8333	83.3333	78.3333
47	95.8333	83.3333	37.5000	87.5000	70.8333	75.0000

One may notice that the results in Table 1 are in accordance with the results shown in Figure 4. We got the answer for question 1. The prediction error depends on the number of predictors taken for analysis. A very moderate number of predictors produced the best results. With increasing number of predictors the qualitity of the prediction is deteriorating. This result needs further confirmation. What concerns the 2nd question: which are the best predictors, we did not found a definite answer. It seems that for the elaborated data the maximum is very flat and different variables may yield results with similar prediction quality.

References

Atkinson A.C., Riani M., Cerioli A. (2004). *Exploring Multivariate Data with the Forward Search*. *Springer Series in Statistics*, Springer New York, pp. XXI+621.

Białek J. (2010). Fractal signatures in investigation of mammographic images. *Master Thesis* (in Polish). Institute of Informatics, Wrocław University, Wrocław.

Deręgowski K., Krzyśko M. (2010). Nonlinear principal component analysis. *Colloquium Biometricum* 40, 105–116.

Efron B., Hastie T., Johnstone I., Tibshirani R. (2004). Least angle regression. *Annals of Statistics* 32 (2), 407–451.

Hastie T., Tibshirani R., Friedman J. (2010). *The Elements of Statistical Learning*. *Data Mining, Inference and Prediction*. 2nd Edition, New–York, Springer.

Peleg S., Noar J., Hartley R., Avnir D. (1984). Multiple resolution texture analysis and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, 518–523.

Sankar D., Thomas T. (2009). Breast cancer detection using entropy based fractal modeling of mammograms. *Int. J. of Recent Trends in Engineering*, Vol. 1, No. 3, May 2009, 171–175.

Sankar D., Thomas T. (2010). Fractal features based on differential box counting method for the categorization of digital mammograms. *Int. J. of Computer Information Systems and Industrial Management Applications* (IJCISIM) 2, 011–019.

Sjőstrand K., Stegmann M. B., Larsen R. (2006). Sparse principal component analysis in medical shape modeling. International Symposium on Medical Imaging 2006, San Diego, CA,USA, Proc. SPIE 6144, 61444X (2006); doi:10.1117/12.651658, (12 pp.).

Verma B. (2008). Novel network architecture and learning algorithm for the classification of mass abnormalities in digitized mammograms. *Artificial Intelligence in Medicine* 42, 67–79.

Woodward D.B., Gelfand A.E., Barlow W.E., Elmore J.G. (2007). Performance assessment for radiologists interpreting screening mammography. *Statist. Med*. 26, 1532–1551.