Journal of Universal Computer Science, vol. 14, no. 2 (2008), 224-236
submitted: 15/6/07, accepted: 17/11/07, appeared: 28/1/08

A Progressive Learning Method for Symbol Recognition

Sabine Barrat and Salvatore Tabbone
(University of Nancy 2
LORIA - Campus scientifique - BP 239 - 54506 Vandoeuvre-lès-Nancy Cedex - France
{barrat,tabbone}@loria.fr)

Abstract: This paper deals with a progressive learning method for symbol recognition which improves its own recognition rate when new symbols are recognized in graphic documents. We propose a discriminant analysis method which provides allocation rules from a training set of labelled data. However a discriminant analysis method is efficient only if the training set and the test data are defined in the same conditions but it is rare in real life. In order to overcome this problem, a conditional vector is added to each instance to take into account the parasitic effects between the test data and the training set. We also propose an adaptation to consider the user feedback.

Key Words: Conditional discriminant analysis, symbol recognition.

Category: I.5.3

1 Introduction

Symbol recognition is a field within graphic recognition for which a lot of efforts have already been made. Several approaches are based on feature descriptors [Adam et al., 2000] and due to the structural aspects of some symbol graph matching techniques [Llados et al., 2001] are suited to symbol recognition. However current symbol recognition methods have good results when we want to recognize few different symbols with low noise and often disconnected from the graphics. In real life, we have to work in large symbol databases and distinguish hundreds of different symbols, often complex and embedded in graphics, and those methods provide weak results. For these reasons, the problem of symbol recognition is far from being solved. Moreover in many cases it is impossible to suppose that symbols can be performed on clearly segmented instances, because the symbols are very often connected to other graphics and/or associated with texts. Therefore the well-known paradox appears : in order to correctly recognize the symbols, we should be able to segment the input data, and reciprocally to correctly segment them, we need to recognize the symbols!

This means that it is usually not possible to perform symbol recognition by simply assuming that a reliable segmentation process is available, that the symbols have been clearly extracted, normalized and noise free. Under these conditions, to improve the recognition, it is necessary to carry out learning methods. In this paper we do not consider structural approaches [Messmer and Bunke, 1996] but statistical methods. A lot of classification methods have been offered and can be divided into two classes [Jain et al., 2000]: supervised and unsupervised classification. Here we focus on a supervised learning method. More precisely we consider the linear discriminant analysis because this method is simple and fast and can be adapted to the recognition of symbols. However, the discriminant analysis presents many drawbacks: in fact it is based on some assumptions which are not always checked due to a large variability of the real data. Consequently, the discriminant analysis methods are efficient only if the training data and the test data are defined in the same conditions, but it is rarely the case for the reasons specified above. This can lead to erroneous trend in the classification, and, to overcome this problem, we use a recent approach called conditional discriminant analysis [Baccini et al., 2001]. It is a modified analysis which improves the learning by a suitable control of the possible trend.

The rest of this article is organized as follows. In the next section we recall the discriminant analysis theory. Then we describe the conditional discriminant analysis process (section 3) and show how to adapt it to a symbol recognition process. The results, obtained on a large database, are presented in section 4 and some conclusions and guidelines for future work are given in section 5.

2 Discriminant analysis

In this section we present the classical discriminant analysis, because it constitutes the core of the conditional discriminant analysis.

2.1 Definitions and notations

The discriminant analysis provides decision rules from a training set of labelled data (supervised learning). Let :

– X_j be a vector representative of an instance :

$$X_j = {}^t(X_{1,j}, X_{2,j}, ..., X_{p,j}),$$

where p is the number of the characteristic vector.

– $\overline{X_l}$ be the barycenter of the class l, i.e the mean vectors of each variable in the class l.

– W be the intraclass covariance matrix supposed identical in each class, of size $p \times p$, symmetrical and regular.

Let a new instance x (1 line, p columns). To be able to affect this new instance, the posterior probability should be maximized, which amounts to minimize the following quantity :

$$||x - \overline{X_j}||^2_{W^{-1}} = {}^t(x - \overline{X_j})W^{-1}(x - \overline{X_j}). \tag{1}$$

This method is reliable only if the conditions of measurements are invariant i.e. if the data are observed under the same conditions during and after the learning phase which is not always warranted. Often, the conditions of measurement depend of significant factors of variability, and unknown trends can appear in experimental conditions. In this case the learning does not well succeed in determining the class of an new instance, thus the learning is partial.

3 Conditional discriminant analysis

To overcome the problem of the trend factor, we are interested in a suitably modified analysis which completes the initial learning by a progressive control of the conditions of use. In many real situations the training set and other data are not observed under the same conditions. In these cases, the measurements taken on the instances depend on factors of trends which would be worth considering. The idea is to add to each instance X, the instance of a random vector Y, representative of the trend due to the experimental conditions. Moreover, one descriptor is not robust enough, therefore it can not cover every types of noises. Adding a vector to each instance will enable the method to be more robust to noises.

This approach, called conditional discriminant analysis, was first proposed by A. Baccini [Baccini et al., 2001] in a different domain, to classical statistical units.

3.1 Definitions and notations

Let us suppose that the estimates of $\overline{X_j}, j \in \{1, 2, ..., s\}$ and of W were made beforehand. Let Y be a matrix with n lines and h dimensions which takes into account the phenomenon of trend. We consider the following assumptions:

1. The intraclass mean of Y is $\overline{Y}$ (the empirical mean), whatever the class j.
2. The variance of $Z = {}^t(X, Y)$ is written :

$$W_Z = \begin{bmatrix} W_X & W_{XY} \\ W_{YX} & W_Y \end{bmatrix}$$

where

$W_X = W$ is the empirical intraclass covariance matrix of X,
$W_{XY} = {}^tW_{YX}$ is the covariance matrix of X and Y. W_Y is the covariance matrix of Y and supposed regular.

3. The intraclass variance of Z is assumed identical for each class and follows a normal law.

To be able to consider the factors of trend, one will carry out a decisional discriminant analysis no longer on X, but on Z, a vector with $p+h$ dimensions. Thus, the resulting changes are :

1. The matrix W is replaced by the matrix W_Z,
2. $\overline{X_j}$ are replaced by $\mu_j = {}^t(\overline{X_j}, \overline{Y})$.

Let us $C = W_X - W_{XY} W_{Y^{-1}} W_{YX}$ supposed regular. According to the principle of the usual decisional discriminant analysis, one must assign an new instance ${}^t(x, y)$ to the class j which minimizes the quantity:

$$\begin{bmatrix} {}^t(x - \mu_j) \\ {}^t(y - \mu_j) \end{bmatrix} W_{Z^{-1}} [(x - \mu_j)(y - \mu_j)]$$

what is equivalent to minimize the expression :

$$||(x - \overline{X_j}) - W_{XY} W_Y^{-1}(y - \overline{Y})||^2_{C_{-1}} \tag{2}$$

In the metric W^{-1}, the new instance ${}^t(x, y)$ is assigned to the class j for which the following expression is minimal [Baccini et al., 2001]:

$$\begin{gathered} ||(x - \overline{X_j}) - W_{XY} W_Y^{-1}(y - \overline{Y})||^2_{C_{-1}} = \\ {}^t((X - \overline{X_j}) - W_{XY} W_Y^{-1}(y - \overline{Y})) \\ (W_X - W_{XY} W_Y^{-1} W_{YX})^{-1}((X - \overline{X_j}) - W_{XY} W_Y^{-1}(y - \overline{Y})) \end{gathered} \tag{3}$$

The significant point in this analysis is the correction of x. However, the replacement of W^{-1} by C^{-1} improves theoretically the analysis.

In fact, the traditional discriminant analysis is based on the diagonalization of $B_X W_X^{-1}$ where B_X is the matrix of covariances between groups and W_X the matrix of covariances within groups. The conditional discriminant analysis is based on the diagonalization of $B_Z W_Z^{-1}$ because we apply traditional discriminant analysis on Z.

3.2 Parameter estimation

It is assumed that the estimates of the parameters W_X, $\overline{X_j}$ and $\overline{\overline{X}}$ are made beforehand from the training set.

We have :

$$W_{XY}^N = \mathbb{E}[(X - \overline{X_j})^t (Y - \overline{Y})] = \mathbb{E}[X^t (Y - \overline{Y})]$$

$$= \frac{1}{N} \sum_{i=1}^{N} X_i^{\,t} (Y_i - \overline{Y}^N), \tag{4}$$

where:

- N is the number of data for which Y is available. We remind you of Y is measured on the training set and the test set.
- $\overline{Y}^N$ is the empirical mean of Y_i.
- W_Y is defined by the empirical covariance matrix of Y on the whole instances where this variable is available.
- Using assumption 2 (§3.1) W_{XY} is computed by means of the whole data, even if we do not know the class of the new instance X.

Each time a new instance is added to the learning process, the formula (4) must be reconsidered for all the previous data. In order to decrease the complexity, we define the following reccurence formula:

$$W_{XY}^{N+1} = \frac{N}{N+1}(W_{XY}^N + X_{N+1}(Y_{N+1} - \overline{Y}^N)).$$

We give in appendix some indications of the demonstration for formula (4).

4 Choice of the parameters

Now that we have explained the principle of the conditional discriminant analysis, we have to choose the most fitted parameters to the symbol recognition problem.

4.1 Vector X

In order to make the discriminant analysis under the best possible conditions, it is necessary to start by choosing relevant variables. In our context, these variables can be obtained by using one or more descriptors which allow us to extract the quantitative variables from the symbols.

To generate relevant variables, which allow for a good discrimination of the data, we should choose robust descriptors to the noise, the deformations, and if possible having invariance properties to some geometrical transformations. Indeed, these invariance properties will correctly classify the same symbol independently of its position and its size in the graphic documents.

For practical reasons, we have chosen the descriptor called R-transform and defined in [Tabbone et al., 2006], but other descriptors can be used in the same way. This descriptor is based on the Radon transform, which is the projection of an image in a particular plan. This projection has interesting geometric properties which make it a good descriptor. According to these geometric properties, a signature of the transform is created. This signature keeps the properties of invariance to some geometric transformations, such as the translation and the scaling (after normalization). In addition the rotation invariance is restored by a cyclic permutation of the signature or directly from its Fourier transform. The Table 1 shows examples of the signatures of a symbol which is scaled and rotated. Thus, for our discriminant analysis, for each symbol of the training set and test set, we compute its signature.

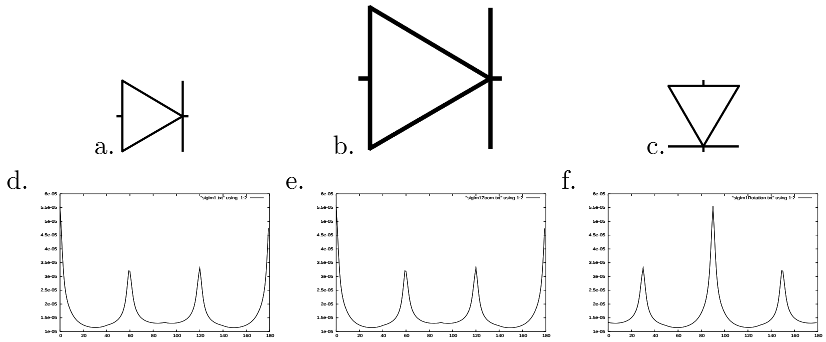

Table 1: Examples of signatures. a), b) and c) are respectively a perfect symbol, the same symbol with a zoom ×2 and the perfect symbol turned of 90 degree. d), e) and f) are their respective signatures.

4.2 Conditional vector *Y*

The vector Y constitutes the key point of the method. The choice of its components is essential for the success of the approach, and can be made in two ways [Baccini et al., 2001]:

1. One can carry out an 'external' choice in considering measurements of one or more indicators independent from the symbol belonging class and most likely well correlated with the parasitic effects of one possible trend.
2. One can also try to make an 'internal' choice in suitably analyzing the data (during and afterward learning) in order to discover possible measurement combinations of X which seem most characteristic of one possible trend while being independent from the group.

These two types of construction of Y are not excluded, and some components of Y could be obtained in the first way and the others with the second.

Whatever the selected type of construction, components must satisfy these two constraints, otherwise the analysis success would be compromised. Moreover the selected variables must be:

– Representative of parasitic effects of the analysis, factors of trends.

– Independent from the class of the considered symbol.

We have chosen the first type (1.). Therefore we need to find and build the vector Y which measures the parasitic effects of the analysis which are not taken into account in the selected descriptors for X. For these reasons, we minimize the following residual of the linear regression between the signature of the unknown symbol X and the signature of each model M representative of the different classes:

$$\sum (X_i - \beta_0 - \beta_1 M_i)^2$$

Thus, the selected Y describes the minimal residual which is supposed to represent the additional information related to the degradations that a symbol can undergo in real conditions.

5 Experimental results

We used symbols from the GREC database [Valveny and Dosch, 2004] for our tests. This database (see fig 1) was created especially for symbol recognition contest.

This database is mainly defined from two application domains, architecture and electronic, because these symbols are most largely used by graphic recognition teams and represent a great number of different forms. We have

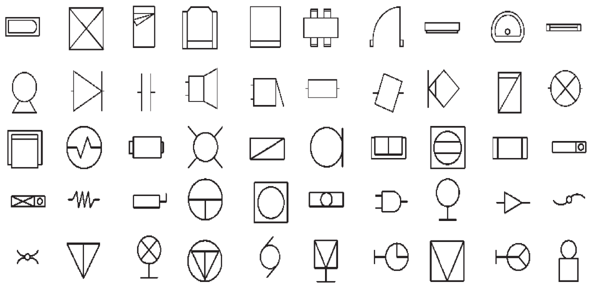

Figure 1: GREC database [Valveny and Dosch, 2004].

50 different symbol models for which we applied some noises based on Kanungo [Kanungo et al., 2000] model. These noises are similar to noise obtained when a document is scanned, printed or photocopied. Thus we apply 4 kinds of different noises, with different intensities, on each of the 50 models. Thus we get for each model 4 classes of noises. Each class of noises contains 100 noisy symbols but with different intensity. In this case we have a database with 20000 different symbols (4 classes of noises $\times 50$ models $\times 100$ intensities of degradation).

Next to simulate occlusions like dimensioning lines which often occur in graphic documents we add random lines to each symbol of the database. More precisely, we create two new sets, each one containing 10000 symbols (one with 200 different black lines $\times 50$ models, the other with 200 different white lines $\times 50$ models). In all we have a database composed of 40000 symbols. For example, the Table 2 presents different degradations applied on the same symbol model (TAB.2.a).

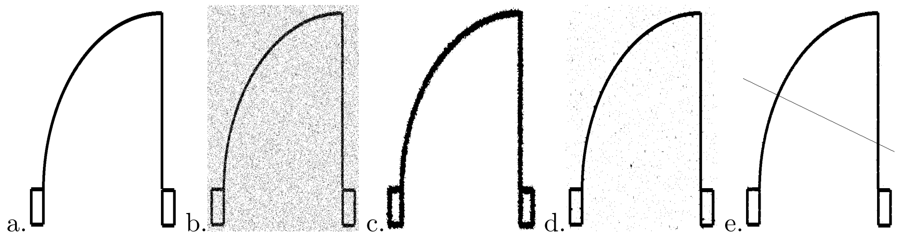

Table 2: Example of a symbol model a) on which we applied different degradations b), c), d) and e).

On this database we defined several tests of learning. We defined training sets composed of two classes of noise and the recognition process is applied to the other classes with different noises. For example a training set contains 5000 symbols : 100 degradations from the 50 models and the test set comprises 10000 symbols belonging to the other classes. Then, we calculated the recognition rate for this test set with the discriminant analysis (DA), the conditional discriminant analysis (CDA) and the conditional discriminant analysis with user feedback (CDAF). In the last case a user gives to the system his/her opinion (correctly or badly classified) at different moments (here every ten symbols) of the recognition process: the training set is updated according to the user opinion. This interactive procedure increases the recognition rate. The Table 3 shows the results obtained with these tests. We can notice that from the beginning of the learning and until the half, the DA and the CDA have a similar behavior: the recognition rate decreases gradually from 92% to 74% for the DA and from 85% to 74% for the CDA. However the DA gives rise to slightly better results (a recognition rate approximately of 5 to 10 percent higher). Approximatively from 5000 tested symbols (which corresponds to the size of the training database), the recognition rate of the DA decreases until 7000 tested symbols and then slightly increases to reach at the end a recognition rate of 76%. On the contrary the recognition rate of the CDA increases gradually until the end of the learning to reach the value of 82% and is thus better than the DA in the second part of the learning. Using the user feedback makes the recognition rate increases gradually during all the learning from a recognition rate of 65% to 87%. Moreover, the results of the CDAF show that de CDA is better with feedback than without, because the feedback takes into account correct allocation only.

6 Conclusion and future works

In this paper we have proposed an original adaptation of a conditional discriminant analysis. The results show the robustness of the approach to the scale compared to the classical discriminant analysis. Our choice for the complementary variable Y has carried on the linear deviations between the perfect model and the symbol to be recognized. We see that this complementary variable takes into account the effects of trends related to the limits of a descriptor to a high number of various symbols and the method of discriminant analysis. Furthermore we have shown experimentally that the user feedback improves the learning. Future works will be dedicated to take into account other disturbances on symbols. We wish to consider nonlinear deviations in the determination of the variable Y and combine several descriptors together in the recognition process.

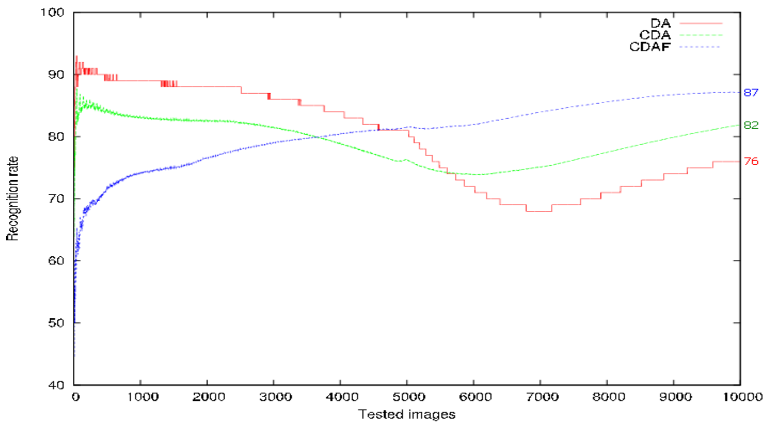

Table 3: Evolution of the recognition rate of the CDA and the CDAF compared to the DA with a learning on 5000 symbols (100 symbols by class, 50 classes in all) and tests on 10000 (200 symbols by class, 50 classes in all).

Acknowledgements

We would like to thank M. Dozzi, professor of Mathematics at the University of Nancy 2, and A. Ruiz-Gazen [Baccini et al., 2001] for their helpful comments on this work.

References

[Adam et al., 2000] Adam, S., Ogier, J. M., Cariou, C., Mullot, R., Labiche, J., and Gardes, J. (2000). Symbol and character recognition: application to engineering drawings. *International Journal on Document Analysis and Recognition*, 3(2):89–101.

[Baccini et al., 2001] Baccini, A., Caussinus, H., and Ruiz-Gazen, A. (2001). Apprentissage progressif en analyse discriminante. *Revue de Statistique Appliquée*, 49(4):87–99.

[Jain et al., 2000] Jain, A. K., Duin, R. P. W., and Mao, J. (2000). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37.

[Kanungo et al., 2000] Kanungo, T., Haralick, R. M., Baird, H. S., Stuezle, W., and Madigan, D. (2000). A Statistical, Nonparametric Methodology for Document Degradation Model Validation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1209–1223.

[Llados et al., 2001] Lladós, J., Martí, E., and Villanueva, J. J. (2001). Symbol Recognition by Error-Tolerant Subgraph Matching Between Region Adjancy Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1137–1143.

[Messmer and Bunke, 1996] Messmer, B. T. and Bunke, H. (1996). Automatic Learning and Recognition of Graphical Symbols in Engineering Drawings. In *Graphics Recognition Methods and Applications*, volume 1072 of *Lecture Notes in Computer Science*, pages 123–134, Springer Verlag.

[Tabbone et al., 2006] Tabbone, S., Wendling, L., and Salmon, J. P.(2006). A new shape descriptor defined on the Radon transform . *Computer Vision and Image Understanding*, 102(1):42–51.
[Valveny and Dosch, 2004] Valveny, E. and Dosch, P. (2004). Symbol Recognition Contest: A Synthesis. In *Graphics Recognition*, volume 3088 of *Lecture Notes in Computer Science*, pages 368–385, Springer Verlag.

Appendix

We give in this section some indications of the demonstration for the formula (4). First we calculate B_Z. By definition $Z = {}^t(X, Y)$ and :

$$B_Z = \frac{1}{g-1} \sum_{j=1}^{g} (\mu_{j,Z} - \mu_Z)^t (\mu_{j,Z} - \mu_Z),$$

where $\mu_{j,Z}$ is the mean vector of Z in the group j,
μ_Z the total mean vector:

$$\mu_{j,Z} = \begin{bmatrix} \mu_{j,X} \\ \mu_{j,Y} \end{bmatrix}$$

where $\mu_{j,X}$ is the mean vector of X in the group j and $\mu_{j,Y}$, the mean vector of Y in the group j. Thus $\mu_{j,Y} = \mu_Y$ since Y is independent from the group.

Thus we have

$$\forall j, \mu_{j,Y} = \mu_j$$

Then

$$\mu_{j,Z} - \mu_Z$$

$$= \begin{bmatrix} \mu_{j,X} - \mu_X \\ \mu_{j,Y} - \mu_Y \end{bmatrix}$$

$$= \begin{bmatrix} \mu_{j,X} - \mu_X \\ 0 \end{bmatrix}$$

Thus we can write

$$B_Z = \frac{1}{g-1} \sum_{j=1}^{g} \begin{bmatrix} \mu_{j,X} - \mu_X \\ 0 \end{bmatrix} ({}^t(\mu_{j,X} - \mu_X), 0)$$

$$B_Z = \begin{bmatrix} \frac{1}{g-1} \sum_{j=1}^{g} (\mu_{j,X} - \mu_X)^t (\mu_{j,X} - \mu_X) & 0 \\ 0 & 0 \end{bmatrix}$$

$=$

$$\begin{bmatrix} B_X & 0 \\ 0 & 0 \end{bmatrix}$$

Next we have to calculate W_Z^{-1} :

We have $W_Z^{-1} = \begin{bmatrix} W_X & W_{XY} \\ W_{YX} & W_Y \end{bmatrix}$

Thanks to the method of matrix inversion with blocks, we obtain:

$W_Z^{-1} = \begin{bmatrix} C^{-1} & -C^{-1}W_{XY}W_Y^{-1} \\ -W_Y^{-1}W_{YX}C^{-1} & (*) \end{bmatrix}$

where (*) is a term which is independent from the group j.

Thus $B_Z W_Z^{-1} = \begin{bmatrix} B_X C^{-1} & -B_X C^{-1} W_{XY} W_Y^{-1} \\ 0 & 0 \end{bmatrix}$

Then we allocate a new observation Z to the group j which minimizes this expression :

$$^t(z - \mu_Z)W_Z^{-1}(z - \mu_Z)$$

$$= (^t(x - \mu_{j,X}), {}^t(y - \mu_Y))W_Z^{-1} \begin{bmatrix} x - \mu_{j,X} \\ y - \mu Y \end{bmatrix}$$

$$= (^t(x - \mu_{j,X})C^{-1}$$

$$-{}^t(y - \mu_Y)W_Y^{-1}W_{YX}C^{-1}, -{}^t(x - \mu_{j,X})C^{-1}W_{XY}W_Y^{-1} + (**))$$

$$\begin{bmatrix} x - \mu_{j,X} \\ y - \mu Y \end{bmatrix}$$

$$= {}^t(x - \mu_{j,X})C^{-1}(x - \mu_{j,X})$$

$$-{}^t y - \mu_Y W_Y^{-1} W_{YX} C^{-1}(x - \mu_{j,X})$$

$$-{}^t(x - \mu_{j,X})C^{-1}W_{XY}W_Y^{-1}(y - \mu_Y) + (***)$$

where $(**)$ and $(***)$ are group-independent terms.

We have

$$^{t\,t}((y - \mu_Y)W_Y^{-1}W_{YX}C^{-1}(x - \mu_{j,X})) =$$

$$^t(x - \mu_{j,X})C^{-1}W_{XY}W_Y^{-1}(y - \mu_Y) =$$

$$^t(y - \mu_Y)W_Y^{-1}W_{YX}C^{-1}(x - \mu_{j,X})$$

because the two equality members are real numbers.

Finally

$$^t(z - \mu_Z)W_Z^{-1}(z - \mu_Z) =$$

$$^t(x - \mu_{j,X})C^{-1}(x - \mu_{j,X})$$

$$-2^t(x-\mu_{j,X})C^{-1}W_{XY}W_Y^{-1}(y-\mu_Y)+(***)$$
$$=||(x-\mu_{j,X})-W_{XY}W_Y^{-1}(y-\mu_Y)||^2_{C^{-1}}$$

where

$$C=W_X-W_{XY}W_Y^{-1}W_{YX}.$$