Supervised features selection in MaZda

Michal Strzelecki

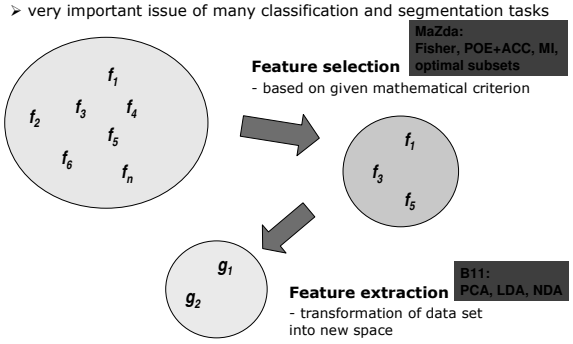
Institute of Electronics, Technical University of Lodz, Poland

Why feature reduction?

- it is not known *a priori* which features are best for given texture analysis - one has to consider as many features as possible,
- it is very difficult to manage with almost 300 features generated by **MaZda**,
- large number of features requires large number of data samples (which are not available normally).

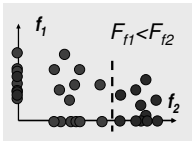
Feature reduction

- very important issue of many classification and segmentation tasks



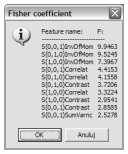
Fisher coefficient

$$F=\frac{D^2}{V^2}=\frac{\frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{K}P_k P_j|\mu_k-\mu_j|^2}{1-\sum_{k=1}^{K}P_k^2}\Big/\sum_{k=1}^{K}P_k V_k^2$$

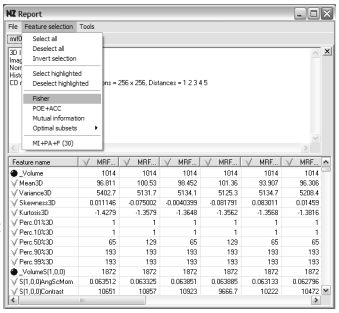


where:
D - between-class variance,
V – within-class variance,
P_k - probability of feature k,
V_k, μ_k - variance and mean value of feature k in given class.

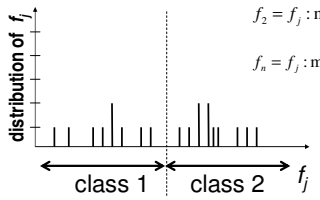
MaZda Report window
Feature selection->Fisher



10 features with the highest Fisher coefficient



Probability of Classification Error and Average Correlation Coefficient (POE+ ACC)



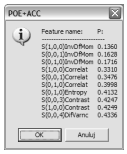
$$f_2 = f_j : \min_j[POE(f_j)+|Corr(f_1,f_j)|]$$

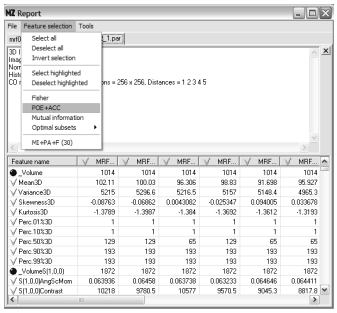
$$f_n = f_j : \min_j[POE(f_j)+\frac{1}{n-1}\sum_{k=1}^{n-1}|Corr(f_k,f_j)|]$$

$$POE(f_j)=\frac{\text{number of mislassified samples}}{\text{total number of samples}}=\frac{2}{20}$$

$$f_1 = f_j : \min_j[POE(f_j)]$$

MaZda Report window
Feature selection->POE+ACC





10 features with the lowest POE+ACC coefficient

Mutual Information (MI)

MI between two random variables *F* and *C* is estimated by:

$$I(F,C)=\sum_f\sum_c P(F,C)\log_2\frac{P(F,C)}{P(F)P(C)}$$

where the *F* and *C* are quantized.

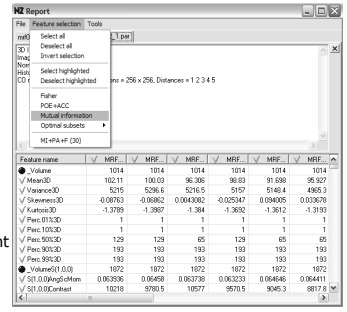
MI=0 when *F* (feature) and *C* (class category) are independent (incorrect classification for all f_i).

$$class(f_i)\neq c_i \quad \text{for all } i$$

Large value of MI means that given feature carries information about class membership (*F* and *C* are correlated).

$$class(f_i)= c_i \quad \text{for some } i$$

MaZda Report window
Feature selection->MI



10 features with the highest Mutual Information coefficient

Optimal subset feature selection
(brute force method)

In this method all feature subsets are analysed. Thus it is possible to find a subset which provides a minimal calssification error (using 1-NN clssifier).
The number of subsets is defined by:

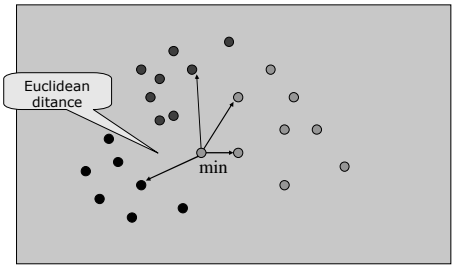
$$S_{Tot}=\sum_{i=1}^{n}\binom{n}{i}$$

where *n* – total nuber of features.
In MaZda 4.0 *n*=284, then S_{Tot} >1e70

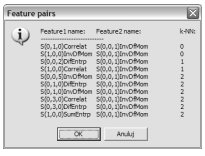
Thus evaluation of subsets with only 1, 2 and 3 features is implemented:

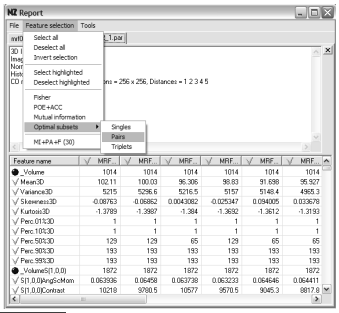
- single features: 284 subsets
- feature pairs: 40186 subsets
- feature triples: 3777484 subsets

1-Nearest Neighbour Classifier, 1-NN)



MaZda Report window
Feature selection->Optimal subsets->Pairs



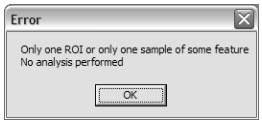


10 feature pairs with the lowest 1-NN classification error

Time consuming (pairs, triples) !!!

Automatic feature selection
(Fisher, POE+ACC, MI, Optimal subsets)

- it is necessary to define at least <u>two classes</u> (two ROIs wit different names),
- each class must contain at least <u>two samples</u>



It is also possible to perform manual feature selection (by selecting/deselecting features in Report window)

13

Analysis examples:

Feature selection:

- four methods were used to obtain best features
- analysis of feature subsets was performed

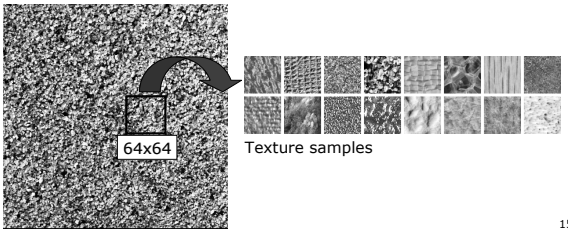
Classification (1NN - one nearest neighbor):

- raw data
- data after linear discriminant analysis (LDA)

14

Experimental Data #1

- 16 textures from Brodatz album, 512x512 pixels
- each image was divided into 64 squares (64x64)
- 1024 samples of 16 texture classes altogether
- for each sample 269 texture features have been calculated using MaZda software (±3σ norm.)
  (gradient matrix, co-occurrence matrix, RL matrix, wavelet, AR model)



15

1-NN classification results for 4 features selected using different methods

optimal subset	MI		Fisher		POE	
raw	raw	LDA	raw	LDA	raw	LDA
0	**16**	**8**	**79**	**59**	**145**	**147**
Number of misclassiedfied samples (total no. of samples: 1024)						

16

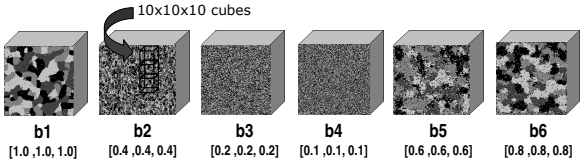
<u>Optimal feature subset:</u>
Sigma, MinNorm, GrMean, S(0,2)Correlat
1-NN error: 0

<u>MI:</u>	<u>Fisher:</u>	<u>POE:</u>
Sigma	S(0,2)DifEntrp	S(0,2)DifEntrp
MinNorm	S(0,3)DifEntrp	Teta4
S(1,0)Contrast	S(0,1)DifEntrp	S(0,5)SumVarnc
Perc.10%	Sigma	S(5,0)SumVarnc
S(0,1)DifVarnc	S(0,2)Contrast	MaxNorm
S(0,2)DifVarnc	Variance	WavEnHH_s-2
GrMean	S(0,4)DifEntrp	S(0,2)AngScMom
S(0,2)Correlat	Perc.10%	Teta3
S(0,3)Contrast	MinNorm	GrNonZeros
S(1,0)Correlat	S(0,2)Correlat	Perc.10%
1-NN error: 0	**1-NN error: 0**	1-NN error: 26
		1-NN error after LDA: 23

17

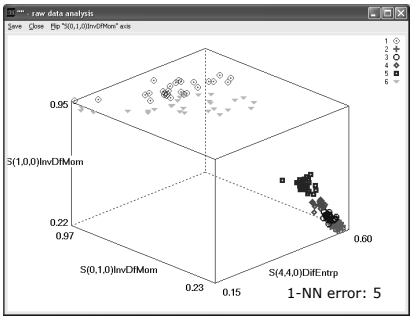
Experimental Data #2

- 6 3D textures generated using MRF model, 130x130x130 pixels
- each image was divided into 27 squares (10x10x10)
- 162 samples of 6 texture classes altogether
- for each sample 284 texture features have been calculated using MaZda software
  (histogram, 3D co-occurrence matrix)



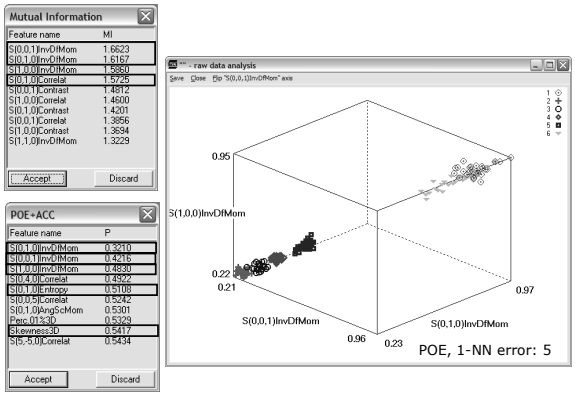
18

Optimal subset with 3 features:
S(1,0,0)InvDfMom, S(0,1,0)InvDfMom,
S(4,4,0)DifEntrp



1-NN error: 5

19



POE, 1-NN error: 5

10 features, 1-NN error: 39 (raw), 7 (LDA)
5 selected features, 1-NN error: 5 (raw), 9 (LDA)

20

Conclusions

- None of the three feature selection methods was able to select optimal subset of features. However some of them found features with the same classification error.
- Quality of feature selection method depends on classification task (data #1 – MI; data #2 – POE).
- Feature subsets may provide better classifiction than the whole set (in case of Fisher, POE, MI) using 1-NN classifier.

21

Exercise 1

- Start MaZda or, if it is already started, close all the report tab-pages
- Load image *texture1.bmp* from *Tutorials\3a_Texture_analysis_1* folder
- Switch to the overlapping mode and create 16 regions of interest of an approximate size 35x35 pixels (observe the status bar for a size information) with *Draw rectangle* tool
- Set analysis options and run the analysis
- In a Report window set class name *Biology* for all the columns
- Load image *texture2.bmp* from *Tutorials\3a_Texture_analysis_1* folder
- Run the analysis (use the same regions and analysis options as previously)
- In a Report window set class name *Mineral* for all the columns of the second tab-page
- Find features that may be useful for discrimination of the two classes

22

Exercise 1 (continued)

- Perform automatic feature selection with Fisher criterion using ***Feature selection->Fisher***
- Save selected features to a file *fisher.sel* using ***File->Save selected***
- Run **b11** using ***Tools->B11 analysis***
- Observe **b11** options window selecting ***Options***
- Select only three top features in Feature selection box
- Observe feature distribution using ***Analysis->Raw data***
- Perform PCA and LDA analysis and observe feature distribution in the new data space

23