Mathematical modelling in Biology

Vadim V. Strijov
strijov@ccas.ru, http://strijov.com

Computing Centre of the Russian Academy of Sciences

In the Seminar Series on Immunopathology:
«Is there a role for Immune Markers and Immune Therapy in disease management?»

31st October 2007, Paris

Outline

1. **Problem statement**
   - Data description
   - Formal problem statement
   - One-dimensional statistics
2. **Classification algorithms**
   - Decision trees
   - Linear classifier
   - Multiple classification
3. **Quality of classification**
   - Result information
   - Cross-validation algorithm
   - Sample size estimations

Problem statement

Consider a problem of classification.

1. We have two groups of patients.
2. Each group is labelled for a class.
3. Each patient is described by a set of markers.

We have to find a model which is based on measured data that shows the groups are significantly different.

Data description

Consider a problem of classification based on the Cardio Immune Data.

Classes $\longrightarrow$ Groups of patients	The patients have classification labels "A1" and"A3".
Objects $\longrightarrow$ Patients	We have measured data for 14 patients in the group "A1" and 17 patients in the group "A3".
Features $\longrightarrow$ Markers	We have 20 markers: K, L, K/M, L/M, K/N, K/O, L/O, K/P, L/P, K/Q, K/R, L/R, L/R/SA, L/T/SA, L/T/SO, U/V, U/W, U/X, U/Y, U/Z

Patients–Markers table, an example

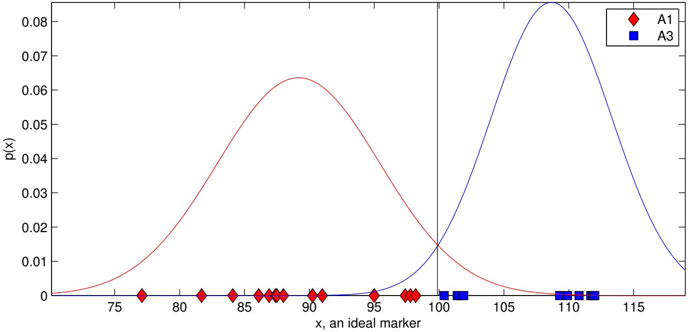
Class	Patient	K	L	K/M	L/M	
A1	C001	58.3	16.7	0.52	0.00	
A1	C004	40.2	6.0	NaN	NaN	
A1	C005	54.3	13.1	NaN	NaN	
A1	C008	48.7	9.8	0.05	0.02	etc.
A3	023	46.6	21.2	0.40	0.08	
A3	026	50.7	26.2	0.12	0.00	
A3	027	45.3	24.5	0.05	0.02	
A3	D037	46.3	13.1	1.23	0.13	
				etc.		

Can we show that the groups are significantly different?

Formal problem statement

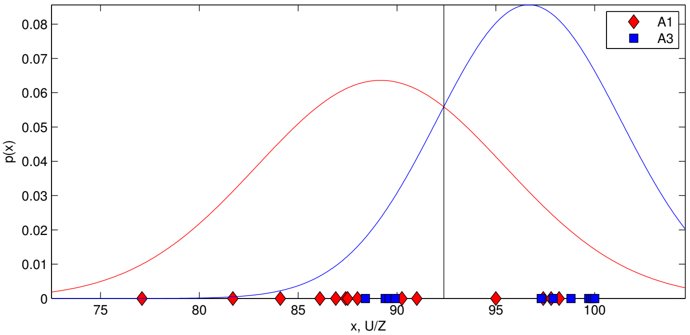
Formally, the problem can be stated as follows: given training data $\{(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_n}, y_n)\}$ produce a classifier $A : X \rightarrow Y$ which maps an object $\mathbf{x} \in X$ to its classification label $y \in Y$.
Here the vector $\mathbf{x_i}$ contains markers for i-th patient, the label y_i contains a class label for i-th patient, and the classifier A is a model. We should find it out to prove our hypothesis on the group difference.

One-dimensional statistics



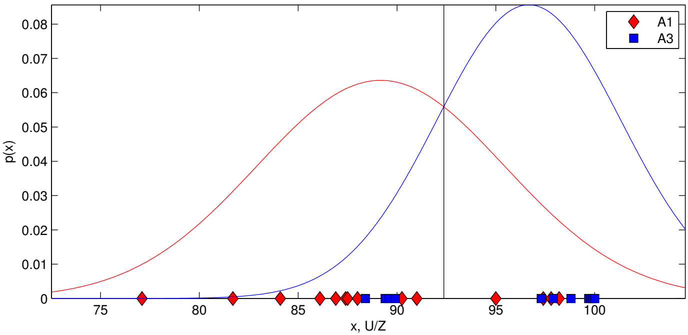
Assume that two groups are separate if they accept the null-hypothesis in one of the following tests: Student's t-test, Welch's t-test or Mann-Whitney's U test.

Real data



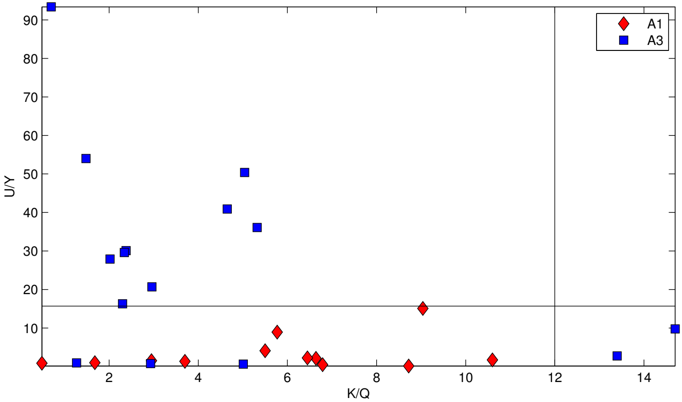
- ✓ It is very simple to visualize one-dimensional data.
- ✓ One-dimensional statistics is well-developed and recognized.
- × And ...

Real data



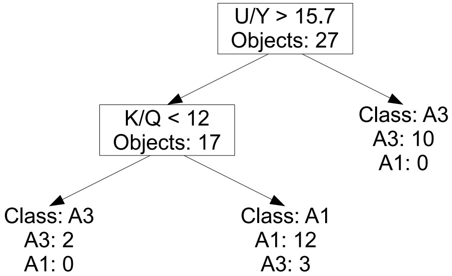
- ✓ It is very simple to visualize one-dimensional data.
- ✓ One-dimensional statistics is well-developed and recognized.
- × And give poor results if one deals with a complex problem.

A classification rule



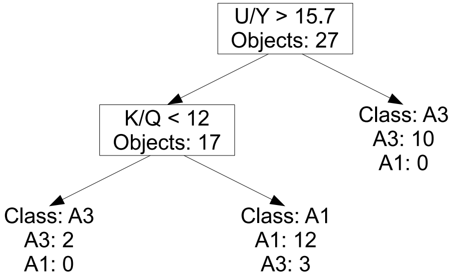
if U/Y > 15.7 then A3 else (if K/Q < 12 then A1 else A3)

Decision tree



if $U/Y > 15.7$ then A3 else (if $K/Q < 12$ then A1 else A3)

Decision tree



if $U/Y > 15.7$ then A3 else (if $K/Q < 12$ then A1 else A3)

- ✓ Different subsets of markers produce trees of different quality.
- ✓ One can use several trees to make a voting algorithm.

Decision forest and voting algorithms

1. If $U/Y < 15.7$ then A1 else A3
2. If $U/Z < 88.2$ then A1 else (if $U/V < 51.9$ then A1 else A3)
3. If $U/Z < 88.2$ then A1 else (if $K/N < 31.9$ then A3 else A1)

Class	Patient	Rule 1	Rule 2	Rule 3	Vote
A1	C014	A1	~~A3~~	A1	A1
	C015	NaN	A1	A1	A1
	D034	A1	A1	~~A3~~	A1
	L107	A1	NaN	~~A3~~	NaN
	etc.	…	…	…	…
A3	023	A3	A3	A3	A3
	026	~~A1~~	A3	A3	A3
	027	A3	NaN	NaN	NaN
	009	~~A1~~	A3	A3	A3
	etc.	…	…	…	…

Profs and coins

Decision tree

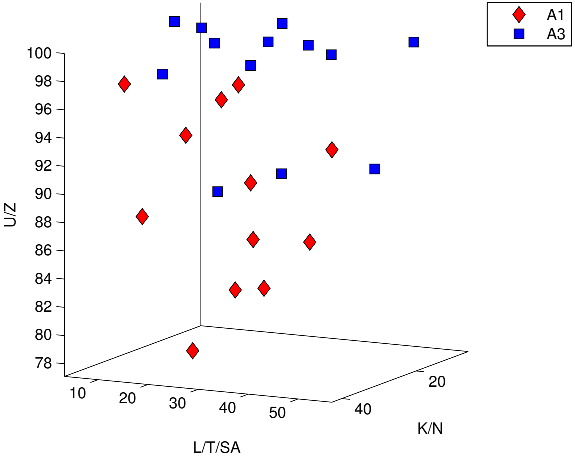
- ✓ uses several markers for classification; so, it makes less mistakes,
- ✓ helps to select the most informative markers,
- ✓ can be used in a voting algorithm,
- × assumes the markers do not depend on each other.

Linearity

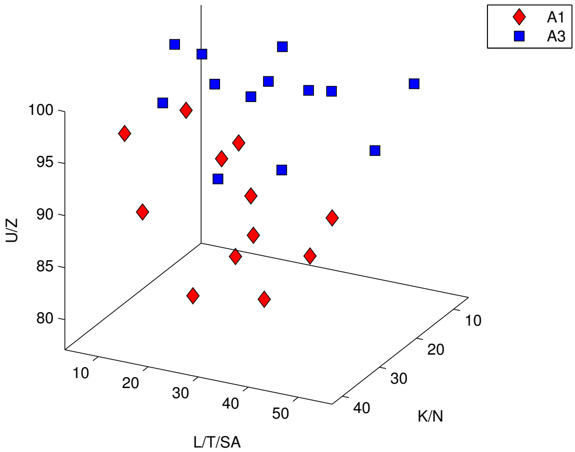
However in this problem we are using the markers do depend on each other. For example, three markers K, N and K/N are linear dependent, since we assume

$$K/N = \alpha K + (1-\alpha)N.$$

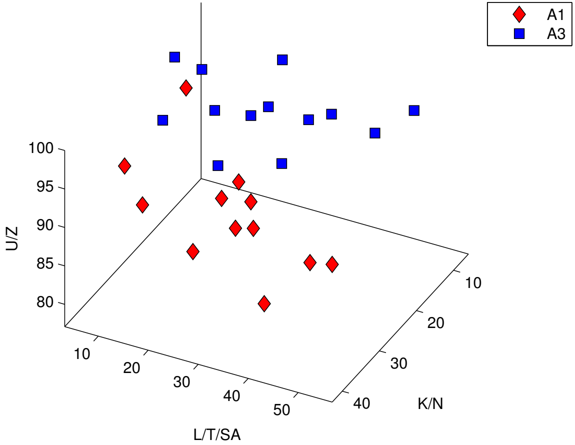
Linear classifier



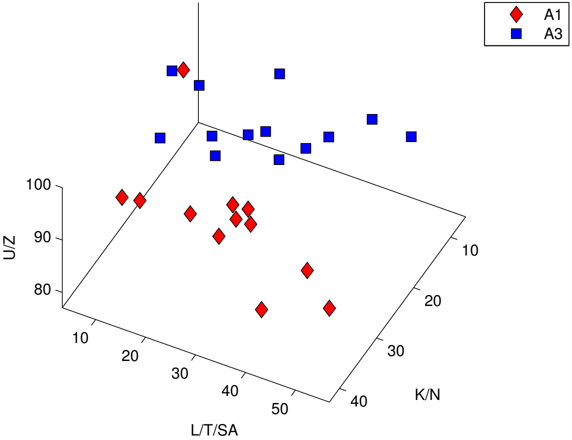
Linear classifier



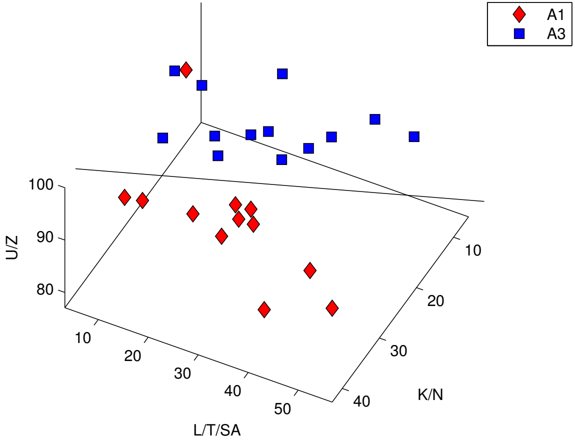
Linear classifier



Linear classifier



Linear classifier



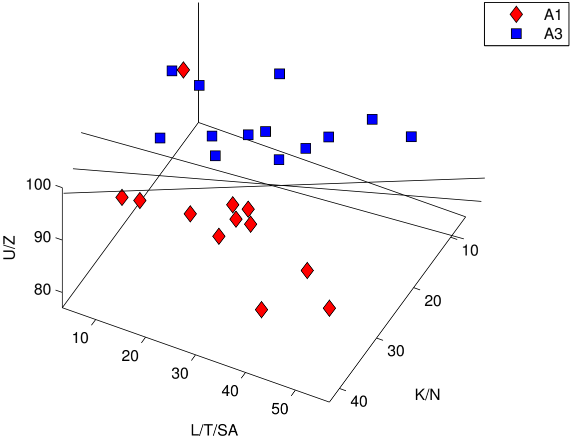
Separation hyperplane

The equation
$$\mathbf{w}^T\mathbf{x} = b$$
describes a separation hyperplane in the marker space.
Let $\mathbf{x}_i$ is a vector of patient's markers and $\mathbf{w}$ are parameters. Then
$$y_i = \operatorname{sign}\left(\sum_j w_j x_{ij} - b\right) = \operatorname{sign}(\mathbf{w}^T\mathbf{x}_i - b)$$
is the class of the i-th patient.

Linear classifier

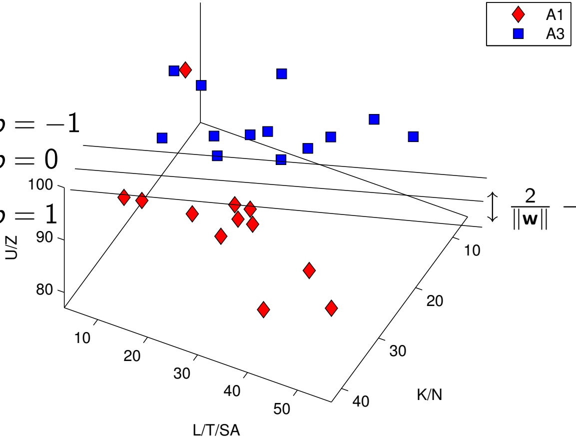
- ✓ assumes the markers depend on each other,
- ✓ copes with a big number of markers,
- ✓ could be easily interpreted by experts,
- ⋙ and can be powered …

Where is the proper separation hyperplane?



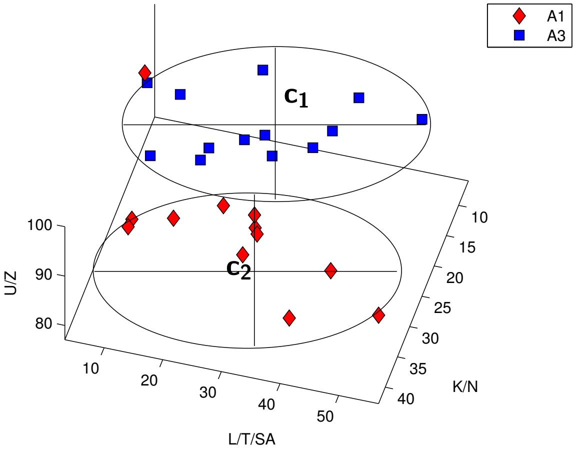
Optimal separation hyperplane

$$\mathbf{w}^T\mathbf{x} - b = -1$$
$$\mathbf{w}^T\mathbf{x} - b = 0$$
$$\mathbf{w}^T\mathbf{x} - b = 1$$



$$\frac{2}{\|\mathbf{w}\|} \longrightarrow \max$$

Support vector machine (radial basis kernel)



$$y_i = \mathrm{sign}(\mathbf{w}^T K(\mathbf{x}_i, \mathbf{c}) - b), \quad \text{where} \quad K(\mathbf{x}_i, \mathbf{c}) \text{ is a Kernel function}$$

Each algorithm implies a hypothesis:

- ✓ markers do not depend on each other $\longrightarrow$ decision trees,
- ✓ objects can be separated by hyperplane $\longrightarrow$ linear classifier,
- ✓ objects can be mapped into a separable space $\longrightarrow$ support vector machines,
- ✓ classes are compact $\longrightarrow$ radial basis functions,
- ✓ marker space are complex $\longrightarrow$ voting algorithms.

To interpret the results

we use the parametric algorithms. They are based on a mathematical model and a set of parameters.

Multiple classification

Now we have 4 groups: A1, A3, B1 and B2.
There are two ways:

1. "one class versus the others", to classify a new patient,
2. "each class versus another", to discover differences between classes.

Multiple classification

Now we have 4 groups: A1, A3, B1 and B2.

Our goal is to differ one group from another.

- ✓ First, run classification algorithms for each pair of classes.
- ✓ Then, obtain classification results and separation hyperplanes.

Together with hyperplanes we have the most informative markers

Pairs	Markers		
A1 vs. A3	L	K/N	U/Z
A3 vs. B1	K/N	K/P	U/W
A3 vs. B2	L	L/P	U/Y

and the marker show that the hyperplanes are placed in the different subspaces for each pair.

Result information

After classification a pair (A1 vs. A3) we obtain the following information:

1. covered patients

$$(\text{C014}, \text{D034}, \text{L107}, \ldots, \cancel{\text{C008}}, 023, 026, \ldots, \text{C015}),$$

2. classification error

$$\frac{|a1|}{|A1|} + \frac{|a3|}{|A3|},$$

3. most important markers

$$(\text{L}, \text{K/N}, \text{U/Z}),$$

4. parameters of the algorithm

$$y_i = \text{sign}(\mathbf{w}^T \mathbf{x}_i - b) = \text{sign}([0.35, 0.72, 0.29]^T \mathbf{x}_i - 34.16).$$

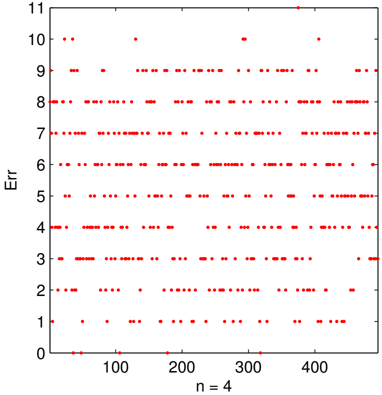
Classification quality: cross-validation algorithm

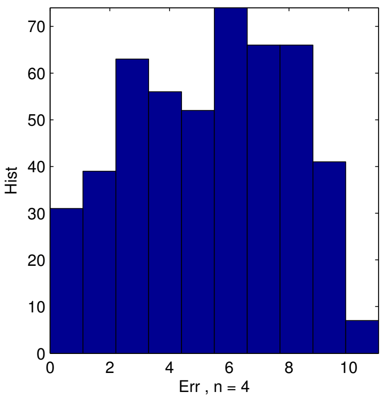
Assume a patient falls in a group by occasion. What will change if we remove several patients from groups?

The concept:

1. Split up the whole set of objects into two subsets:
   1) training,
   2) test.
2. Training set defines the parameters of the algorithm.
3. Test set defines the number of errors.

Robustness of the classification

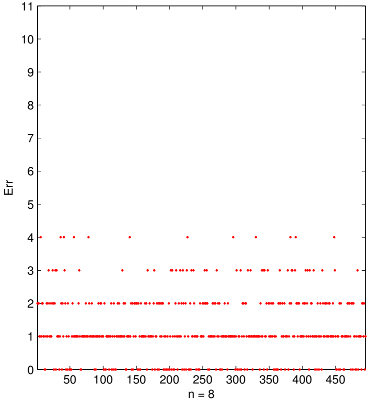


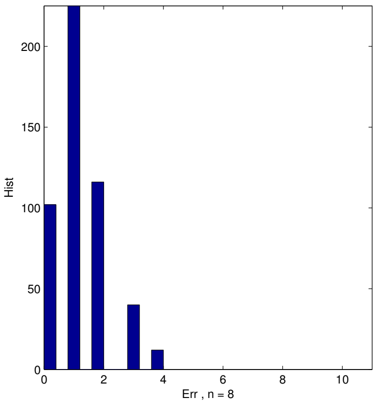


X-axis: experiment number
Y-axis: error, misclassified patients

X-axis: error
Y-axis: histogram of given error

Robustness of the classification

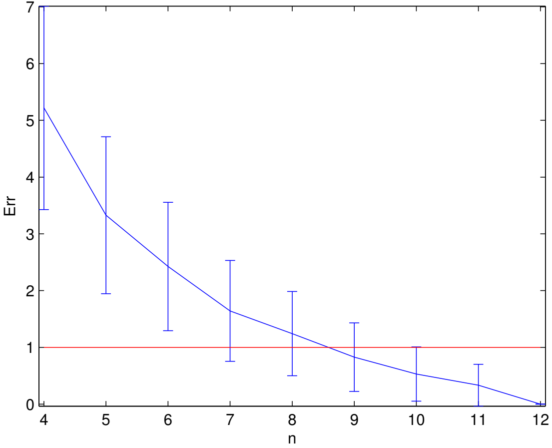




X-axis: experiment number
Y-axis: error, misclassified patients

X-axis: error
Y-axis: histogram of given error

Sample size estimations



X-axis: number of patients in a training set in each group
Y-axis: expected number of errors and the confidence interval

Conclusion: creation a mathematical model from data

- We have to use markers together to investigate a difference between groups.
- The ideal classification algorithm is simple and gives minimal error.
- We can assess the algorithm quality and the minimal sample size.

There are plenty classification algorithms to choose from!