Partial Least Squares – Discriminant Analysis
PLS-DA

Prof. Dmitry Kirsanov
Applied Chemometrics Laboratory
Institute of Chemistry
St. Petersburg State University

d.kirsanov@gmail.com

Calibration and classification

Chemometrics

Quantitative tasks

Numerical prediction of various system properties

PCR, PLS, etc

Qualitative tasks

Exploratory analysis
Classification
Recognition

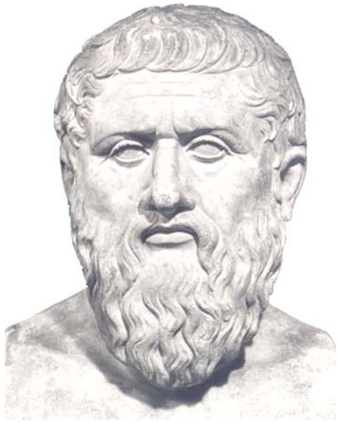
PCA, SIMCA, etc

What is classification

- **Classification**

  separation of *objects* into *classes*

  according to analogies in their *properties*



Plato
(V-IV age BC)

- ***Objects*** **– whatever (material objects, processes, texts, etc.)**

- ***Properties*** **– qualitative or quantitative characteristics of the objects (variables in PCA)**

- ***Classes*** **– variety of objects with similar properties**

Classification algorithms

Unsupervised

No *a priori* knowledge if there are some groups of similar samples

The main mechanism – search for similarity in parameters of the samples

The main goal – to understand if there are some groups of samples and what is the reason (variables) responsible for grouping

Supervised

There is *a priori* knowledge on the grouping in the initial set of samples

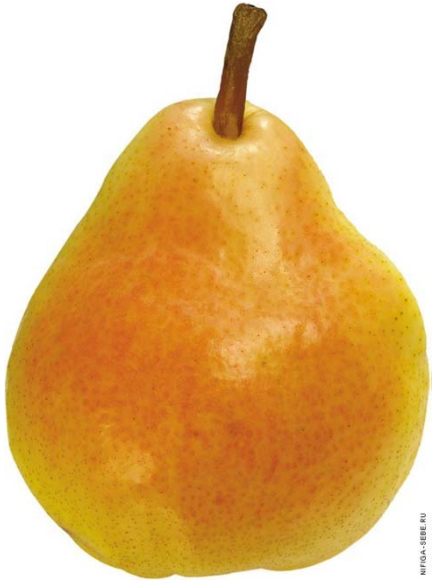
The main mechanism – building the model relating the sample parameters with class belonging

The main goal – classification of the new samples

What is classification?

Quite often one parameter is not enough to define the class

Parameters and groups can be of any type

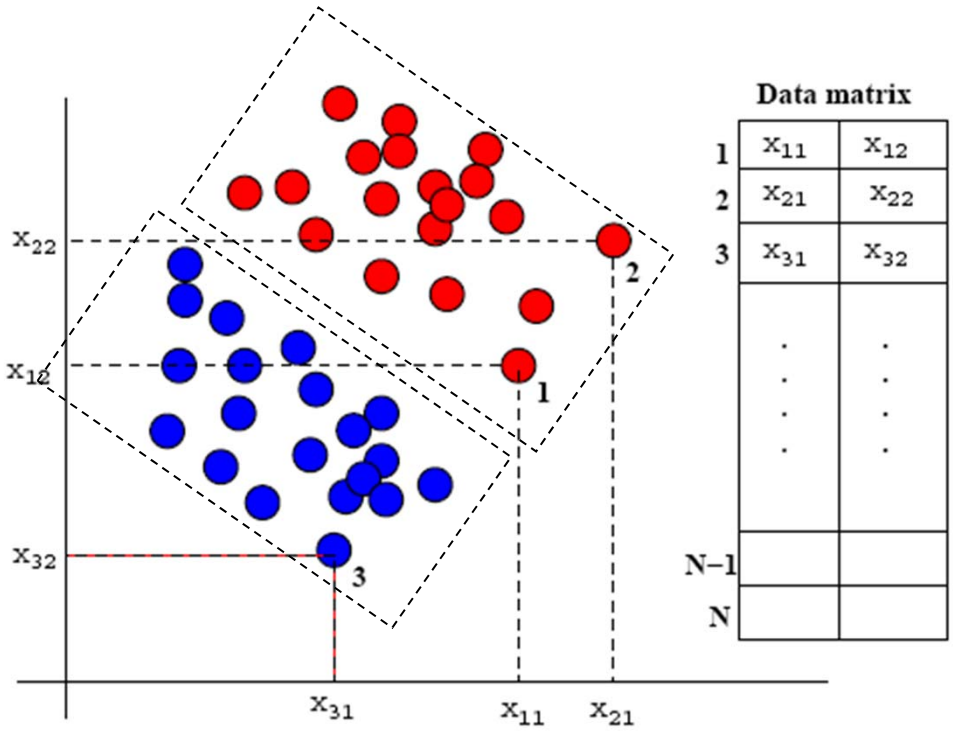


Geometrical interpretation

Vector of parameters – variables making N-dimensional coordinate system (N – number of parameters)

Samples – points in the parameter space

Groups, or classes – limited subspaces in the parameters space



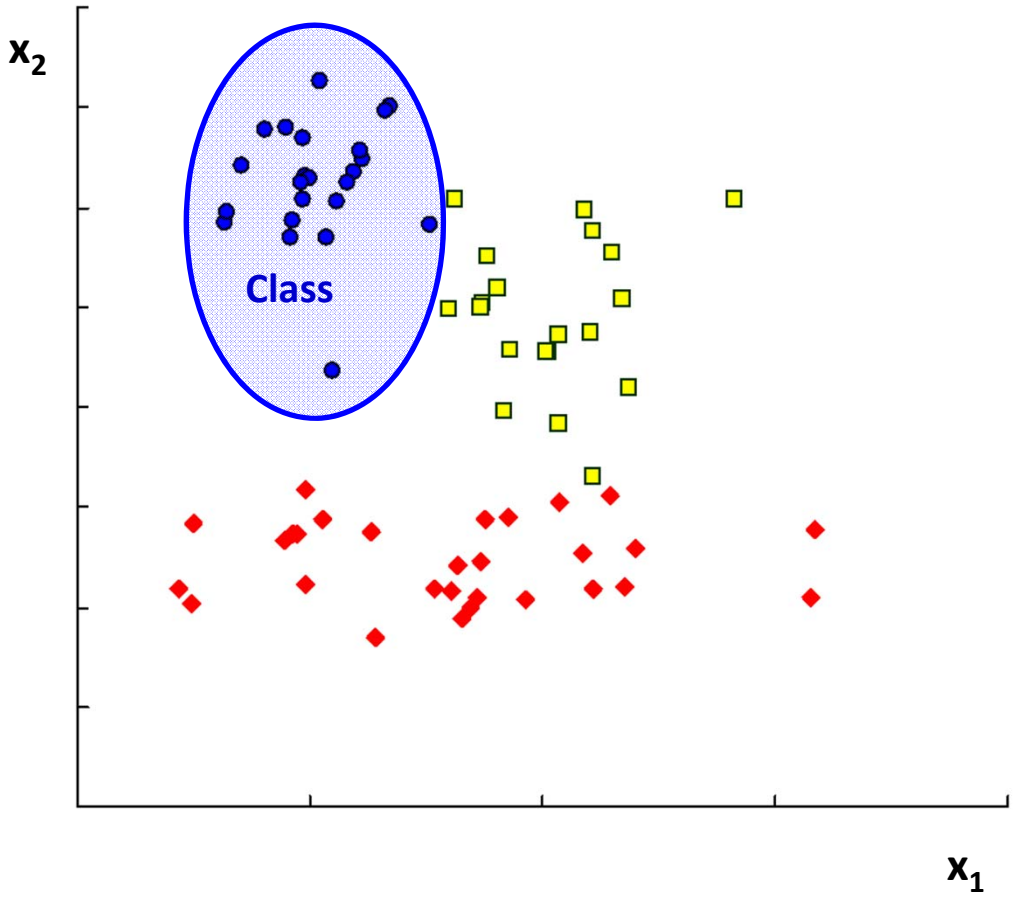
Distance

- **Distance (distance function) – quantitative measure of similarity of the objects.**
- **The smaller the distance, the more similar the objects are.**
- **Distance should have certain properties:**

$$D \geq 0; \quad D(X,X) = 0; \quad D(X,Y)=D(Y,X); \quad D(A,B) \leq D(A,C)+D(C,B)$$

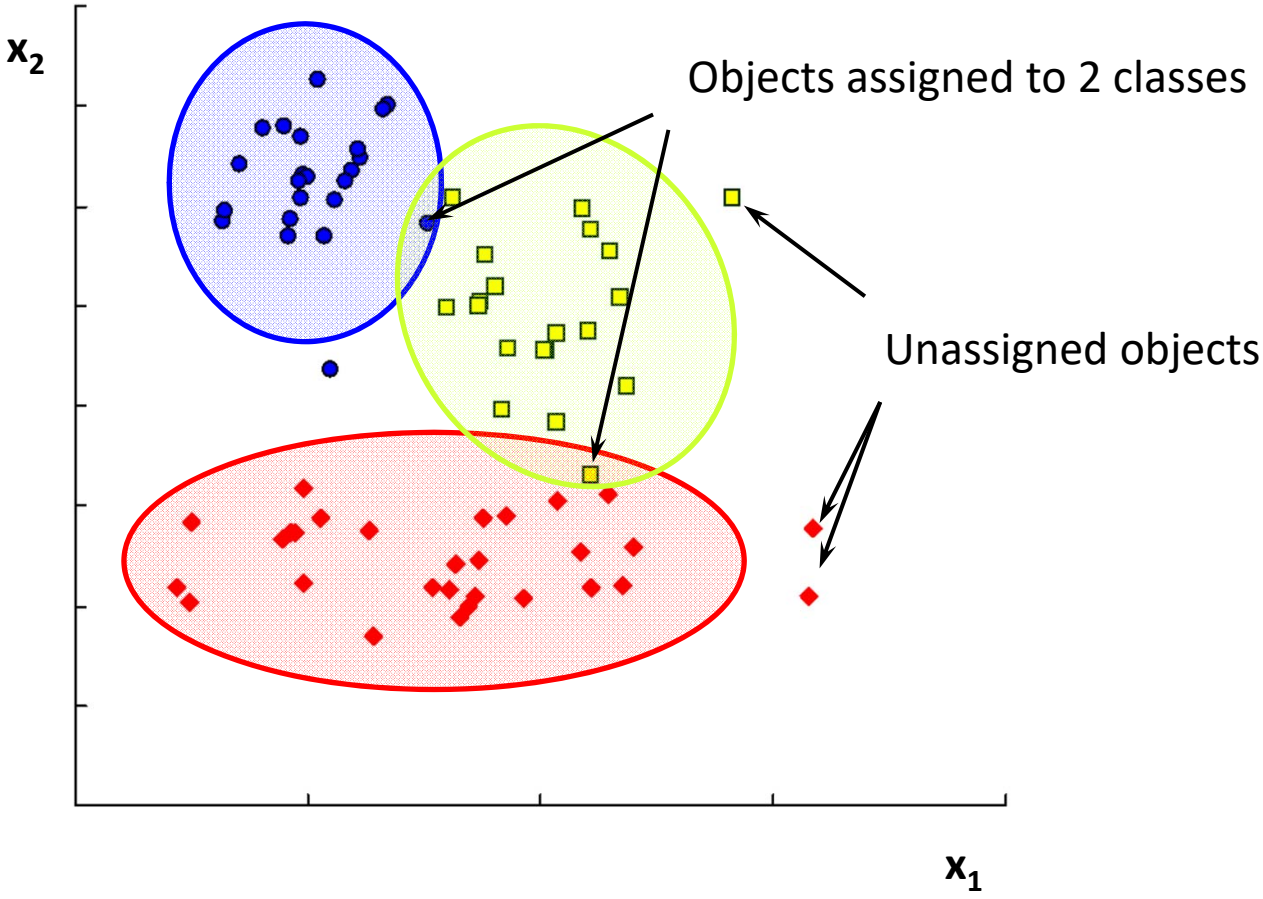
- **Distance is calculated using the properties of the objects**
  - **Changing the calculation way changes the classification results**

One-class classifier



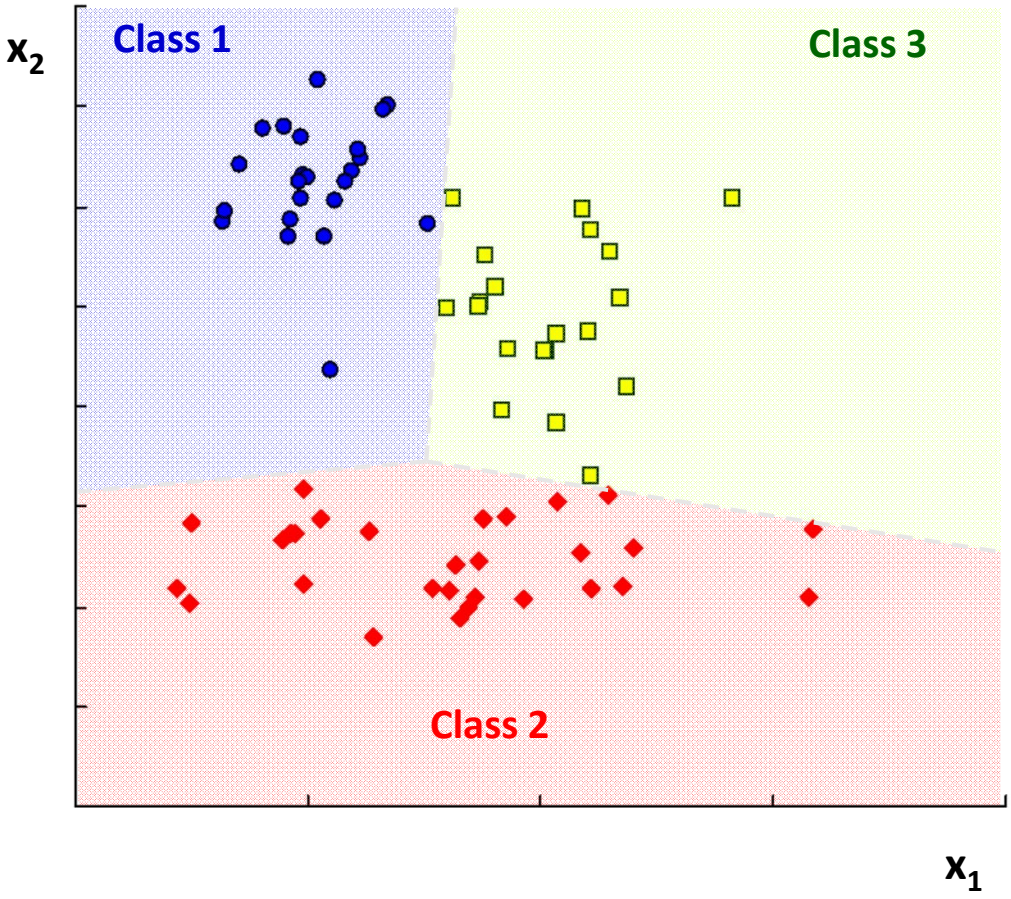
One-class classifier decides if the object belongs to one particular class. The number of other classes does not matter.

Multiclass classifier



Multiclass classifier can be constructed as N independent one-class classifiers.
N classes → *N* models.

Discrimination



Discrimination:

each object is assigned to the class – *completeness condition,*
and only to a single one – *purity condition*

Classification errors

- **One-class classifier:**
- Statistical criterion with null hypothesis H_0 : the object belongs to the class

Confusion matrix		Training set	
		Yes	No
Classifier	Yes	***TN***	***FN***
	No	***FP***	***TP***

- **First type error *FN*** – rejection of the sample, belonging to the class
  - Acceptance of “friends”
  - Connected with classifier sensitivity

- **Second type error *FP*** – assignment to the class of the object which does not belong to the class
  - Rejection of “aliens”
  - Connected with classifier specificity

Classification metrics

Sensitivity = TP/(TP + FN)

Specificity = TN/(TN + FP)

Accuracy = (TP + TN)/(TP + FP + TN + FN)

TP – true positive samples (P classified as P)

FN – false negative (P classified as N)

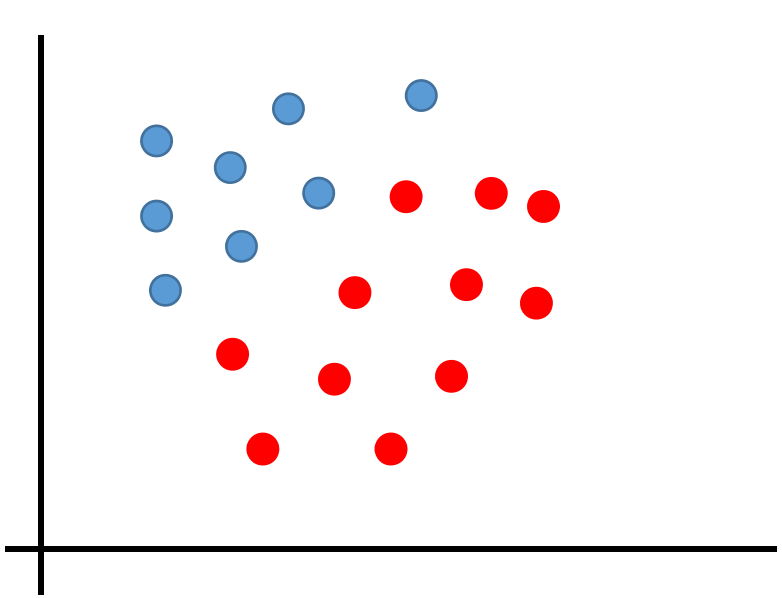
TN – true negative

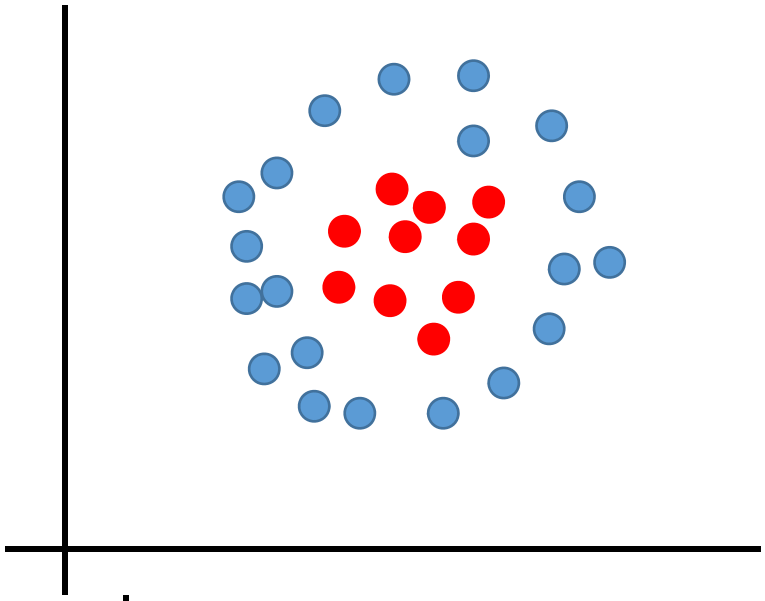
FP – false positive samples (N classified as P)

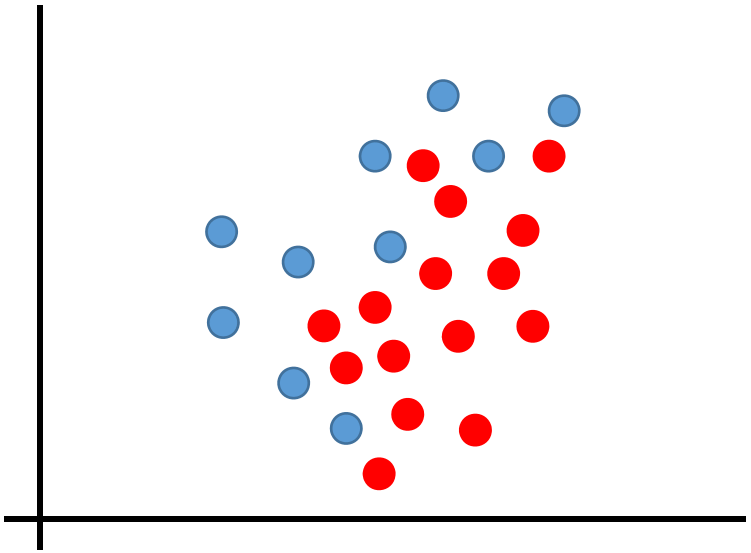
Classification error

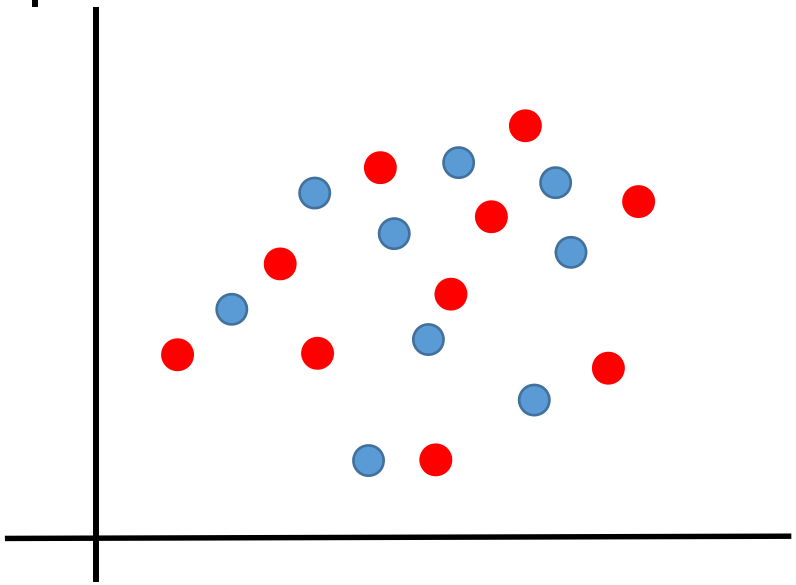
- **Two class discrimination**
  - **First type errors for class 1 become second class errors for class 2**
  - **Second type errors for class 1 become first type error for class 2**
  - **Thus, there are two formally independent metrics**

Geometrical interpretation

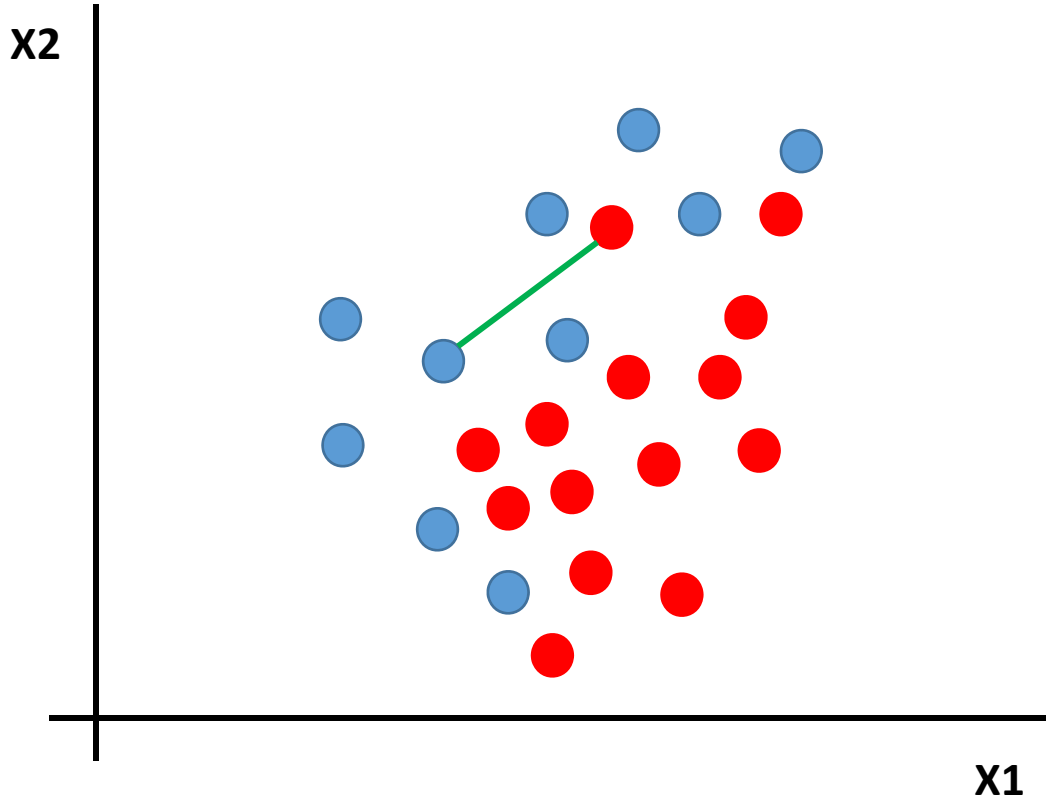








k Nearest Neighbors



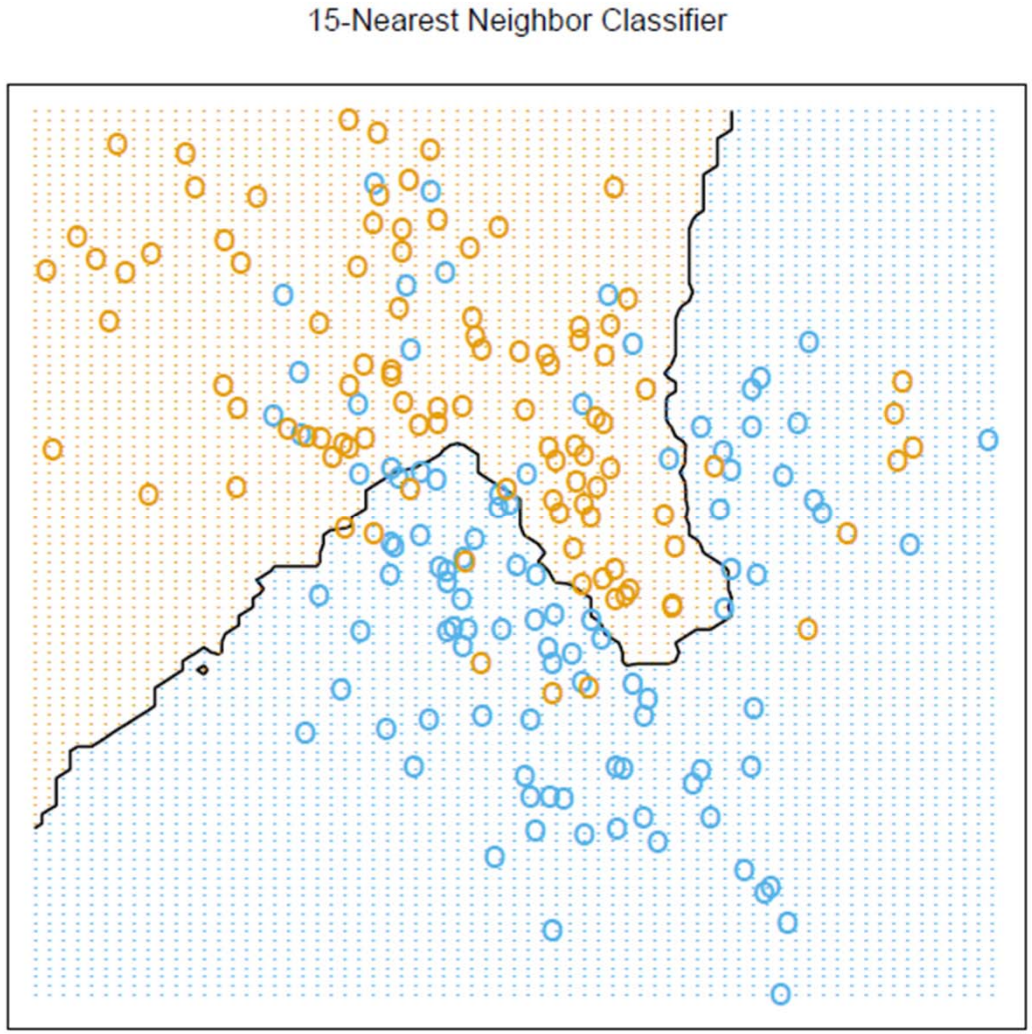
- counting the neighbors of the sample

- assigning the sample to the class where most of the neighbors belong

Tell me who is your friend and I will tell you who you are

$$d=sqrt\ ((x_1^1-x_1^2)^2+(x_2^1-x_2^2)^2)$$

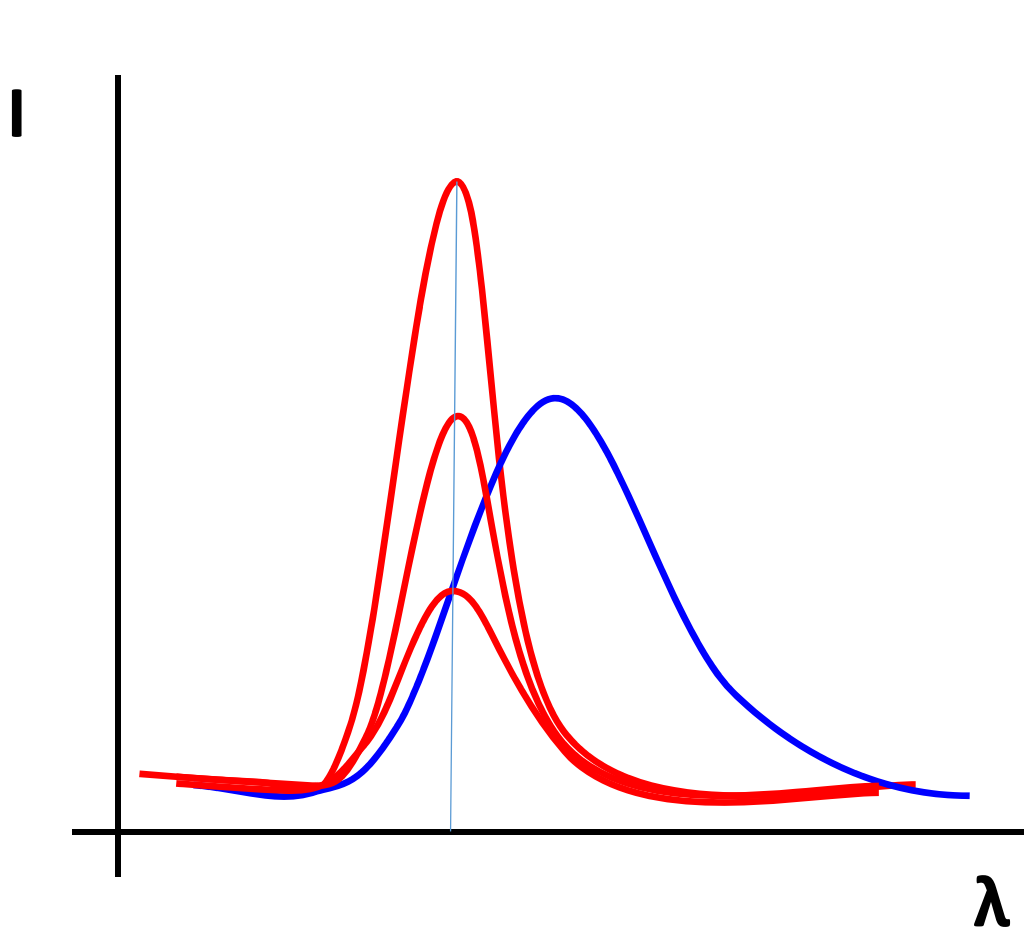
kNN example

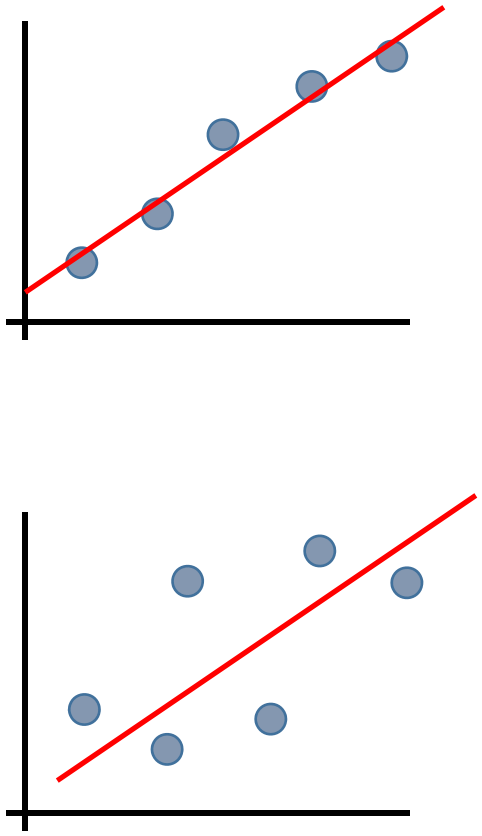


PLS regression

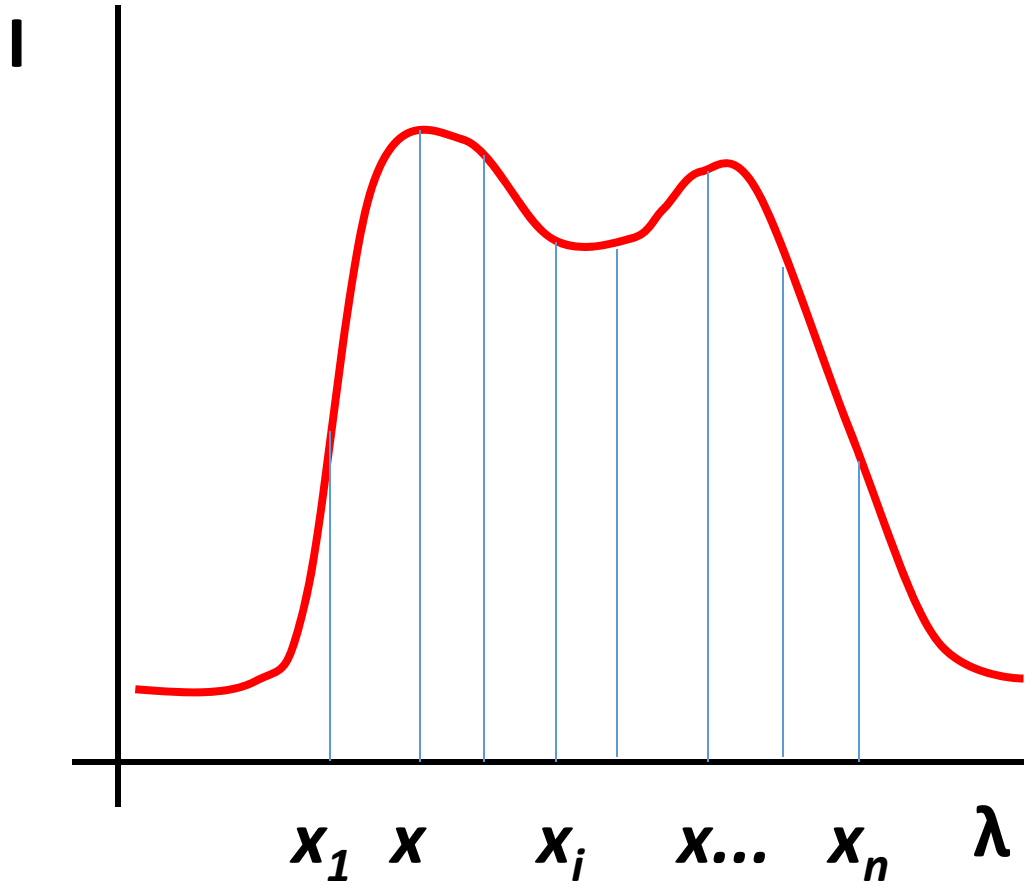
- **Most powerful and universal multivariate calibration algorithm**
- **Widely applied in modern chemometrics**
- **Works both for calibration (qualitative) and classification (quantitative) tasks**
- **Default method for many applications**

Univariate calibration

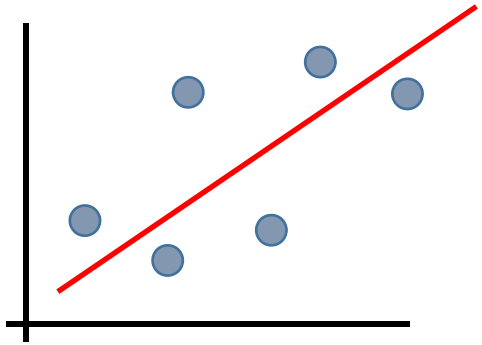




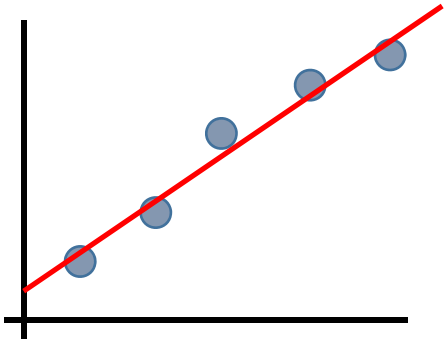
Multivariate calibration



$y=a+bx$

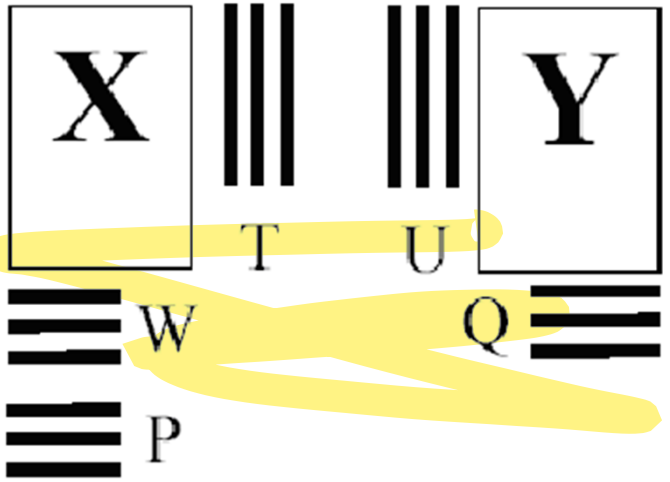


$Y=A+BX$



PLS regression

- Modeling of both X and Y
- 2 sets of scores and loadings
- additional matrix W (loading-weights)
- criterion: max covariance between T and U



$$X = TP^T + E$$

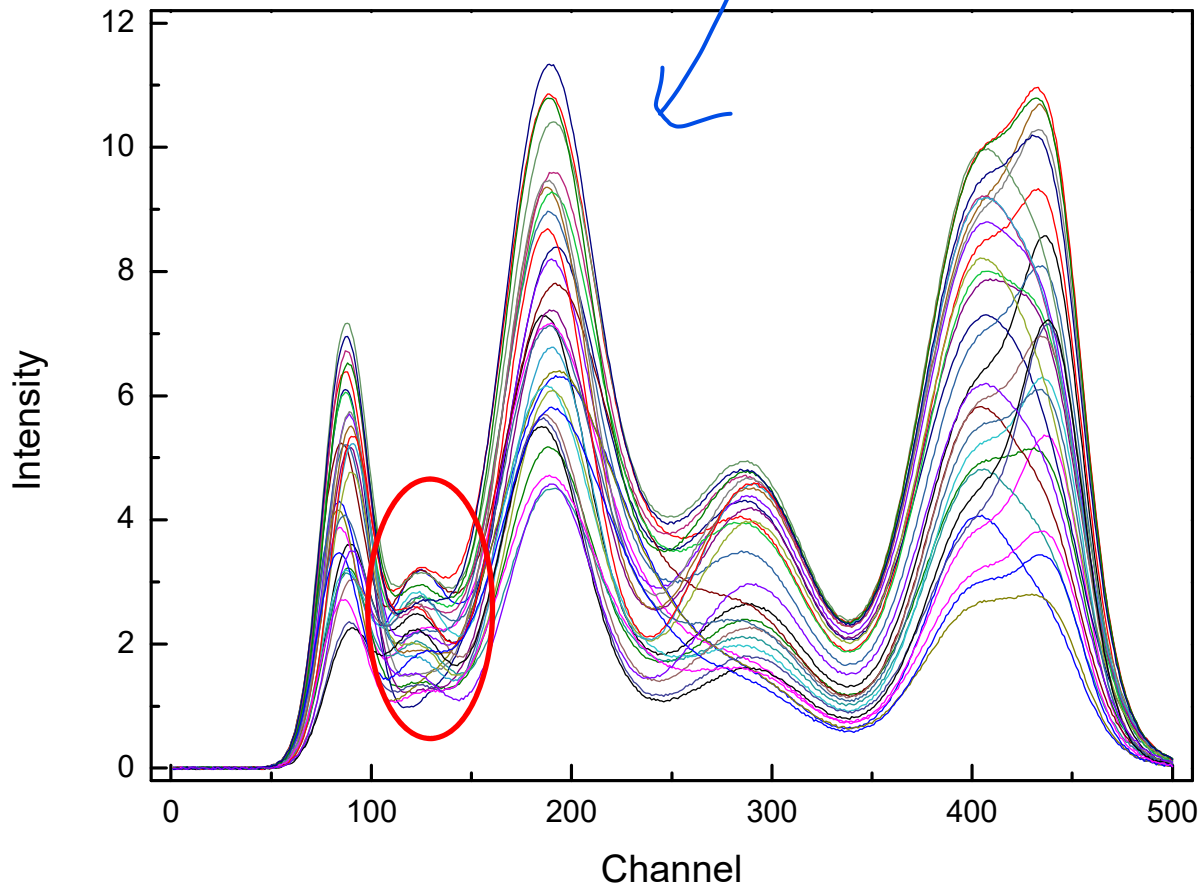
$$Y = UQ^T + F$$

$$W = \max(\mathrm{cov}(T,U))$$

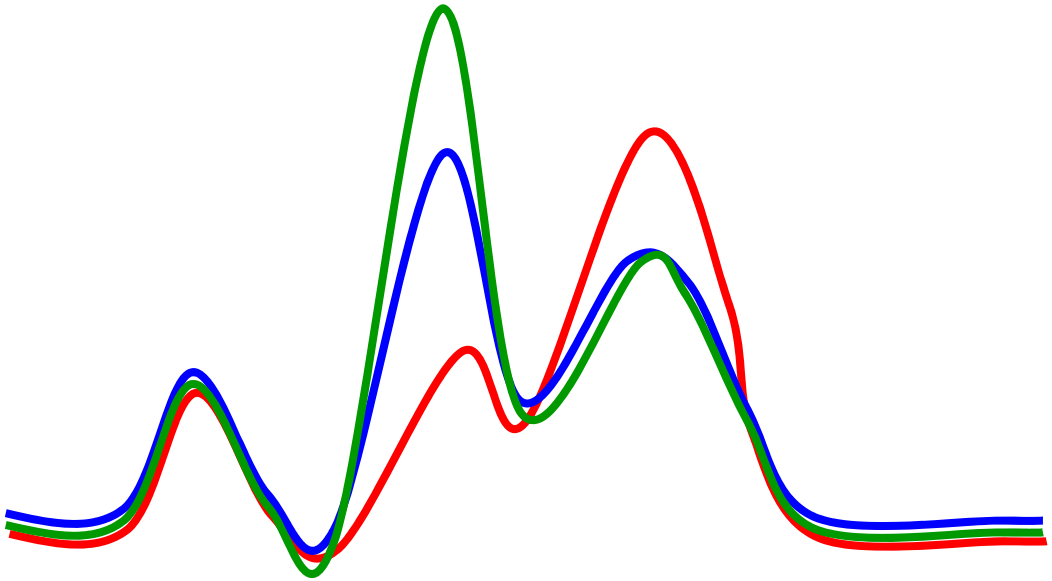
$$B = W(P^TW)^{-1}Q^T$$

The main idea behind PLS

I don't get it!?



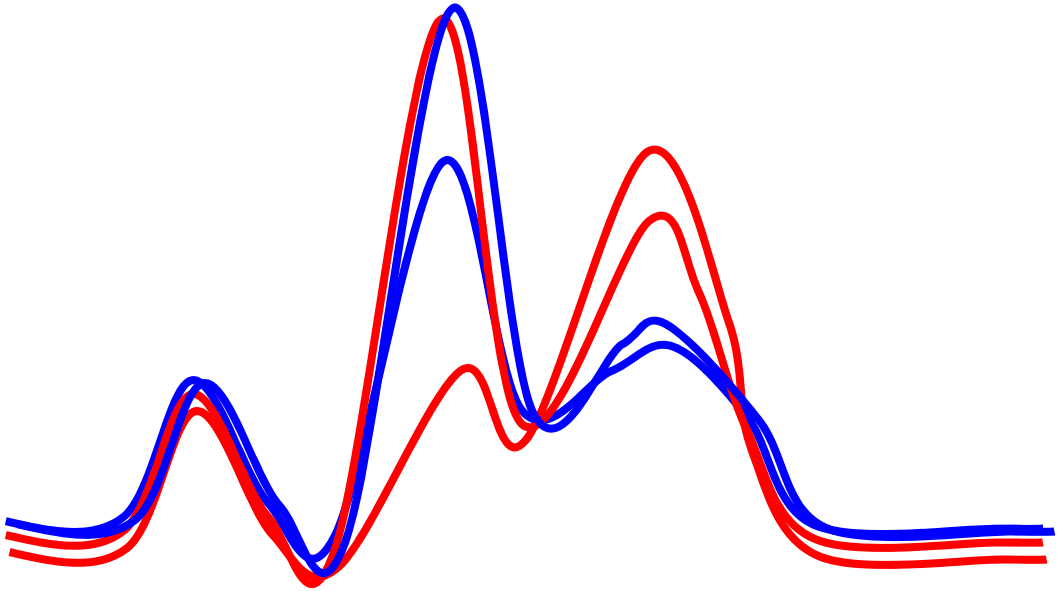
The main idea behind PLS



Concentration
5
10
15

PLS looks only for that part of variance in X, Which is *correlated* with variance in Y

PLS-DA encodes the classes with zeroes and ones



Class
0
0
1
1

What if there are several classes?
“One vs all” approach

Prediction with PLS

$$\hat{Y} = X_{new}B$$

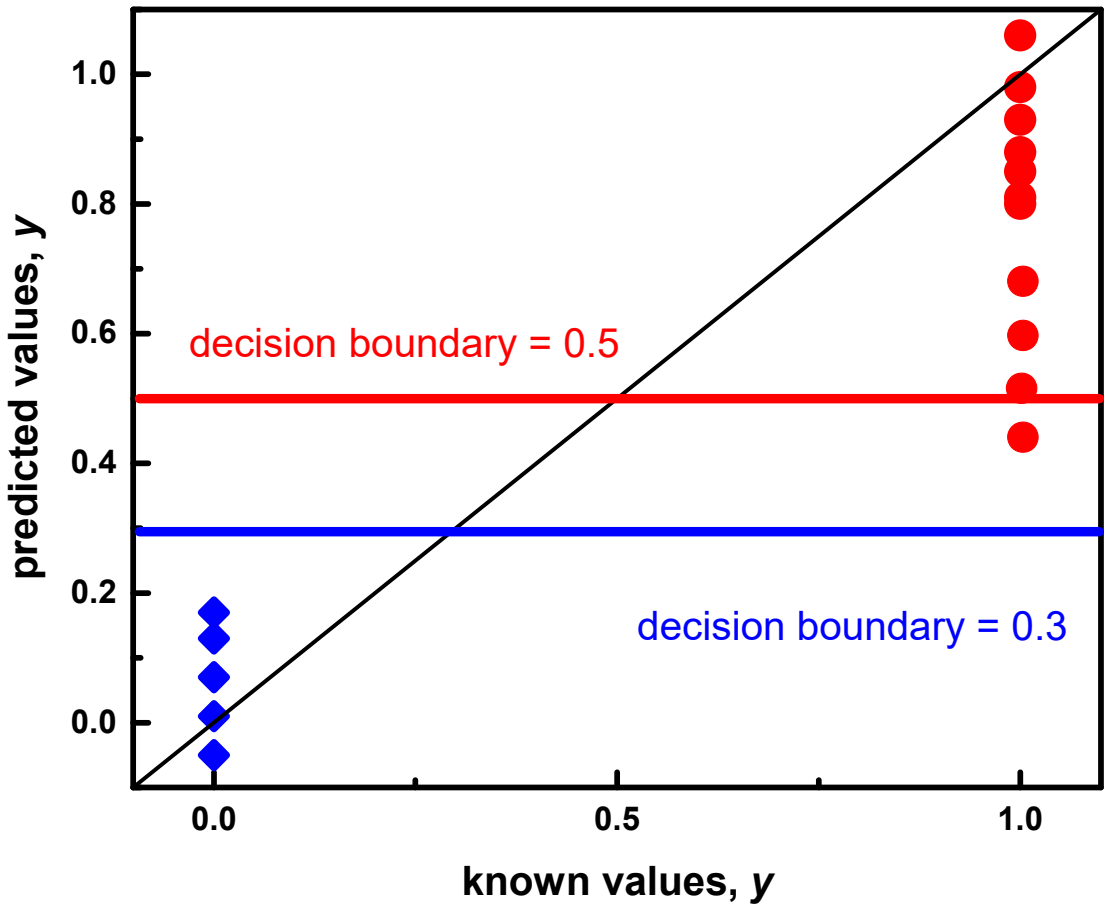
$$B = W(P^TW)^{-1}Q^T$$

$$W = \max(\text{cov}(T,U))$$

$\hat{Y}$ in PLS-DA will predict the class for a new sample at certain *decision boundary* value

PLS modeling output

- **Decision boundary**

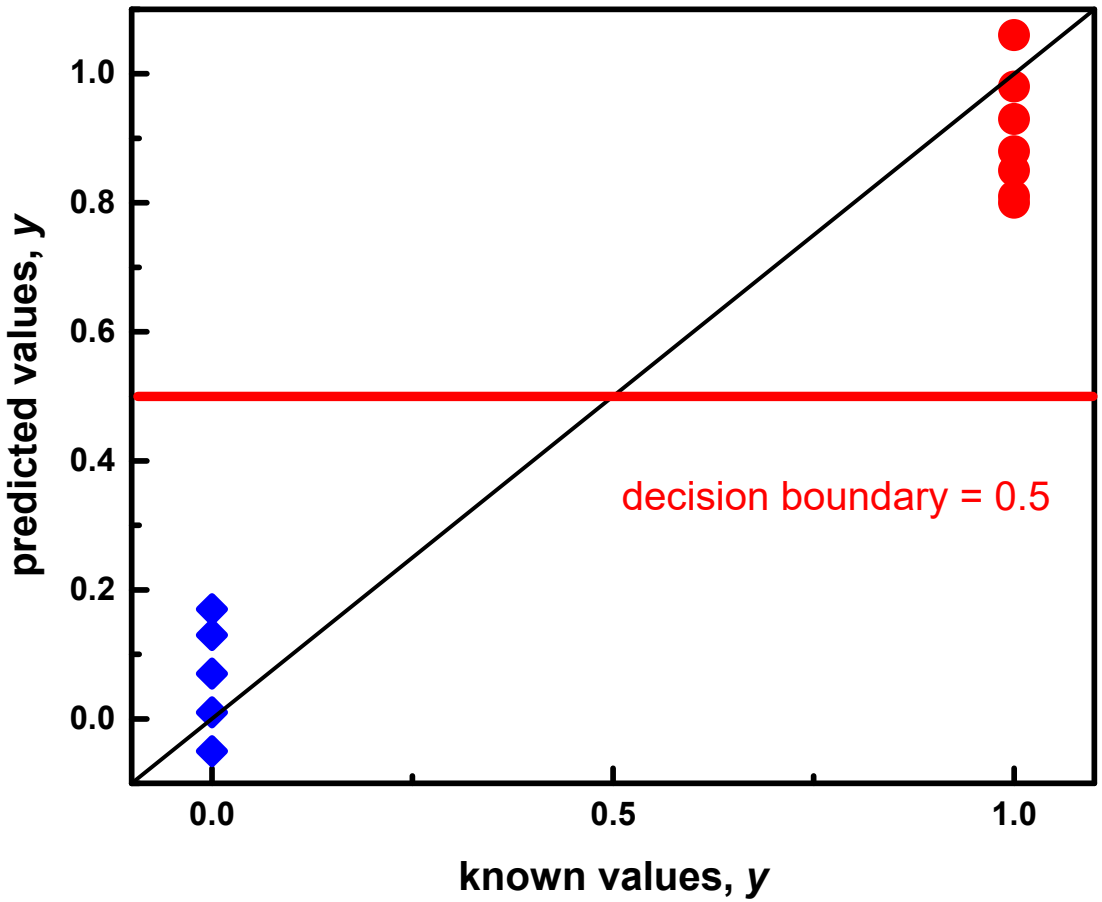


PLS modeling output

- **Measured vs predicted plot**
- **Explained variance plot**
- **Regression coefficients plot**

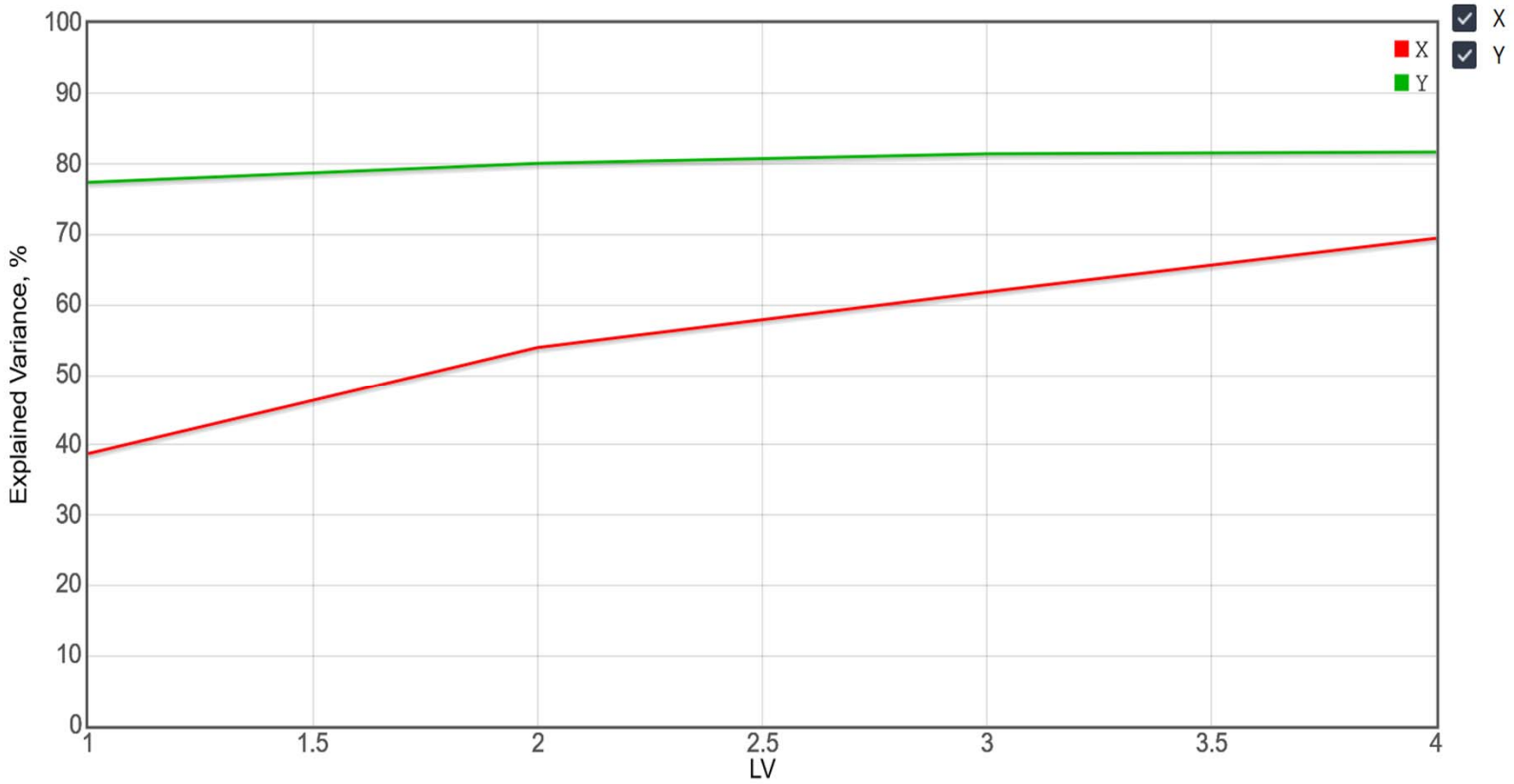
PLS modeling output

- **Measured vs predicted plot**

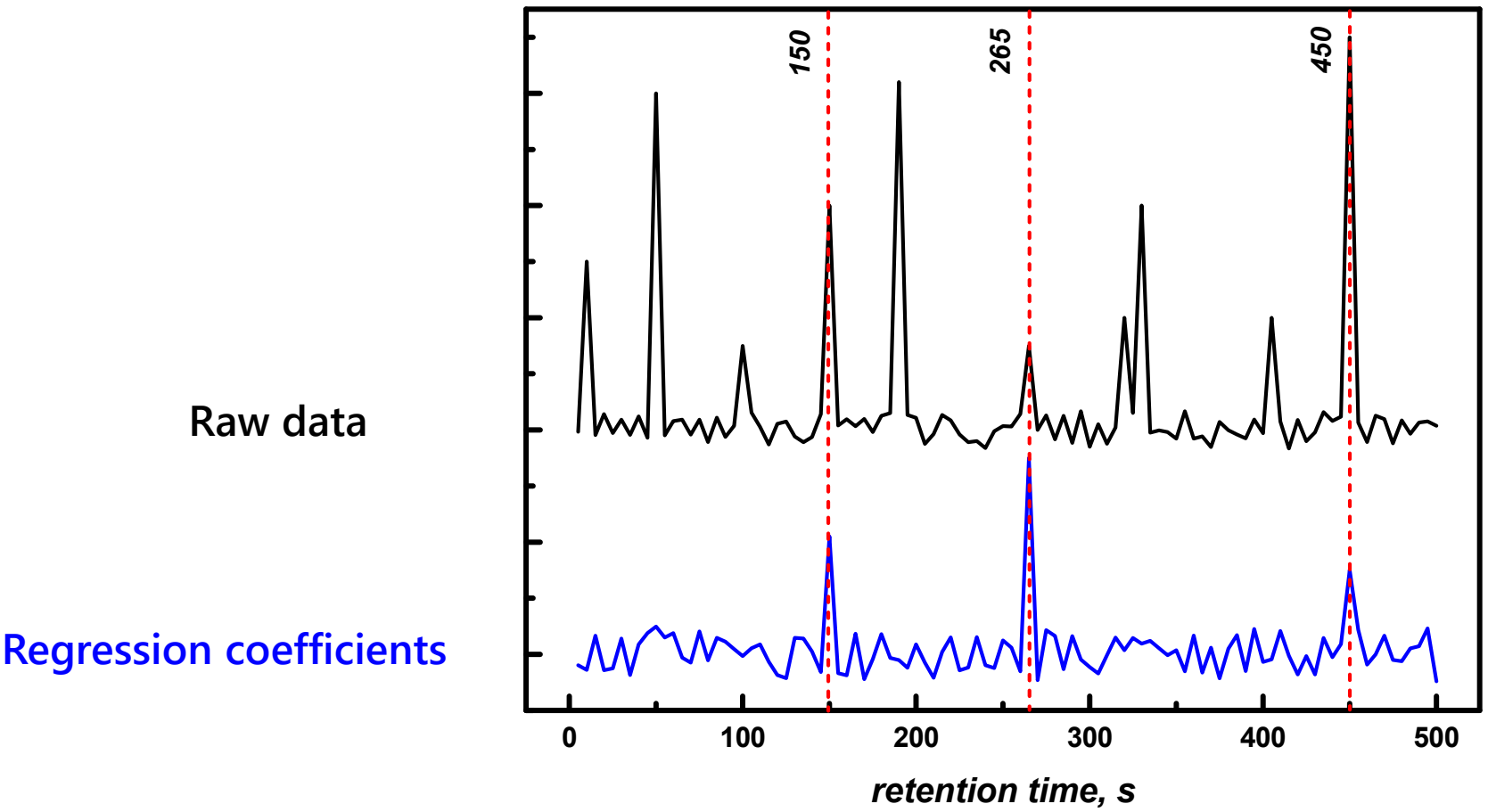


PLS modeling output

- **Explained variance plot**



Regression coefficients



PLS-DA is not one-class classifier

It is not possible to acquire a representative sample set of something which is not a target class

PLS-DA example

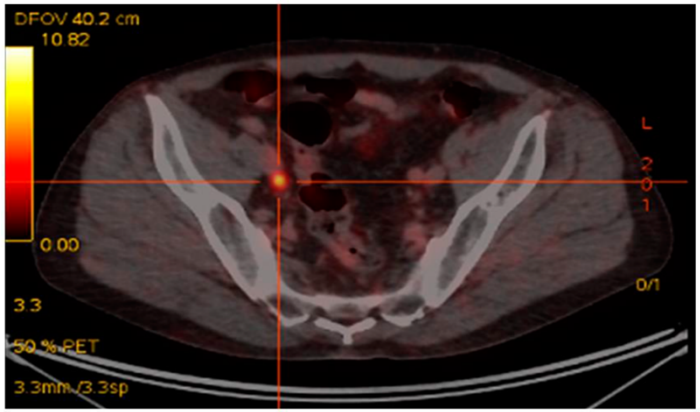
Distinguishing urine samples from patients with prostate cancer using potentiometric multisensor system

Prostate cancer

- Prostate cancer (PC) is the second most common cancer in males.
- One of the leading causes of cancer mortality.
- About 1 man in 9 will be diagnosed with prostate cancer during his lifetime.
- Prostate cancer develops mainly in older men. About 6 cases in 10 are diagnosed in men aged 65 or older, and it is rare before age 40.
- About 1 man in 41 will die of prostate cancer.

Diagnostic methods

Endorectal ultrasonography (ERUS)
Computed Tomography (CT)
Magnetic resonance imaging (MRI)
Positron-emission tomography (PET)



he **actual diagnosis** of prostate cancer can only be made with
prostate biopsy.

iomarkers

evel of prostate specific antigen (PSA) in blood with ELISA
t 4.0 ng/ml cutoff, the PSA test sensitivity and specificity are
4.9% and 63.1%, respectively.
high level of serum PSA is not necessarily evidence of PC and can also be caused by benign
rostatic hyperplasia (BPH), inflammation of the
rostatic gland, urinary retention or rectal palpation.

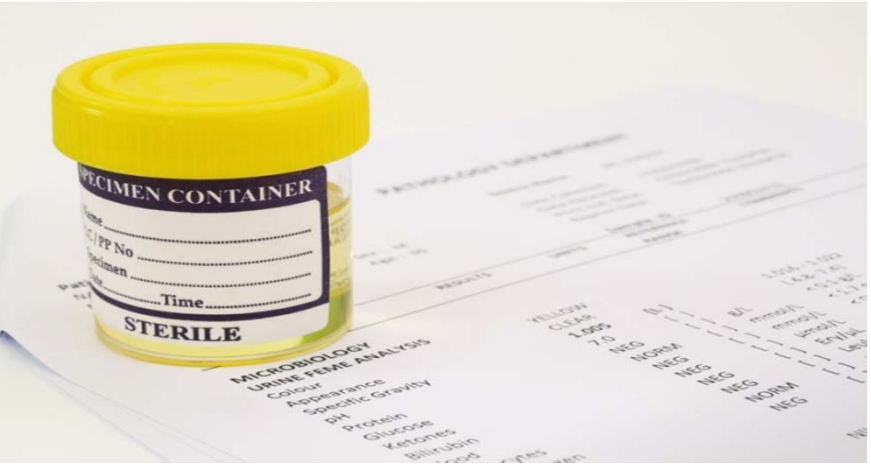
lso: sarcosine, alpha-methylacyl-CoA racemase

re credit: *Diagnostics* 2018, *8*(1), 16

Objective of this study

to check if potentiometric multisensor system can distinguish between the urine samples from patients with PC and those without

Experimental - Samples

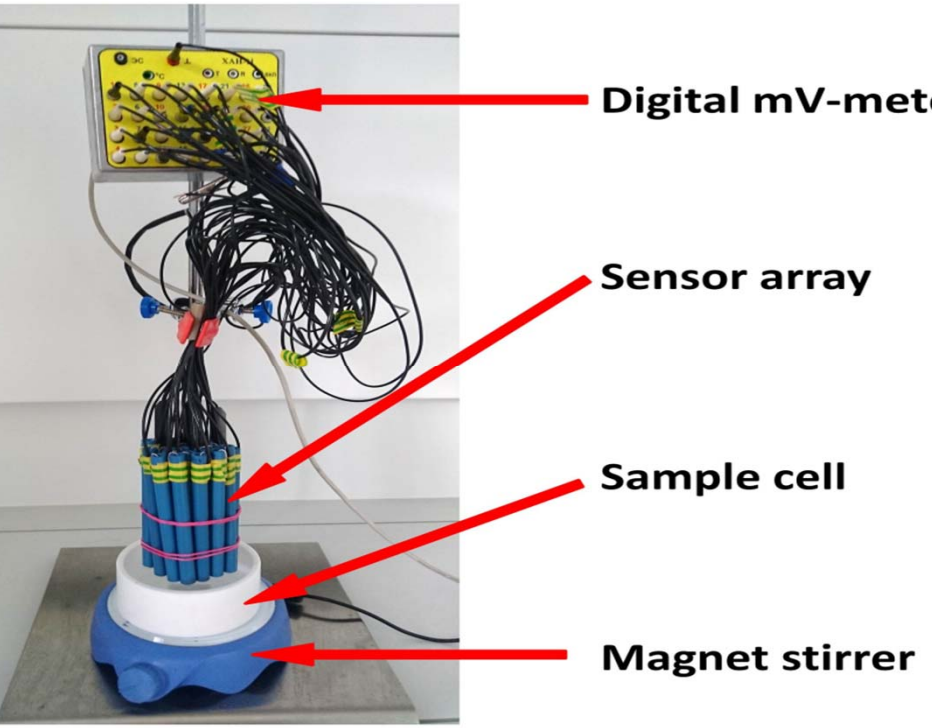


89 urine samples provided by Urology Clinic of S.M
Kirov Military Medical Academy (St. Petersburg, Russia)

43 samples from patients with PC diagnosis confirme
by prostatic puncture biopsy (P).
46 were from anonymous male volunteers with no P
diagnosis (N).
No further details on the samples were available due t
the considerations of ethics and confidentiality.

he samples frozen in 100 ml standard urine containers
orage in the lab refrigerator at -21^0 C.
he samples were thawed before the analysis using water bath at room temperature.

Experimental – Sensor array



28 potentiometric sensors with PVC-plasticized, chalcogenide glass polycrystalline membranes.

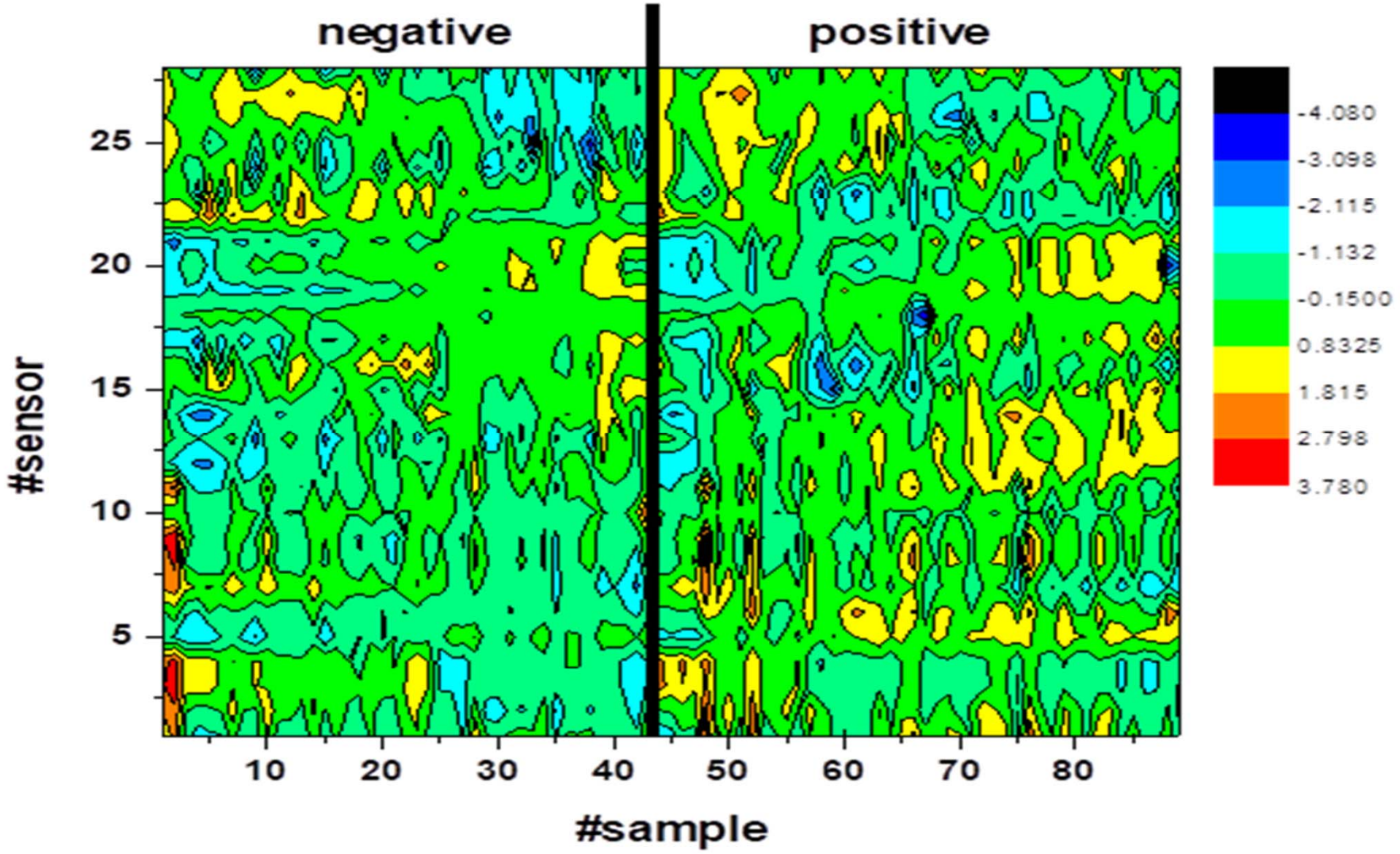
Standard Ag/AgCl reference electrode

Measurement time in each urine sample was 3 min in a Teflon sample cell under stirring.

Each sample was diluted 10 times (10 ml of urine in 9 ml of bidistilled water) before the potentiometri measurements.

Talanta 131 (2015) 556–561

Raw sensor responses – Heat map



SIMCA

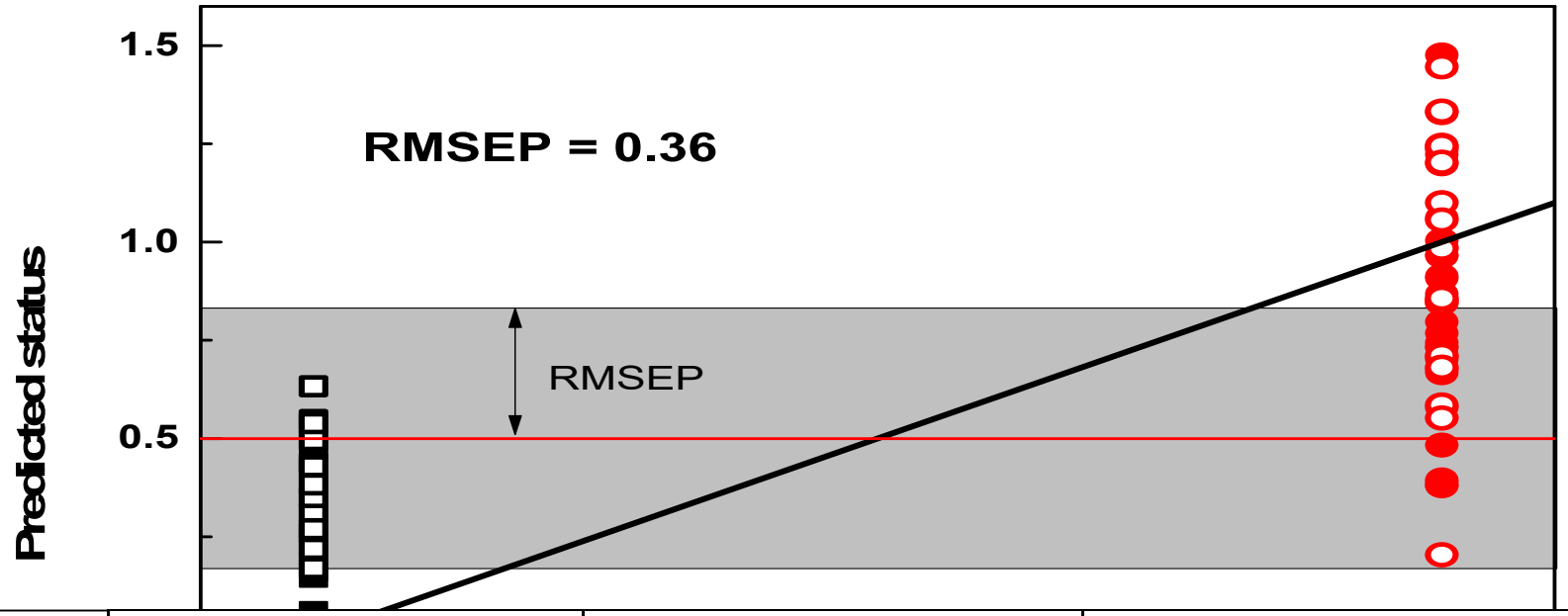
SIMCA was employed in a one-class mode: PCA model was calculated for *P* class and then all the calibration samples from *N* class as well as the test set were projected on this model.

The model was based on 6 LVs accounting for 92% of total variance

The following results were obtained:

Calibration	**Test set**
Sensitivity 100%	Sensitivity 80%
Specificity 7%	Specificity 13%
Accuracy 55%;	Accuracy 47%

PLS-DA



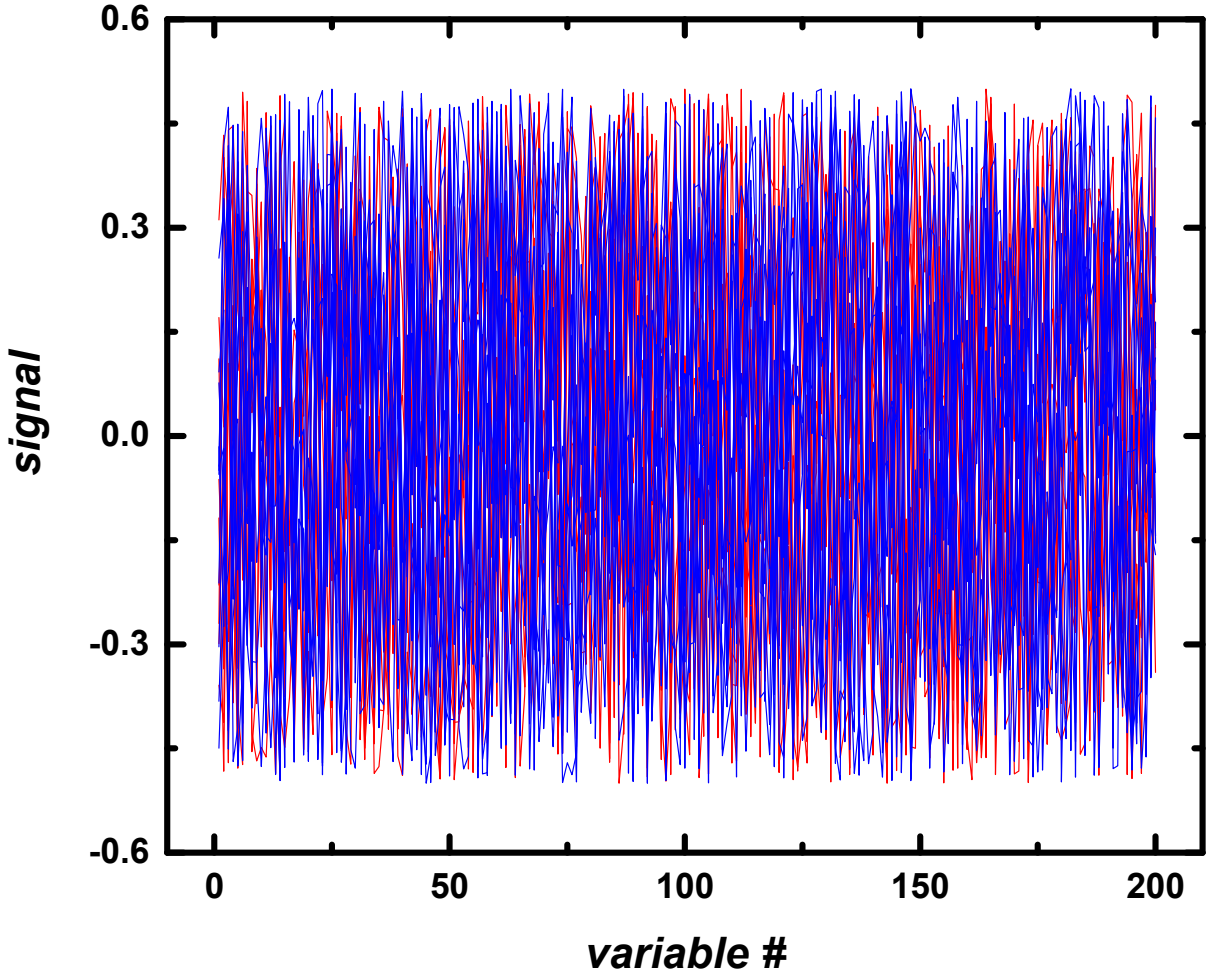
Decision rule	Sample set	Sensitivity	Specificity	Accuracy
ixed decision boundary	calibration	84	100	92
	test	93	80	87
Decision boundary ± RMSEP	calibration	42	29	36
	test	60	13	37

Randomly assigned classes

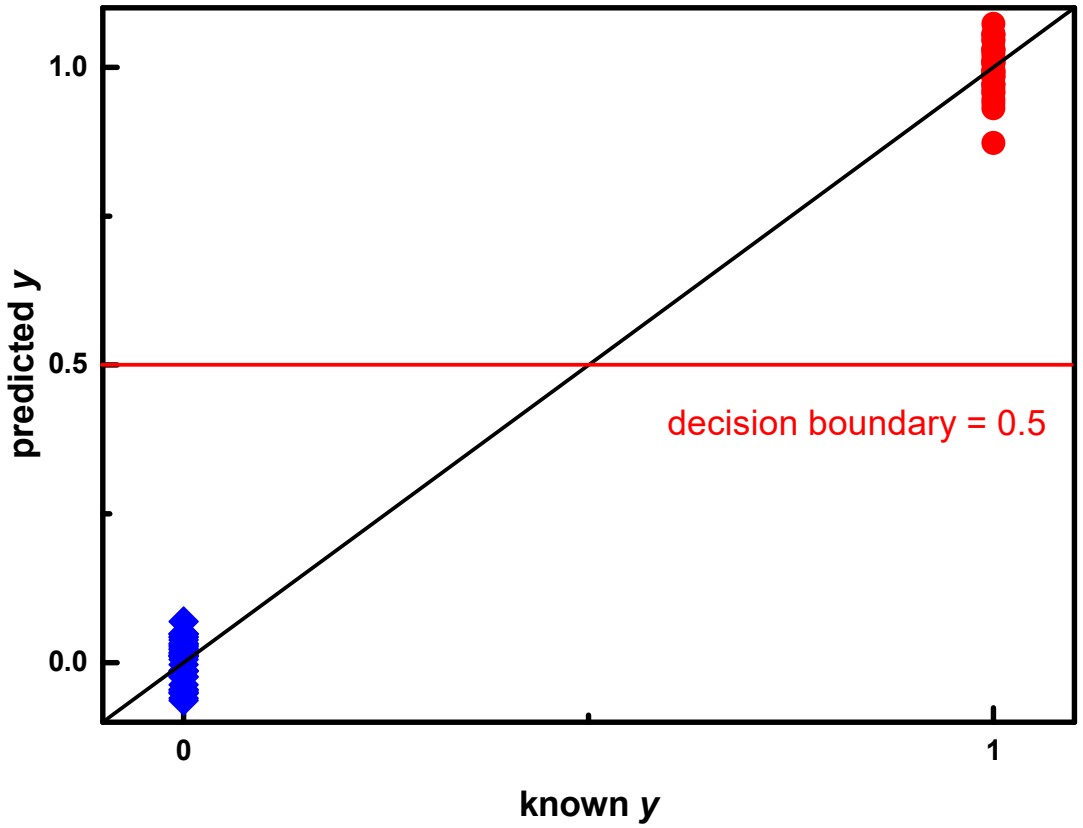
0 variables

0 samples

ndomly assigned to two classes

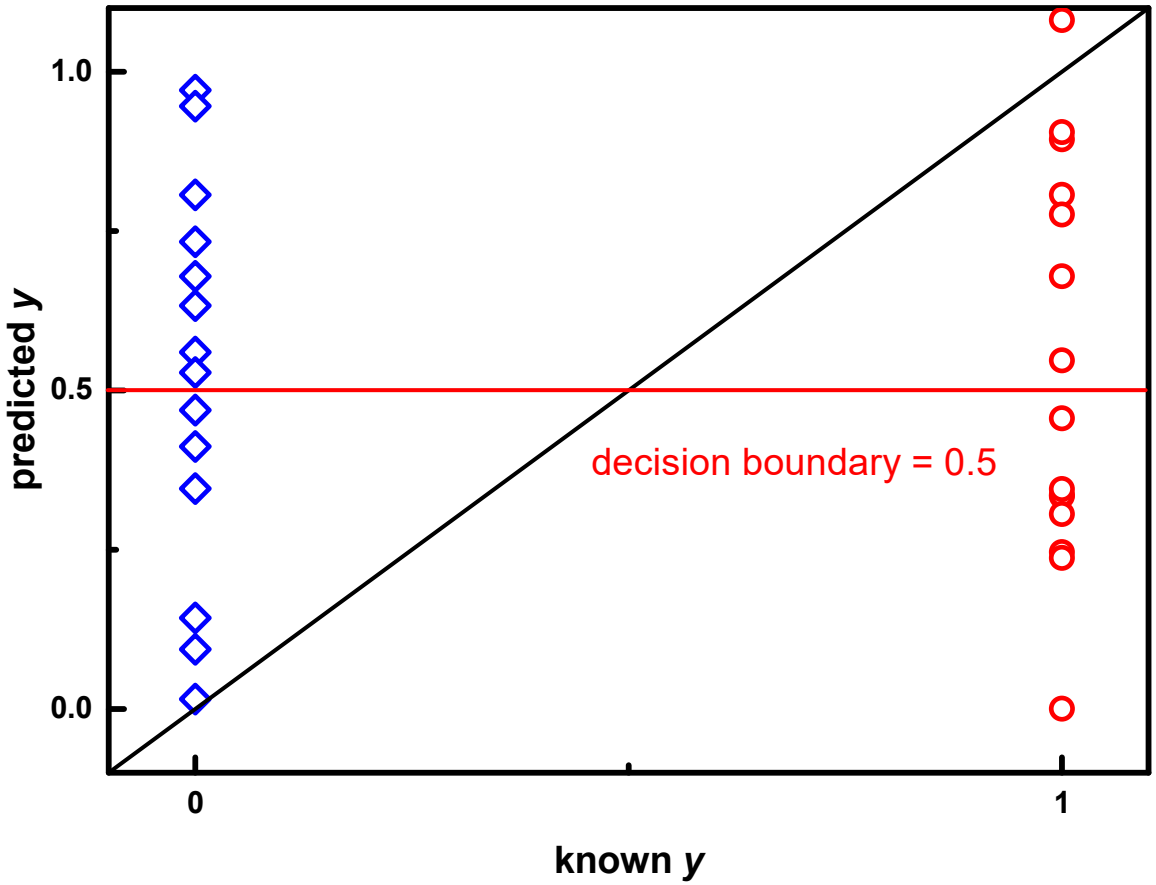


PLS-DA calibration with randomly assigned classes and noise data



PLS is eager to please

PLS-DA validation with randomly assigned classes and noise data



How to make a good model?

- **appropriate experimental design**
- **visual study**
- **preprocessing (smoothing, filtering, scaling/weighting if needed)**
- **model interpretation, outliers detection**
- **choice of appropriate number of LVs, variable selection**
- **validation, validation, validation**

Thank you for your attention!