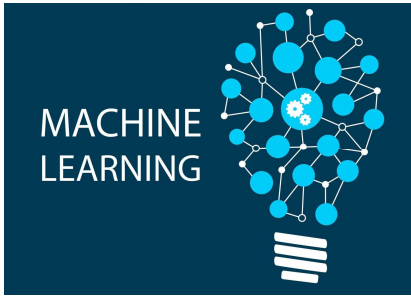
Local PLS regression
– Overview and extensions



matthieu.lesnoff@cirad.fr

cirad

SELMET

ChemHouse

SensorFINT 08/09/2022 Sète, France

Livestock Systems and Animal Products Management

The SELMET research unit is an international group focused on livestock production in Mediterranean and tropical areas. The unit has approximately 100 permanent and contractual staff drawn from CIRAD, INRAE and the Institut Agro Montpellier. SELMET offers research, teaching and expertise to support sustainable transitions in livestock activities.

Lab NIR + Chemical

Forages and feed

- chemical composition
- digestibility

- Introduction
- General principles
- Illustrations
- Some details
- Some extensions

Introduction

"Local PLSR" non-linear prediction pipeline

... using linear models (PLSR)

Direct adaptation of well-known statistical models:

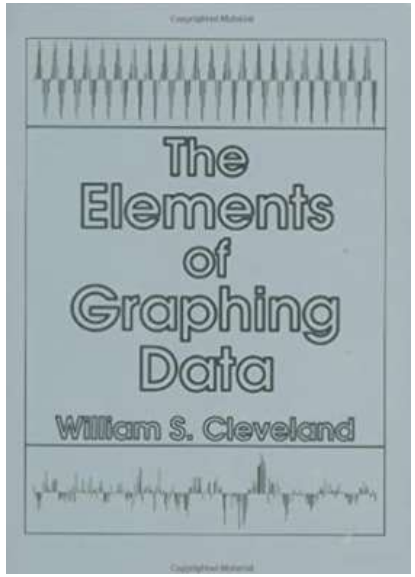
locally weighted regression (lwr)

Ex: Algorithm LOWESS (smoother)

Cleveland, W.S., 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association 74, 829. https://doi.org/10.2307/2286407

Cleveland, W.S., Devlin, S.J., 1988. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association 83, 596–610. https://doi.org/10.1080/01621459.1988.10478639

Cleveland, W.S., Grosse, E., 1991. Computational methods for local regression. Stat Comput 1, 47–62. https://doi.org/10.1007/BF01890836LOWESS,



Locally weighted regression lwr

- Weighted ordinary least squares (OLS)
- Weights = f(distance to the observation to predict)
- (+ robustness rules)

n observations
= Training data set
$\{x_i, y_i ; i = 1, \ldots, n\}$

$$\sum_{i=1}^{n} \frac{1}{n} (y_i - \boldsymbol{x}_i' \boldsymbol{b})^2$$ Ordinary LS

$$\sum_{i=1}^{n} w_i (y_i - \boldsymbol{x}_i' \boldsymbol{b})^2$$ Weighted LS

x_{new} to predict

$$\sum_{i=1}^{n} w_{i[x_{new}]} \left(y_i - \boldsymbol{x}_i' \boldsymbol{b}_{[x_{new}]}\right)^2$$ Locally weighted LS

The model changes for each new observation to predict

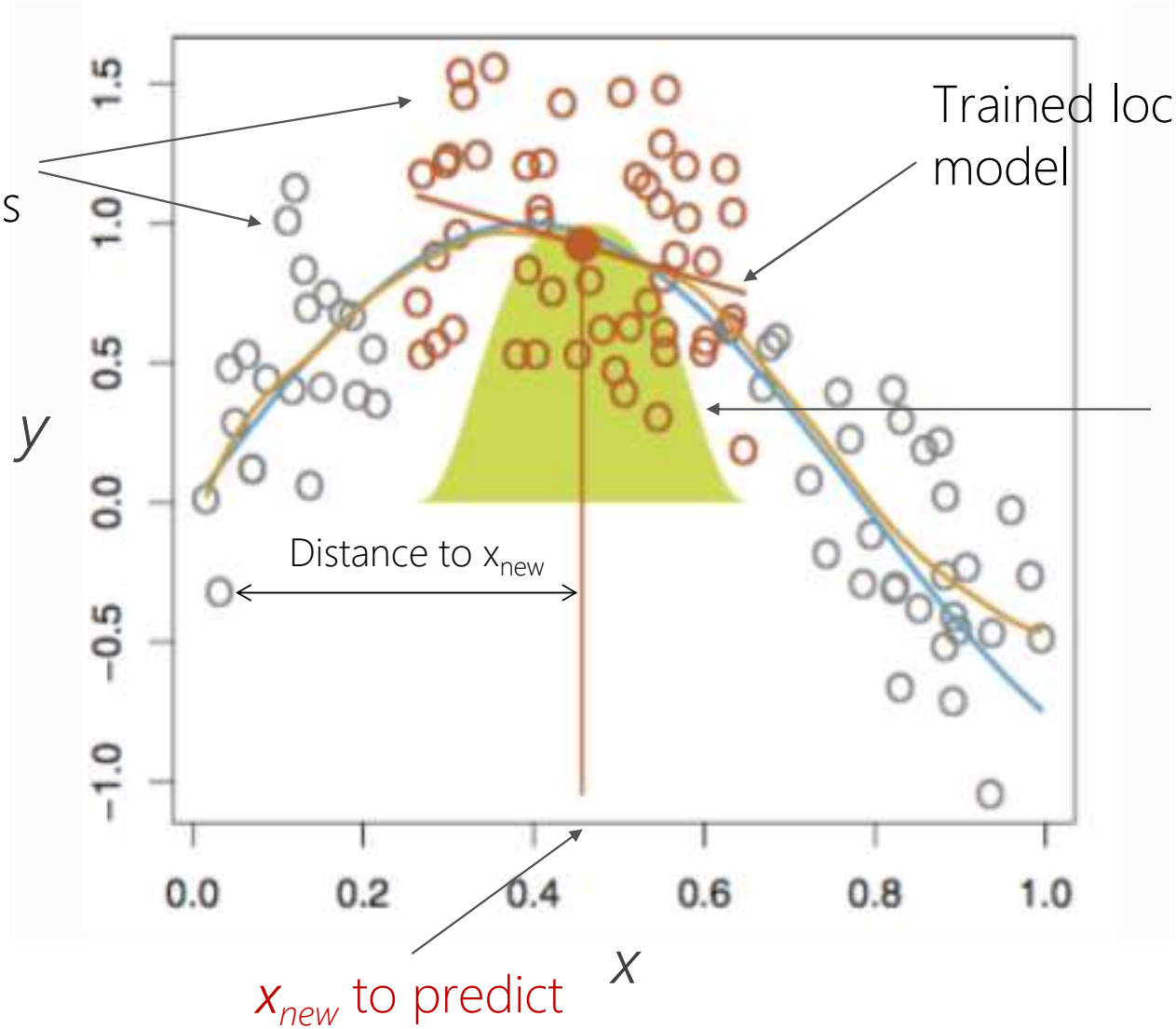


Image: https://xavierbourretsicotte.github.io/loess.html

lwr = LW-MLR locally weighted multiple linear regression

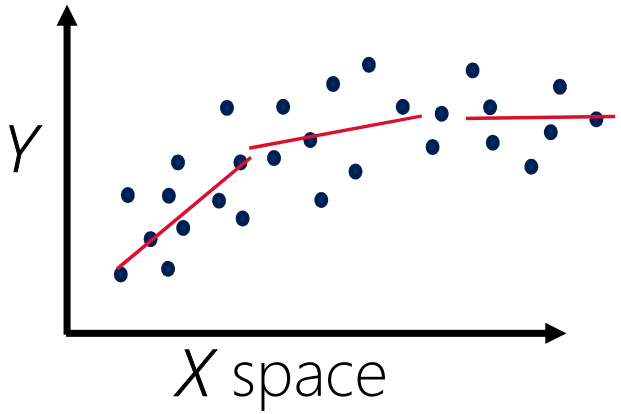
⇓ extension

"Local PLS" Naes & Isaksson 1990 ("LWR"~local-PCR), Shenk et al. 1997

- Regression LW-PLSR
- Discrimination LW-PLSDA

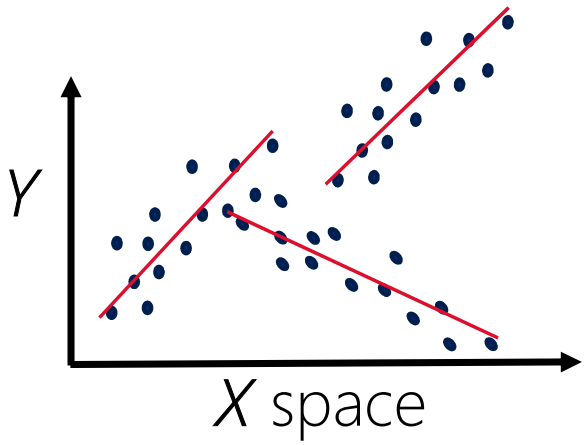
Useful when non linearity between *X* and *Y*

Curvature



Curved trend

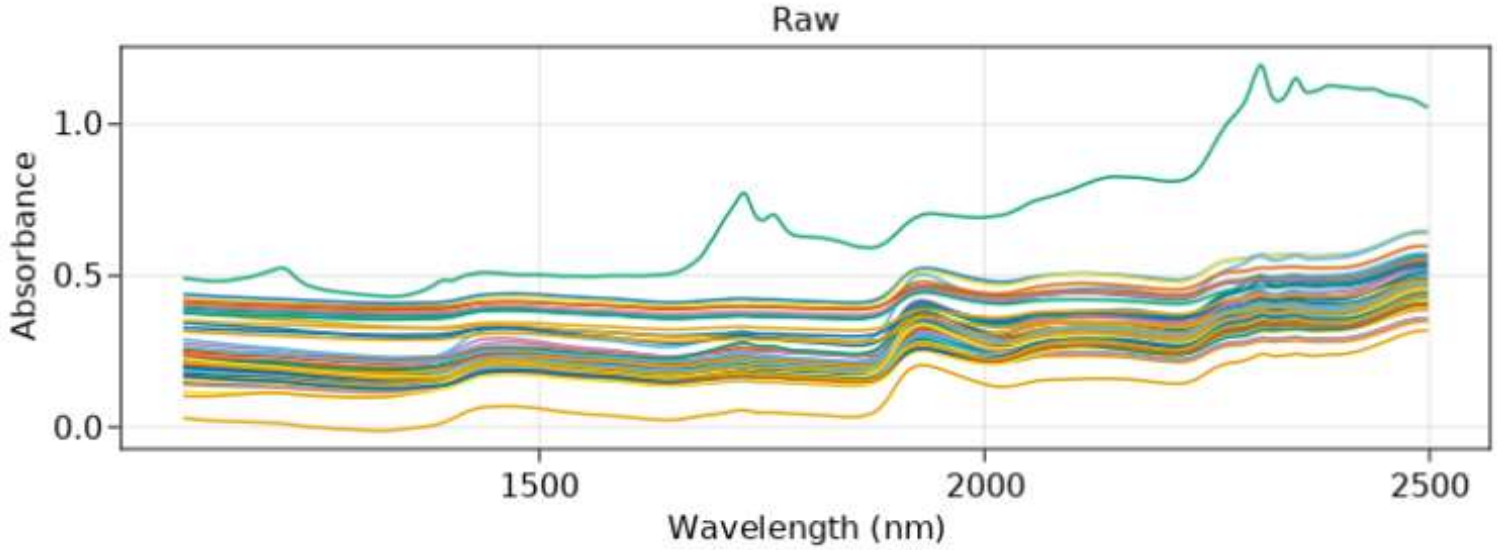
Clustering, heterogeneity



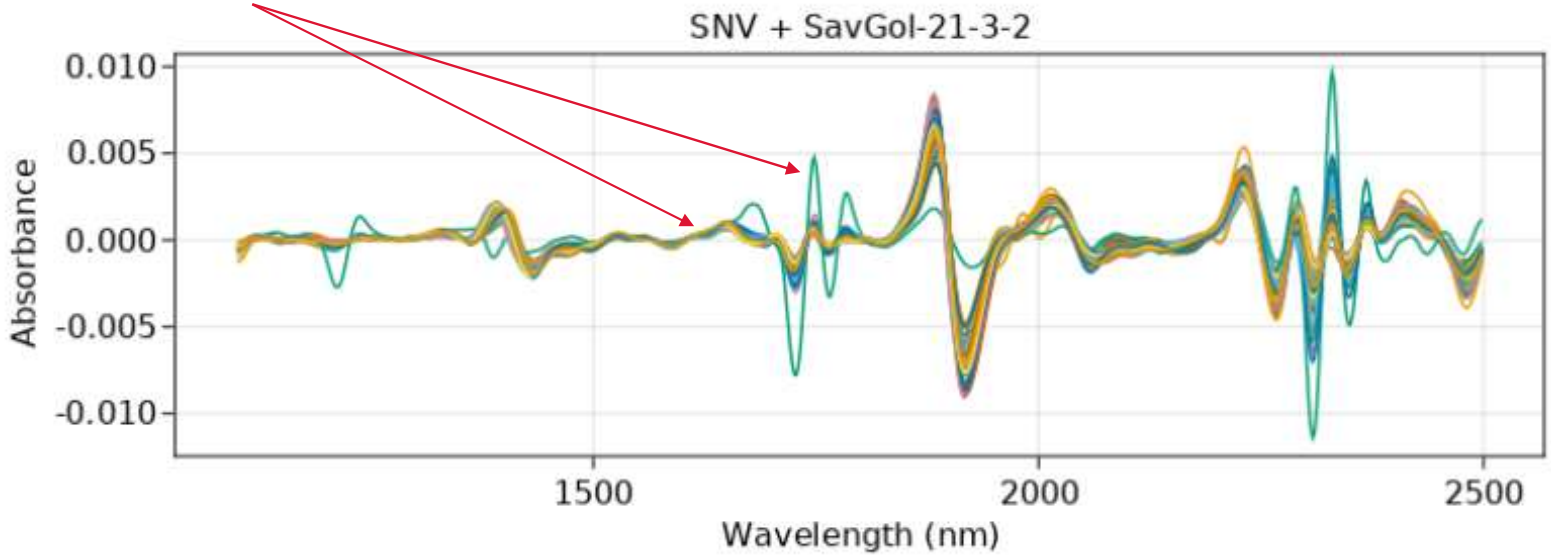
Relation between *X* and *Y* varies between the clusters

Animal feces
y = Digestibility of feed

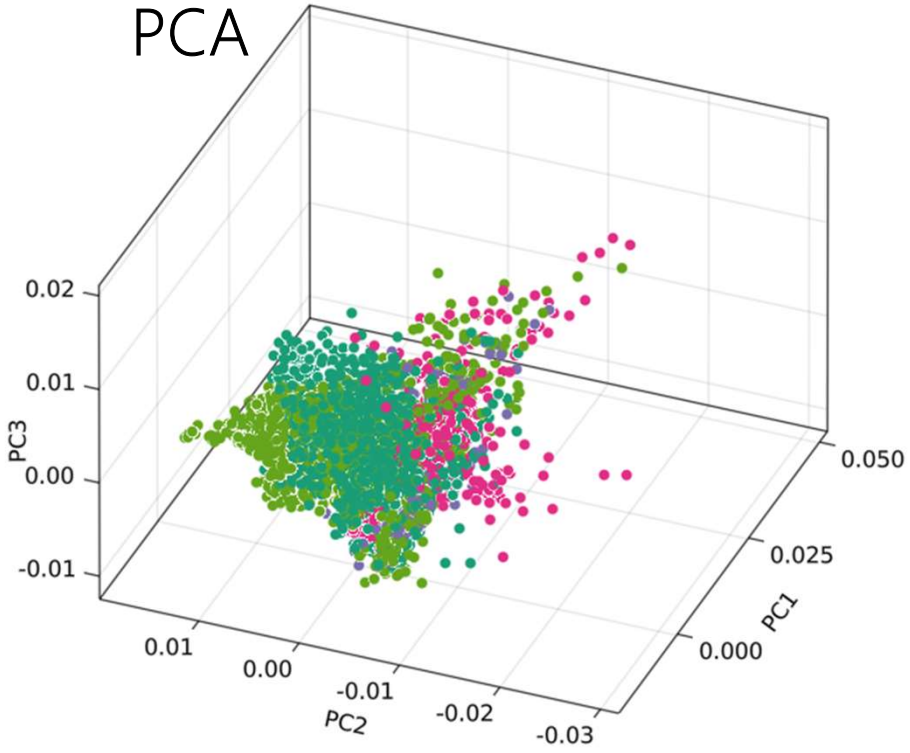
Foss NIR System



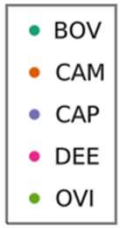
Heterogeneity



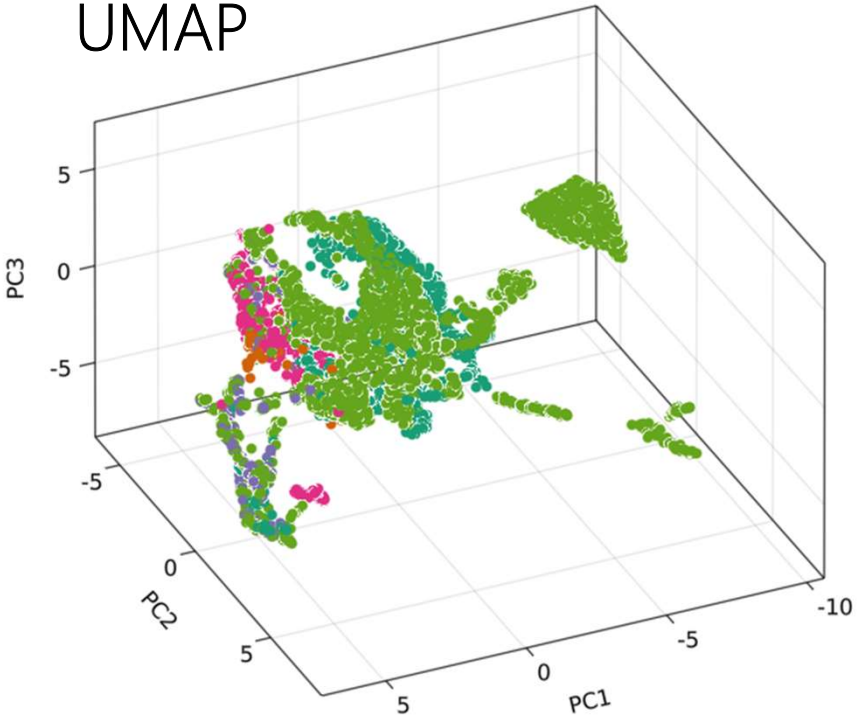
PCA

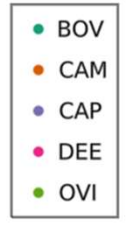


Feces origin



UMAP





With such heterogeneity, we can expect non-linearity
Non-linear models are requested

- Many different pipelines of local PLSR/DA can be built

- Their comparative performances is often data-dependent

General principles of local PLSR

Attempt of typology of algorithms

1. kNN-PLSR = usual “local PLSR”

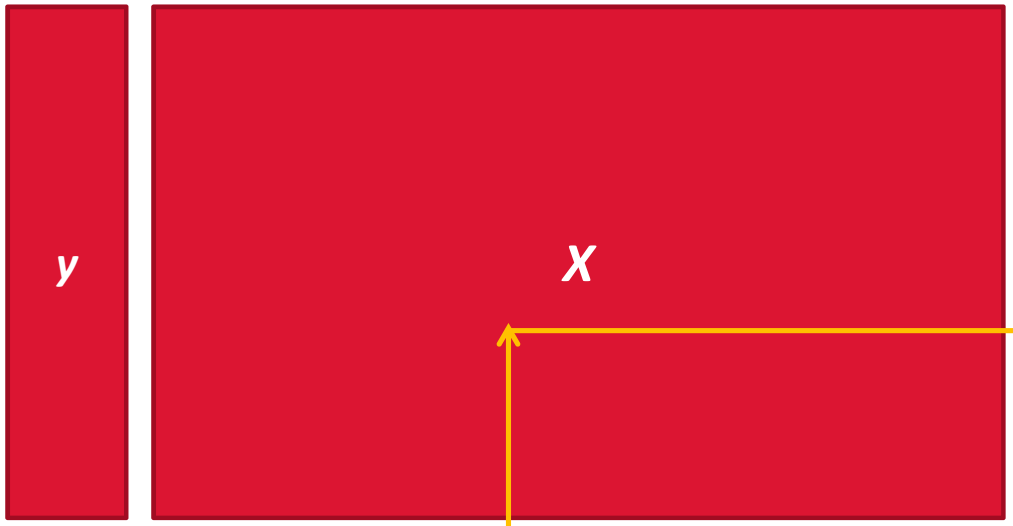
   kNN = k nearest neighbors

2. LWPLSR

3. kNN-LWPLSR (1+2)

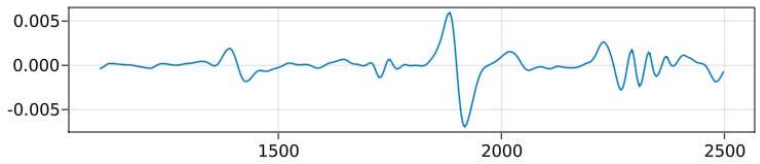
All have the same theoretical background

1. kNN-PLSR = usual "local PLSR"

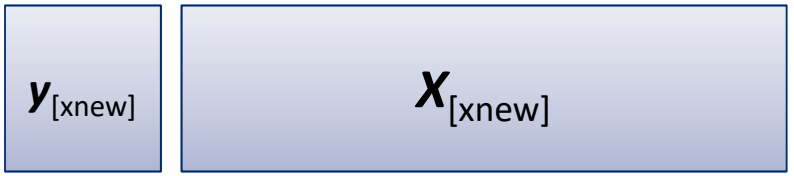


Search/selection
of *k* nearest neighbors
of x_{new} in *X*

New spectrum
to predict x_{new}



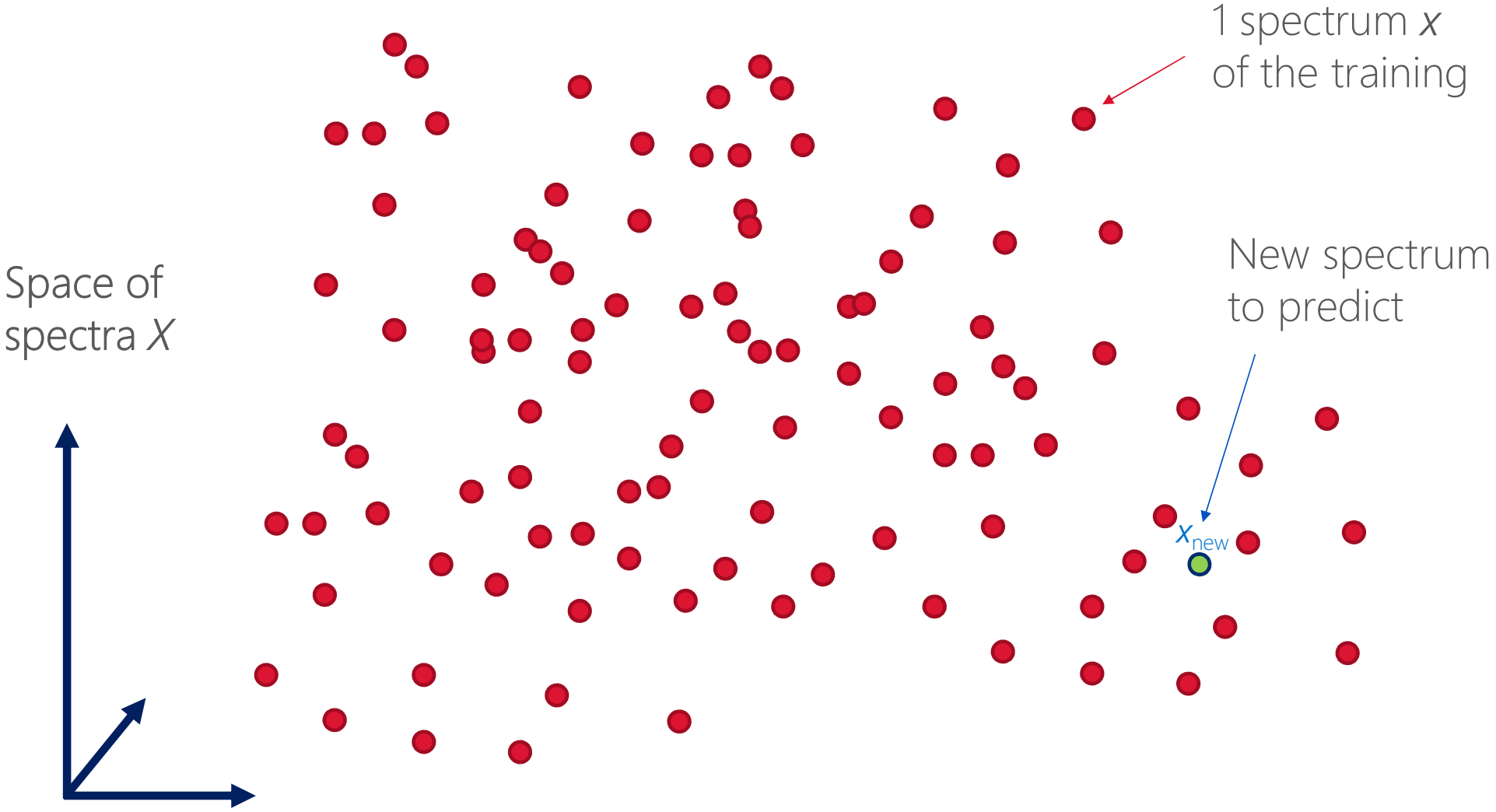
Neighborhood of x_{new} ~Homogeneous

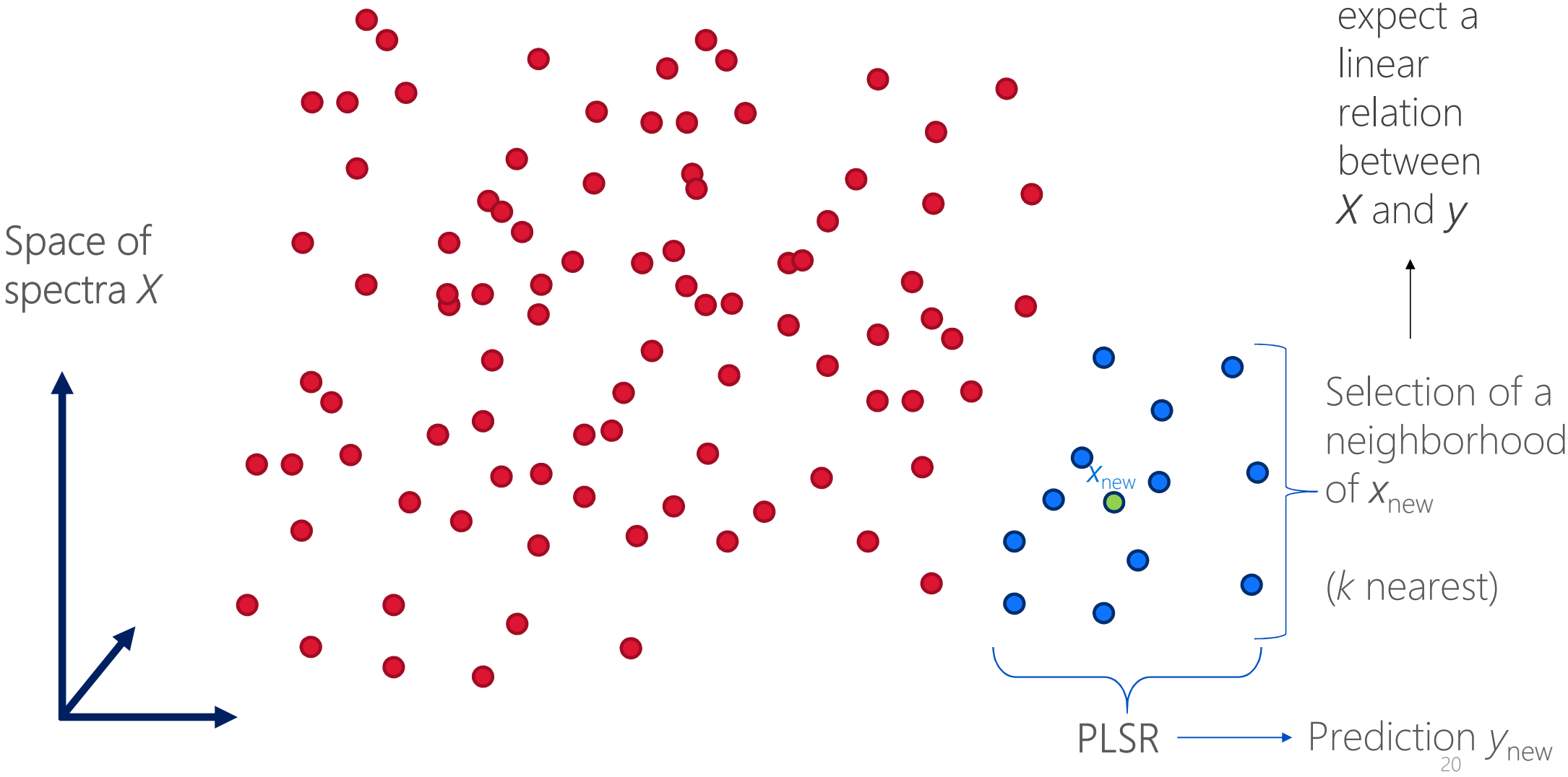


PLSR

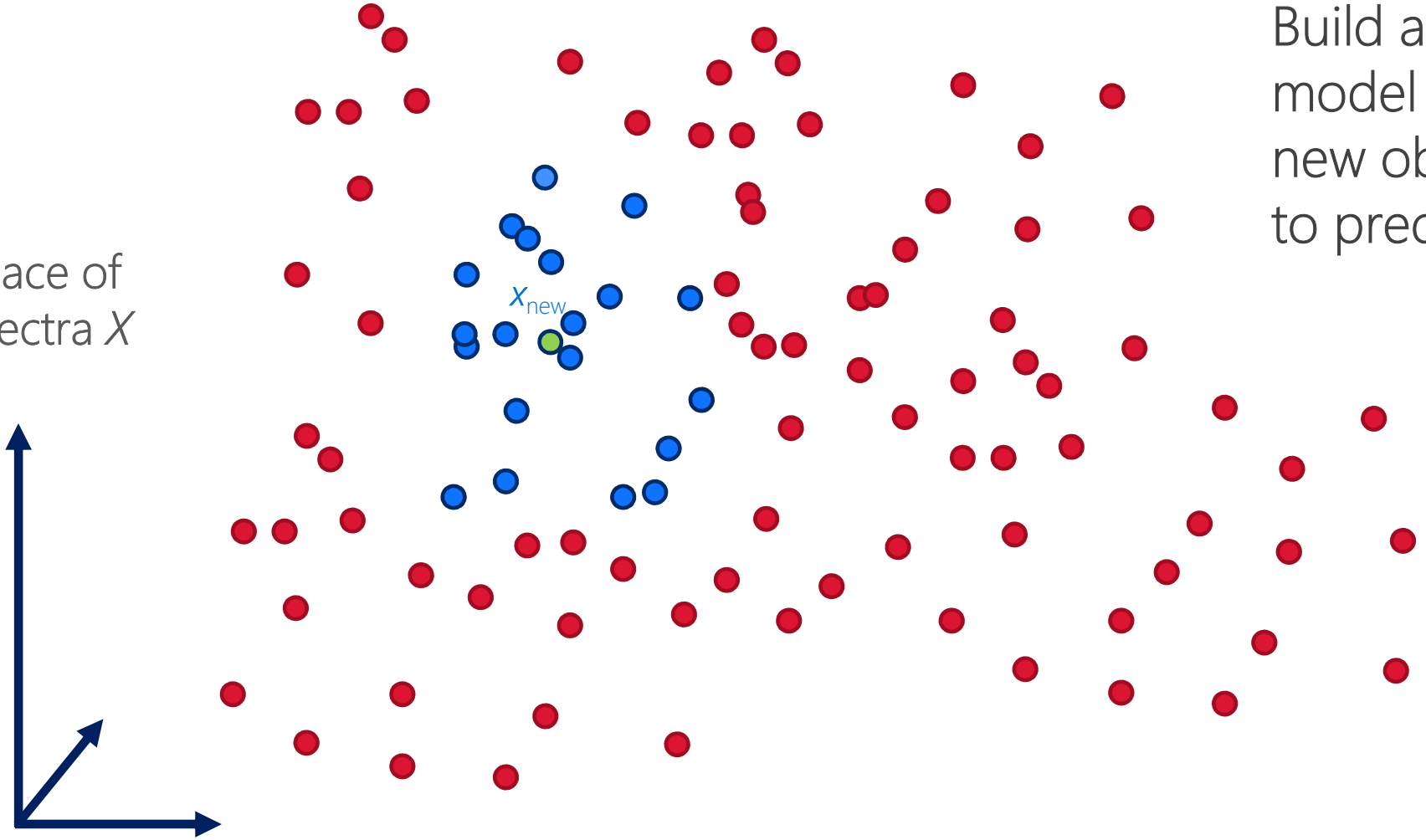
Prediction

y_{new}


1 spectrum x
of the training
Space of
spectra X
New spectrum
to predict
x new

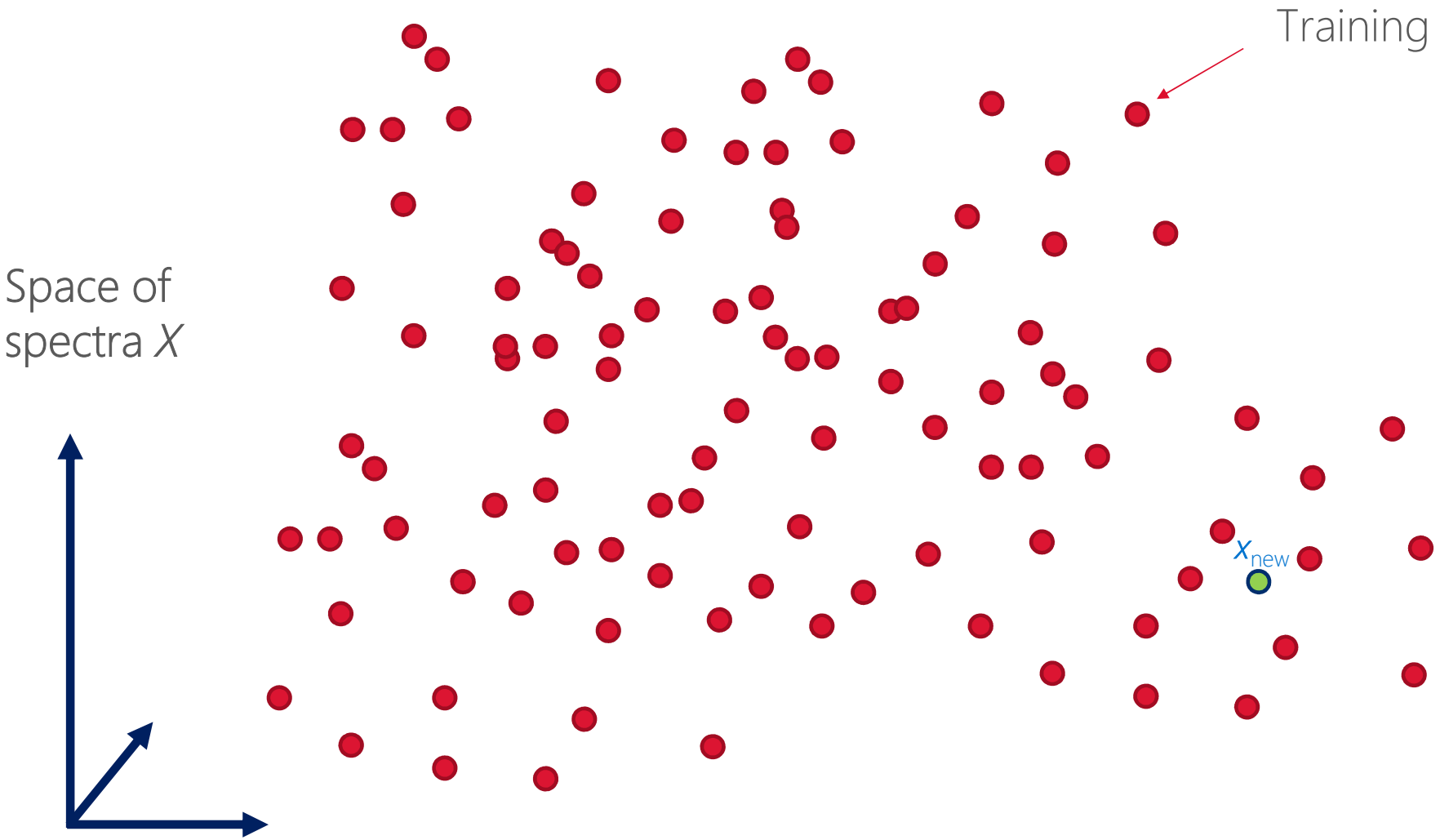

we
expect a
linear
relation
between
X and y
Space of
spectra X
Selection of a
neighborhood
of x_new
(k nearest)
x_new
PLSR
Prediction y_new

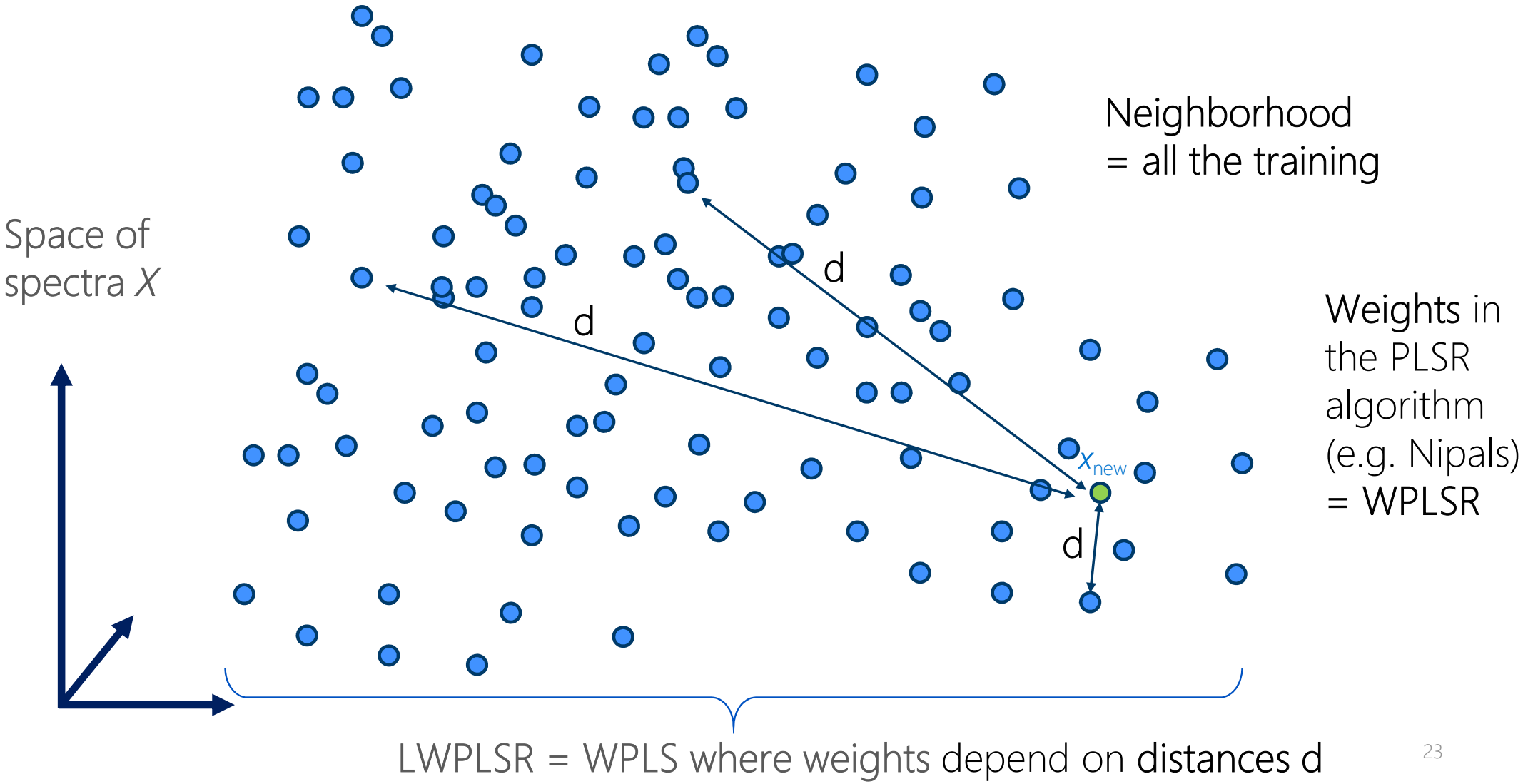
Space of spectra X



Build a new model for each new observation to predict

2. LWPLSR (common routines in Matlab, Python)




Space of spectra X
Neighborhood = all the training
d
d
d
x_{new}
Weights in the PLSR algorithm (e.g. Nipals) = WPLSR
LWPLSR = WPLS where weights depend on distances d

WPLSR

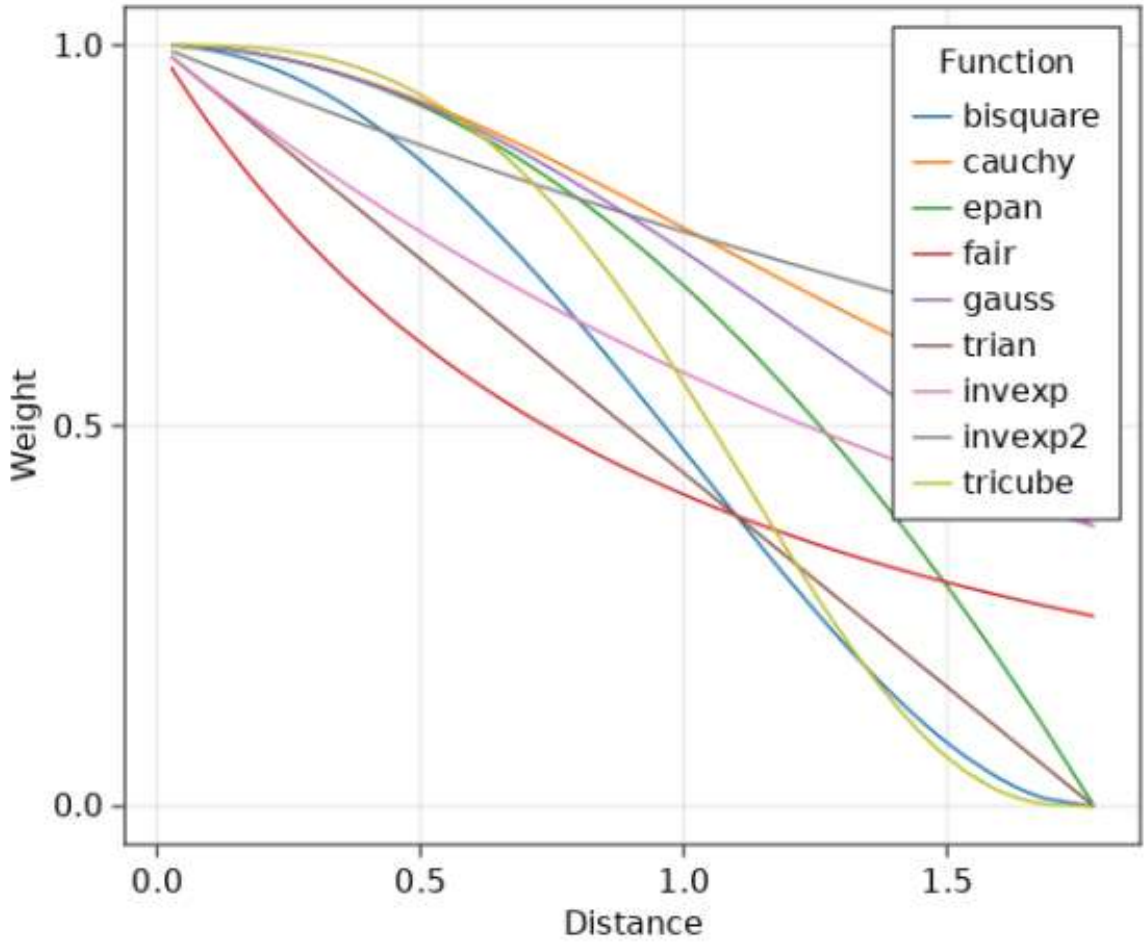
- Means, covariances, regressions are computed in row-metric $\{w_1, \ldots, w_n\}$

	PLSR	WPLSR
• max $\mathrm{Cov}(t_j, y)^2$	$\boldsymbol{t_j'\left(\frac{I}{n}\right)y}$	$\boldsymbol{t_j'Wy}$
	$= \sum_{i=1}^{n} \frac{1}{n} t_{ij} y_i$	$= \sum_{i=1}^{n} w_i t_{ij} y_i$
• $\mathrm{argmin}_{\beta} \lVert y - T\beta \rVert^2$	$\sum_{i=1}^{n} \frac{1}{n} (y_i - \boldsymbol{t_i'\beta})^2$	$\sum_{i=1}^{n} w_i (y_i - \boldsymbol{t_i'\beta})^2$

LWPLSR

Weights $\{w_1, \ldots, w_n\}$ depend on distances $\{d_1, \ldots, d_n\}$

Many choices of weight functions



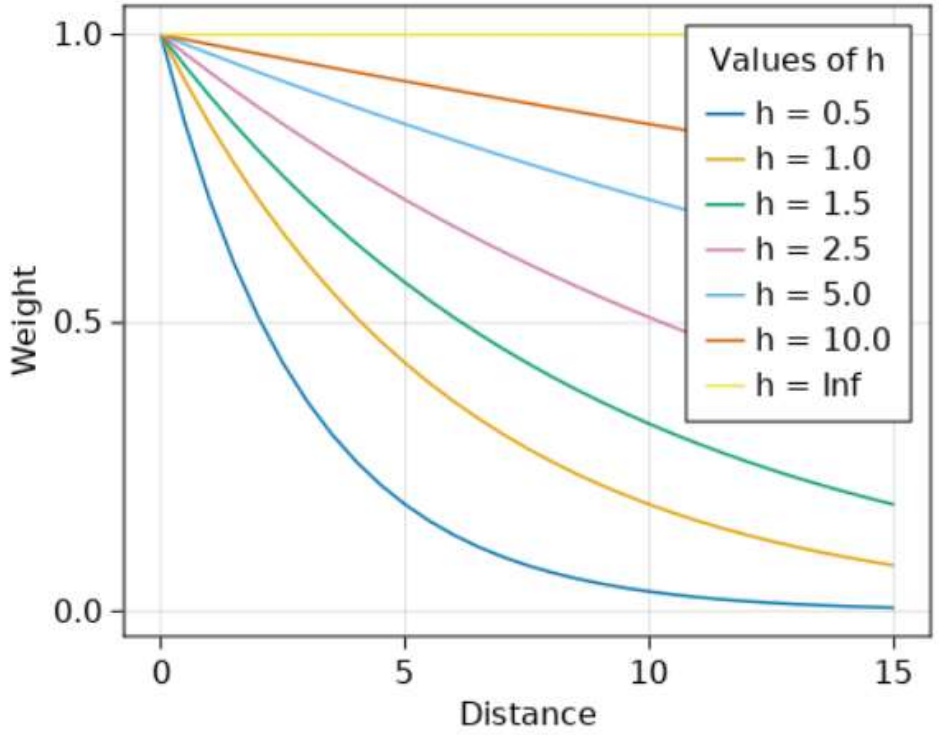
Ex: function wdist of Jchemo = versatile "Gaussian" weight function

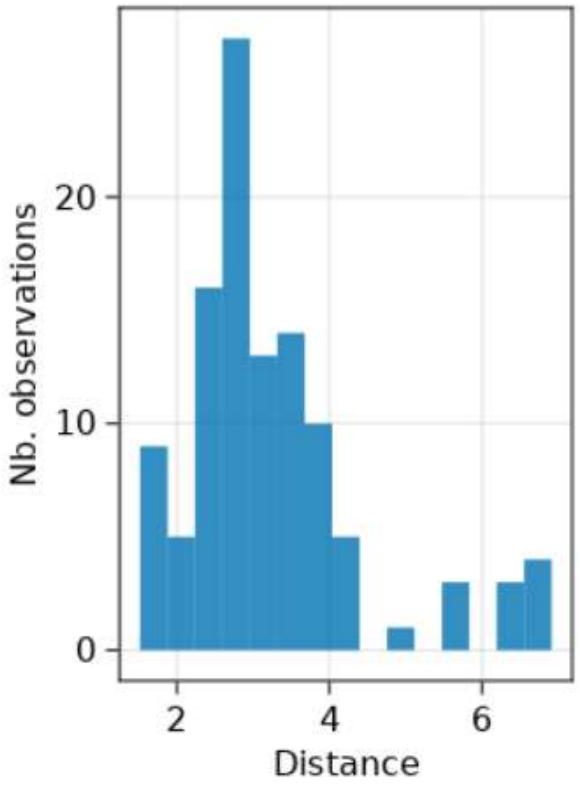
Adaptation from
Kim S, Kano M, Nakagawa H, Hasebe S. Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection. Int J Pharm. 2011;421(2):269-274. https://doi.org/10.1016/j.ijpharm.2011.10.007

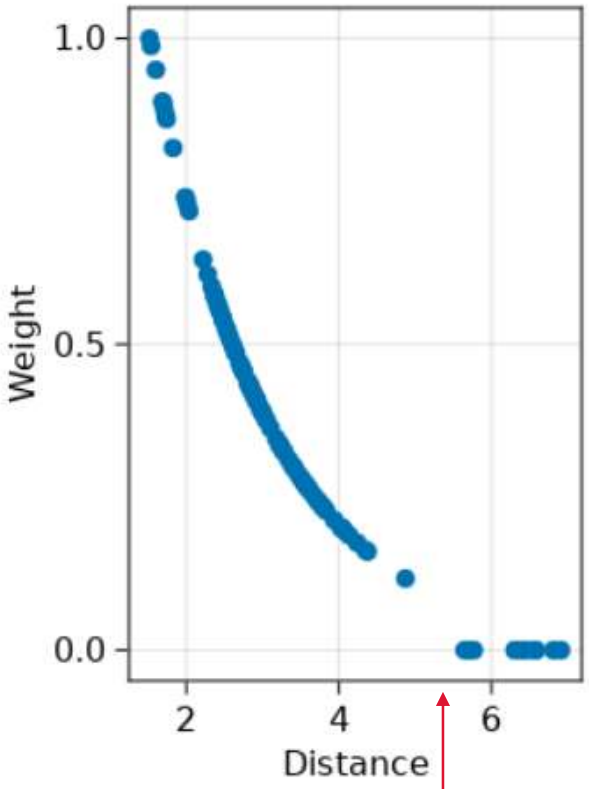
j = 1, ..., k neighbors

$$w_j = exp\frac{-d_j}{h \times mad\{d_1, \ldots, d_k\}}$$

$w_j = w_j$ / maximum$\{w_1, \ldots, w_k\}$

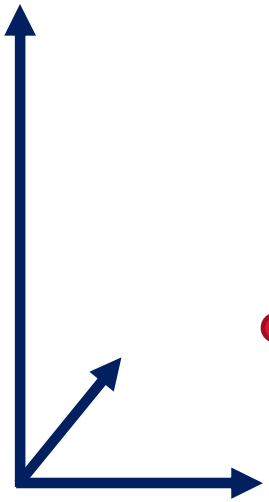


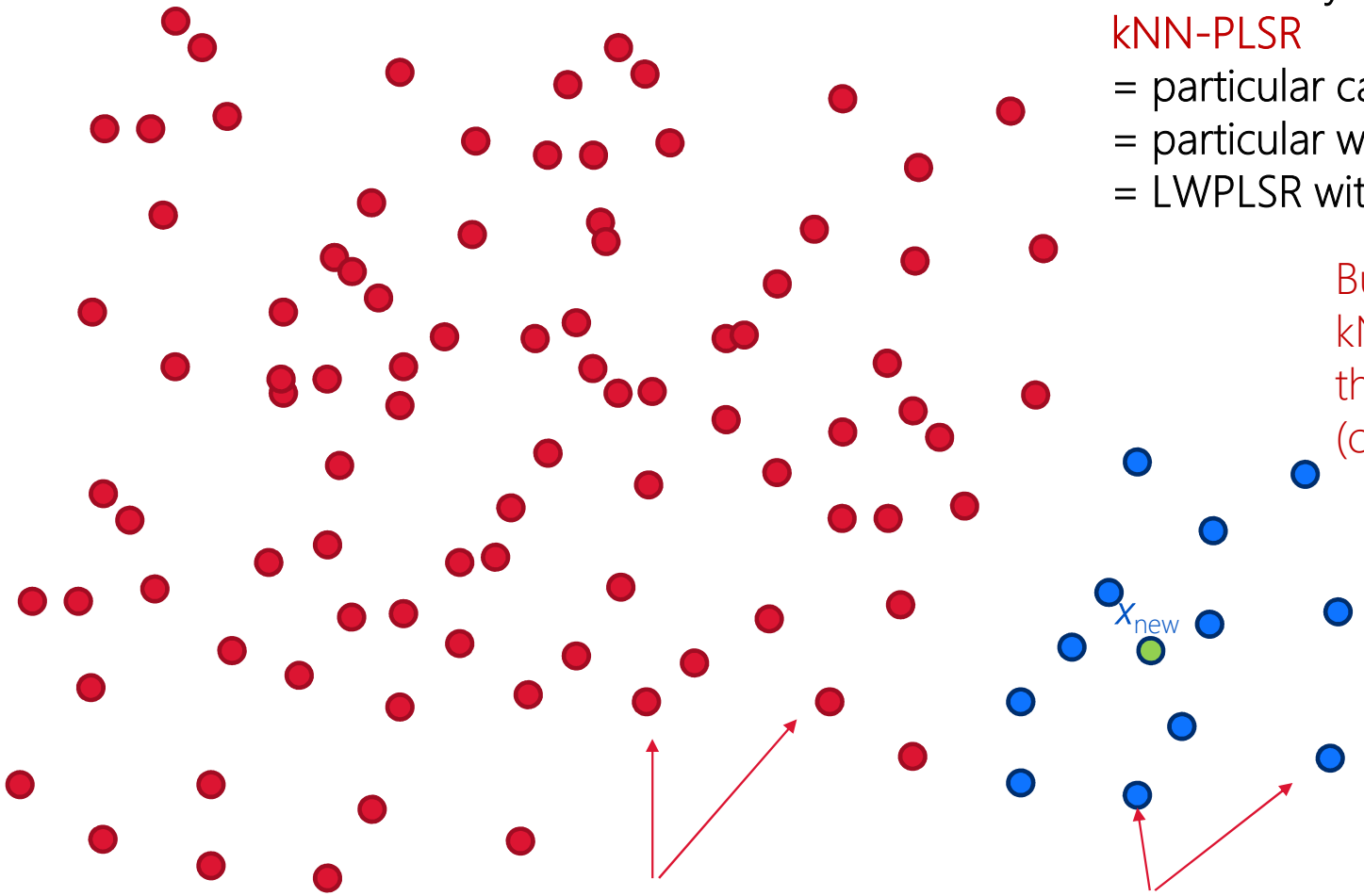




cutoff
= median(d) + 4 × mad(d)

Space of spectra *X*





Weights = 0

Weights = 1 / *k*

Remark

Theoretically

kNN-PLSR

= particular case of LWPLSR

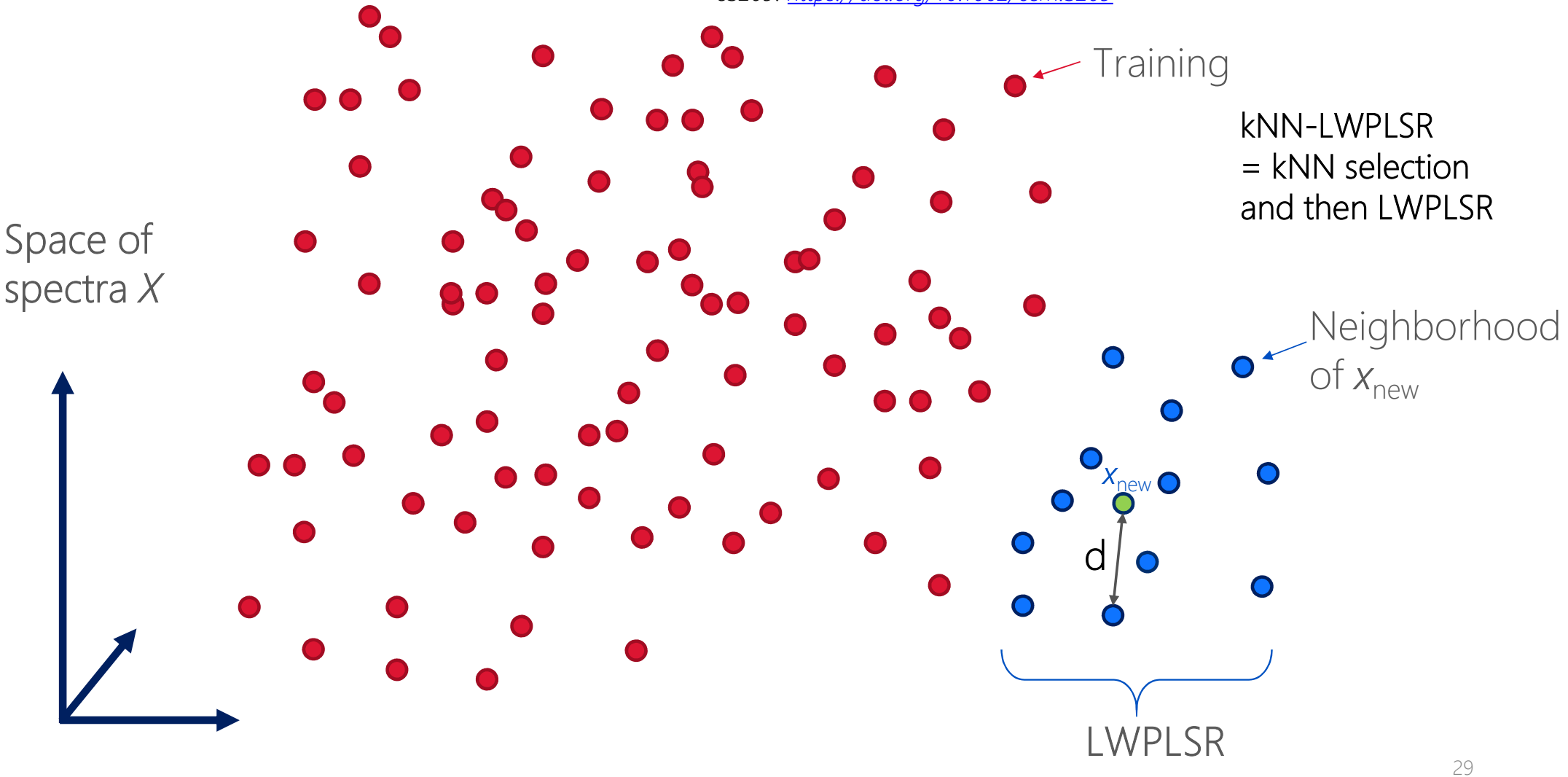
= particular weight function

= LWPLSR with discrete weights

But the preliminary kNN selection makes the **algorithm faster** (computation time)

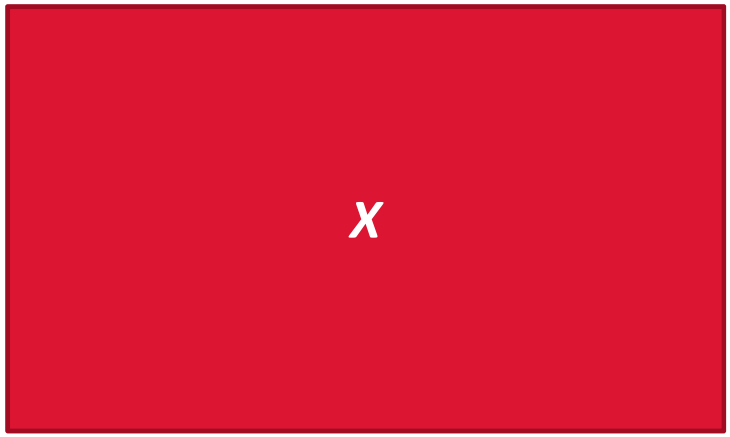
3. kNN-LWPLSR (mixing 1+2)

Lesnoff, M., Metz, M., Roger, J.-M., 2020. Comparison of locally weighted PLS strategies for regression and discrimination on agronomic NIR data. Journal of Chemometrics n/a, e3209. https://doi.org/10.1002/cem.3209



Algorithm	Neighbors selection	WPLS algo
kNN-PLSR	Yes	No
LWPLSR	No	Yes
kNN-LWPSR	Yes	Yes

1)

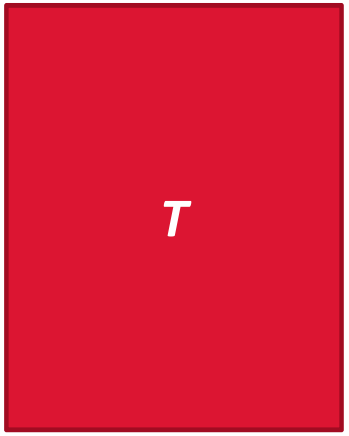


- Euclidean distance
- Correlation
- Any dissimilarity

Other option: **Dimension reduction**

- PCA
- Random projections
- PLS
- Etc.

2)



- Euclidean distance
- Correlation
- Mahalanobis
- Any dissimilarity

Local PLSR is different from clusterwise PLSR

- Preda, C., Saporta, G., 2005.
  Clusterwise PLS regression on a stochastic process.
  Computational Statistics & Data Analysis 49, 99–108.
  https://doi.org/10.1016/j.csda.2004.05.002

- "Simca" methods etc.

Clusterwise PLSR

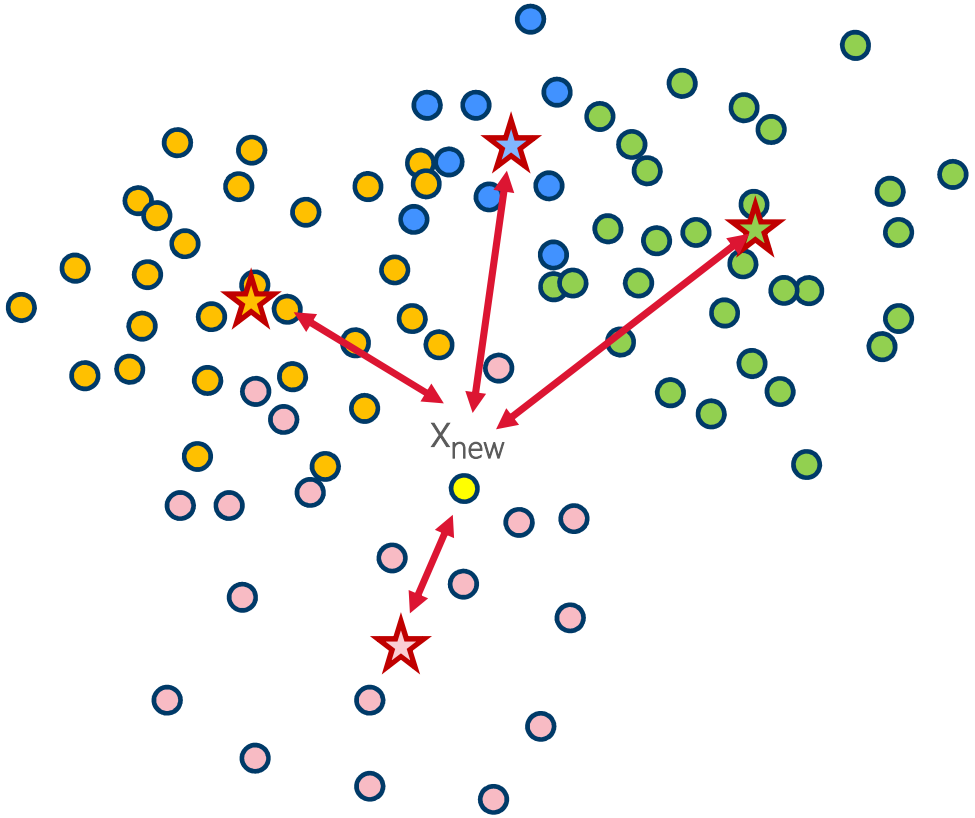
1. Clustering: a priori or unsupervised (e.g. kmeans)

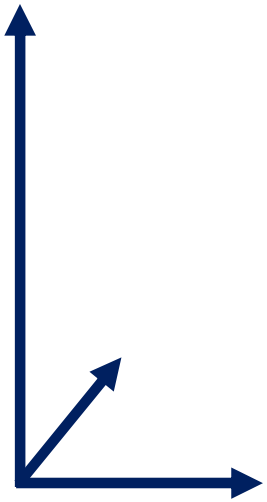
2. Build one PLSR model per cluster

3. Predict the new observation using the model fitted on the nearest cluster

Faster than kNN-LWPLSR but in general less performant

Space of spectra *X*





4 clusters

1. Build one PLSR model per cluster
2. Compute the distance between the observation and the center of the clusters
3. Predict the observation from the model of the nearest cluster

Illustrations

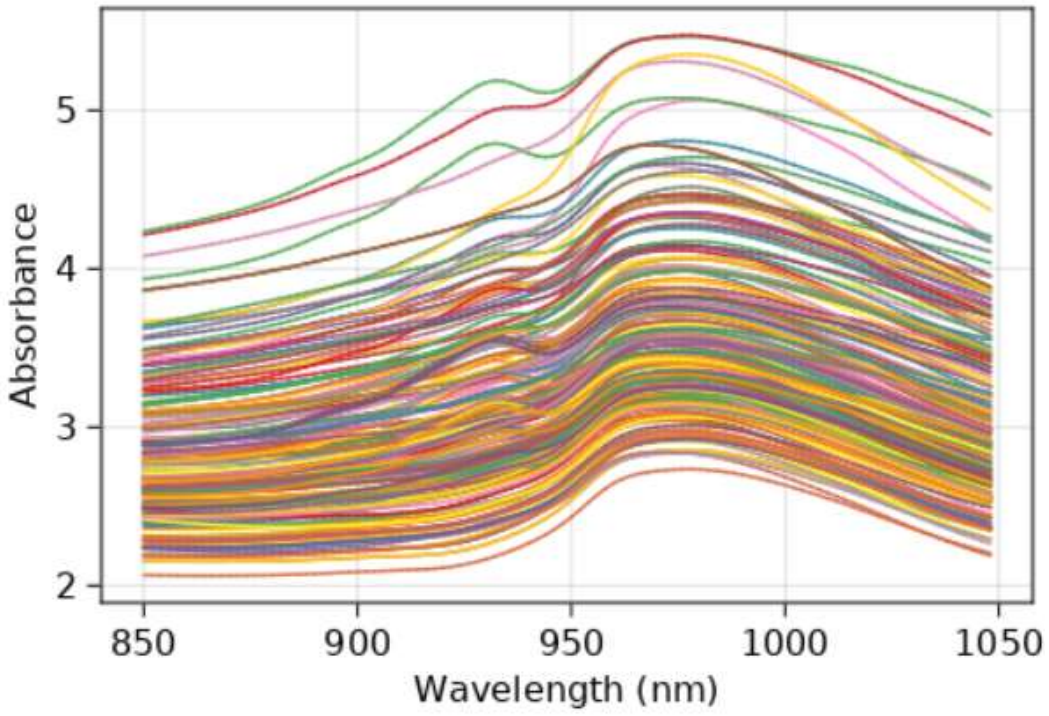
Illustration 1. Tecator NIR data

http://lib.stat.cmu.edu/datasets/tecator

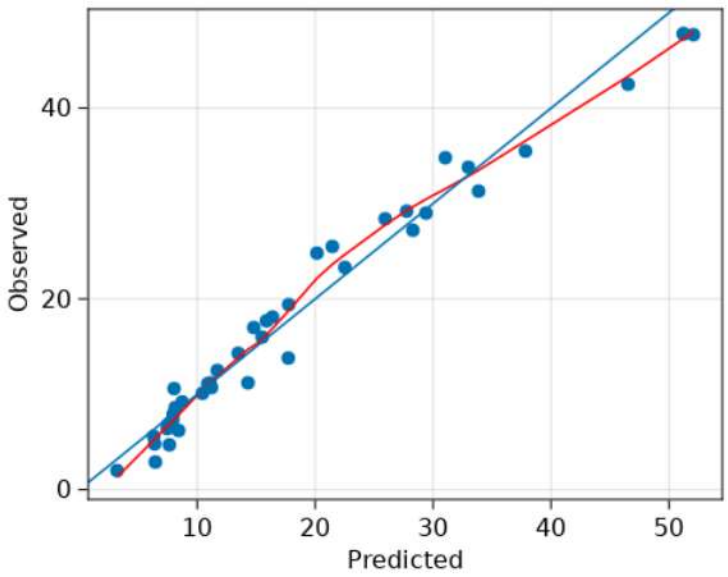
Data recorded on a Tecator Infratec Food and Feed Analyzer (wavelength range 850 - 1050 nm).

Samples of finely chopped pure meat
y = moisture, fat and protein contents.

$n = 215$



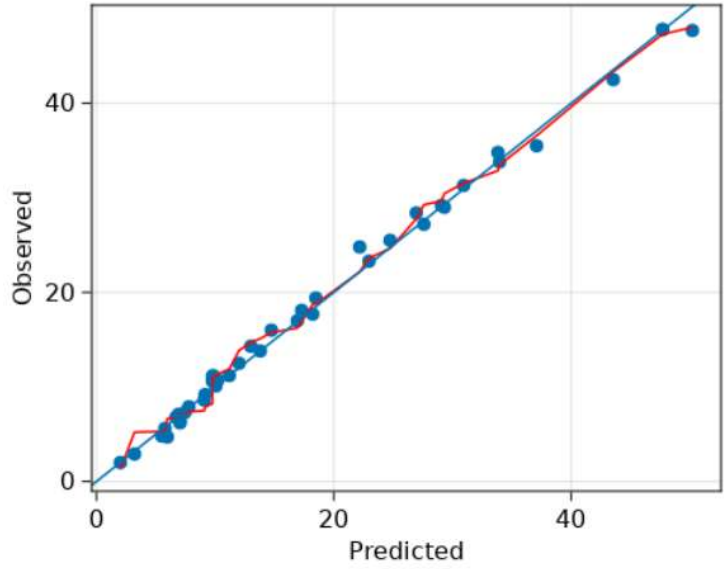
y = fat content
*n*train = 172
*n*test = 42



PLSR 9 LVs

$RMSEP_{Test}$ = 1.99

Non-linearities are suspected



LWPLSR

$RMSEP_{Test}$ = .89

Illustration 2.

Lesnoff, M., Metz, M., Roger, J.-M., 2020. Comparison of locally weighted PLS strategies for regression and discrimination on agronomic NIR data. Journal of Chemometrics

3 NIR datasets
Forages and feed

TABLE 1 The three data sets (PROT, DMD, and BARL) used for methods comparisons

Data Set	N	Response *y*	Type of Material	Source
PROT	4075	Protein content (mean = 31.9; min = 2.8; max = 76.6, sd = 20.3)	Animal feed, rapeseed, corn gluten, grass silage, soya, wheat, milk powder, maize, sun flower	CRA-W, Belgium
DMD	1148	In vitro dry matter enzymatic digestibility (mean = 52.2; min = 9.9; max = 95.0, sd = 16.8)	Cereal, grass, legume, tree, grassland mixing forages (mainly from tropical drylands)	CIRAD, France
BARL	7515	Absence/presence of barley (proportion of samples with barley = 0.76)	Commercial compound feeds with variable formulations and animal destinations	ETSIAM,[20] Univ Cordoba, Spain

FIGURE 1 Principal component analysis (PCA) score plots (components 1 and 2) of the preprocessed spectra for the three data sets (PROT, DMD, and BARL)

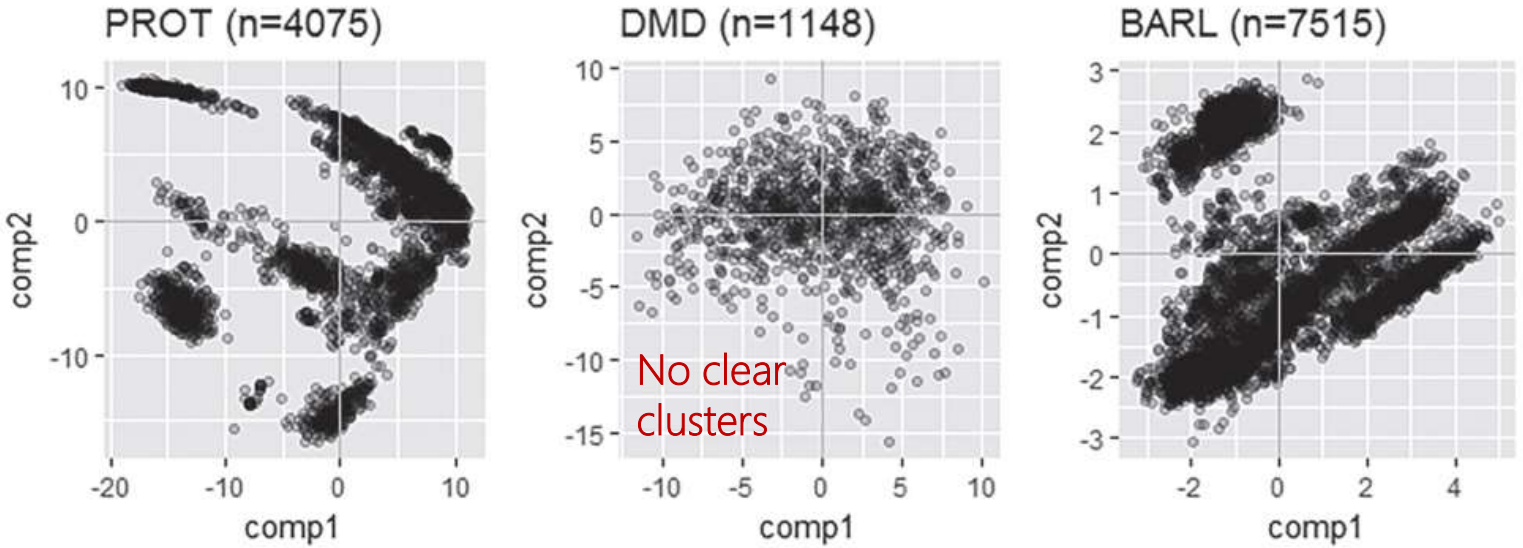


TABLE 2 Prediction error rates for the three data sets (PROT, DMD, and BARL)

(a) PROT

Method	VAL1				VAL2				VAL3				Mean	
	h	k	n_{comp}	RMSEP	h	k	n_{comp}	RMSEP	h	k	n_{comp}	RMSEP	RMSEP	
PLSR	–	–	17	1.45	–	–	19	1.38	–	–	17	1.42	1.41	PLSR
KNN-L	–	100	9	0.77	–	100	9	0.77	–	100	9	0.76	0.76	
LW	0.4	–	12	0.75	0.5	–	15	0.72	0.4	–	12	0.80	0.75	LWPLSR
KNN-LW	0.9	200	11	0.70	2.0	200	11	0.70	1.0	200	10	0.75	0.72	

(b) DMD

Method	VAL1				VAL2				VAL3				Mean	
	h	*k*	n_{comp}	RMSEP	*h*	*k*	n_{comp}	RMSEP	*h*	*k*	n_{comp}	RMSEP	RMSEP	
PLSR	–	–	12	5.37	–	–	6	6.65	–	–	12	5.14	5.68	PLSR
KNN-L	–	50	5	4.95	–	50	4	5.58	–	50	4	5.03	5.19	
LW	0.4	–	7	4.67	0.4	–	8	4.69	0.5	–	6	4.24	4.53	LWPLSR
KNN-LW	0.8	600	7	4.73	0.6	600	8	4.78	0.8	600	6	4.19	4.57	

(c) BARL

Method	VAL1				VAL2				VAL3				Mean	
	h	*k*	n_{comp}	ERRP	*h*	*k*	n_{comp}	ERRP	*h*	*k*	n_{comp}	ERRP	ERRP	
PLSR	–	–	14	0.147	–	–	11	0.172	–	–	14	0.173	0.164	PLSDA
KNN-L	–	100	11	0.077	–	50	8	0.064	–	50	5	0.068	0.069	
LW	0.5	–	10	0.041	0.3	–	7	0.035	0.4	–	9	0.036	0.037	LWPLSDA
KNN-LW	1.0	200	7	0.043	1.0	400	9	0.032	0.8	300	8	0.035	0.037	

Illustration 3. Mango data NIR + Dry matter (DM) content prediction

Anderson et al 2020, 2021

https://data.mendeley.com/datasets/46htwnp833/1

11691×9 DataFrame

Row	set	region	date	type	cultivar	temp	season	pop	dm
	String	String	String	String	String	String	Int64	Int64	Float64
1	Cal	NT	2/10/2015	Hard Green	Caly	Mid	1	2	16.7925
2	Cal	NT	2/10/2015	Hard Green	Caly	Mid	1	2	16.7925
3	Cal	NT	2/10/2015	Hard Green	Caly	Mid	1	2	16.071
4	Cal	NT	2/10/2015	Hard Green	Caly	Mid	1	2	16.071
5	Cal	NT	2/10/2015	Hard Green	Caly	Mid	1	2	16.394
6	Cal	NT	2/10/2015	Hard Green	Caly	Mid	1	2	16.394
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11686	Val Ext	QLD	25/01/2019	Ripen	HG	No	4	112	18.9869
11687	Val Ext	QLD	25/01/2019	Ripen	HG	No	4	112	18.9869
11688	Val Ext	QLD	25/01/2019	Ripen	HG	No	4	112	19.7118
11689	Val Ext	QLD	25/01/2019	Ripen	HG	No	4	112	19.7118
11690	Val Ext	QLD	25/01/2019	Ripen	HG	No	4	112	19.4786
11691	Val Ext	QLD	25/01/2019	Ripen	HG	No	4	112	19.4786

11679 rows omitted

Dim1 \ Dim2	2015	2016	2017	2018	2019
Cal	2869	1207	2529	808	0
Tuning	514	530	1550	236	0
Val Ext	0	0	0	1130	318

Anderson, N.T.; Walsh, K.B.; Flynn, J.R.; Walsh, J.P. Achieving Robustness across Season, Location and Cultivar for a NIRS Model for Intact Mango Fruit Dry Matter Content. II. Local PLS and Nonlinear Models. Postharvest Biology and Technology 2021, 171, 111358, doi:10.1016/j.postharvbio.2020.111358.

Fast calibration of kNN-LWPLSR (Jchemo)

```
julia> metric = "mahal"
       nlvdis = 20 ; h = 5 ; k = 500 ; nlv = 12
       fm = lwplsr(Xptrain, ytrain; nlvdis = nlvdis,
           metric = metric, h = h, k = k, nlv = nlv) ;
       pred = Jchemo.predict(fm, Xptest).pred ;
       rmsep(pred, ytest)
1x1 Matrix{Float64}:
 0.8593636054514092

julia> cor2(pred, ytest)
1x1 Matrix{Float64}:
 0.9035426752267037
```

Results on Val ext.

Table 3

Dry matter content prediction statistics of models based on calibration and tuning data sets combined (n = 10,243 spectra) used in prediction of an independent validation set (n = 1448 spectra). Fearns significance test on SEP was employed for comparisons to LPLS. Difference level to LPLS (underlined) SEP is indicated, either to $p < 0.001$ (***) or 0.05 (*). Results not significantly different ($p < 0.05$) in terms of SEP to the LPLS result are shown in bold. There were no significant differences in similar comparisons of Bias.

Model	R^2	RMSEP (%)	SEP (%)	Bias (%)
Individual Models				
LOVR	0.900	0.881	0.855*	−0.210
LPLS	*0.903*	*0.887*	*0.846*	*−0.232*
ANN	0.889	0.892	0.890***	0.056
LPLS-cv	0.899	0.892	0.863***	−0.224
LPLS-S	0.895	0.896	0.873***	−0.199
GPR	0.889	0.898	0.892***	−0.103
MBL	0.890	0.903	0.901***	−0.059
DR-LGB	0.872	0.976	0.969***	−0.115
PLS	0.859	1.014	1.012***	−0.065
LOCAL	0.857	1.023	1.021***	−0.065
SVR	0.859	1.048	1.022***	−0.229
Cubist	0.833	1.135	1.124***	−0.156
Ensemble Models				
ANN_GPR_LPLS-S Ensemble	**0.904**	**0.839**	**0.835**	**−0.082**
Hone-E	**0.902**	**0.850**	**0.847**	**0.078**
DR-E	0.873	0.963	0.956***	−0.120

Models used: Local Optimized by Variance Regression (LOVR), Local Partial Least Squares (LPLS), Artificial Neural Network (ANN), Local Partial Least Squares with latent variables selected by cross-validation (LPLS-cv), Local Partial Least Squares Scores (LPLS-S), Gaussian Process Regression (GPR), Memory Based Learner (MBL), DataRobot Light Gradient Boosting (DR-LGB), Partial Least Squares (PLS), Shenk Local (LOCAL), Support Vector Regression (SVR), Cubist, Hone Create Stacked Ensemble (Hone-E), and DataRobot ElasticNet Ensemble (DR-E).

Illustration 4.
Discrimination of 11 varieties of rice harvested in northern Italy

Bevilacqua, M.; Marini, F. Local Classification: Locally Weighted–Partial Least Squares-Discriminant Analysis (LW–PLS-DA). Analytica Chimica Acta 2014, 838, 20–30, doi:10.1016/j.aca.2014.05.057.

Table 7
Comparison among LW–PLS-DA, global PLS-DA, k_{NN} and kernel-PLS-DA on the four data sets: model parameters and classification accuracy on the external test sets.

	LW–PLS-DA			Global PLS-DA		k_{NN}		Kernel-PLS-DA		
	% Overall correct classification	Nr NN	Nr LV	% Overall correct classification	Nr LV	% Overall correct classification	k	% Overall correct classification	σ	Nr LV
Data set "sphere"	100	50	2	62	1	99.8	2	100	50	9
Data set "cross"	99.2	54	1	33.7	3	99.3	3	98.8	0.5	15
Data set "linear"	99.5	124	3	76.7	3	99.0	5	99.0	0.5	13
Data set "rice"	93.1	56	8	57.5	7	11.4	19	12.5	0.2	15

?

?

Some details

Pedagogical way to program (centered) WPLS

(1) Transform the data

```
xmeans = colmean(X, weights)
ymeans = colmean(Y, weights)
center!(X, xmeans)
center!(Y, ymeans)
W = Diagonal(weights)        ## Metric
sqrtw = sqrt.(weights)
sqrtW = Diagonal(sqrtw)      ## here use sparse matrix coding
## Can be costly for large matrices
X .= sqrtW * X
Y .= sqrtW * Y
## end
```

(2) Then => use common PLSR iterations on X and Y
(metric W disappears)

(3) Finally, come back to the good scale for the outputs : * 1 ./ sqrtW

Alternative: metric W can be apparent in the PLSR iterations

Little more "complex" to write but often less computation time

Ex: Function plskern of packages rchemo (R) and Jchemo (Julia)

"Improved kernel algorithm #1"

Dayal, B.S., MacGregor, J.F., 1997. Improved PLS algorithms. Journal of Chemometrics 11, 73-85

An example of benchmark

```
n = 10^6 ; p = 500  ; q = 10
X = rand(n, p) ; Y = rand(n, q)
nlv = 25

julia> @time plskern(X, Y[:, 1]; nlv = nlv) ;
       @time plskern(X, Y[:, 1:10]; nlv = nlv) ;
  6.968429 seconds (44 allocations: 3.979 GiB, 0.57% gc time)
  7.934715 seconds (220 allocations: 4.181 GiB, 7.97% gc time)
```

Some history around LWPLSR (not exhaustive)

- *Naes, T., Isaksson, T., Kowalski, B., 1990. Locally weighted regression and scatter correction for near-infrared reflectance data. Analytical Chemistry 664–673.* ➔ "LWR": Not exactly a local PLSR (Use Cleveland et al.). Global PCA, Mahalanobis distance, Tricube function, lwr on neighborhood global PCA scores

- *Aastveit AH, Marum P. Near-infrared reflectance spectroscopy: different strategies for local calibrations in analyses of forage quality. Appl Spectrosc. 1993;47(4):463-469. https://doi.org/10.1366/0003702934334912* ➔ Variant of LWR. PCA scores are recomputed on the neighborhood

- *Schaal S, Atkeson CG, Vijayakumar S. Scalable techniques from nonparametric statistics for real time robot learning. Appl Intell. 2002;17 (1):49-60. https://doi.org/10.1023/A:1015727715131* ➔ WPLS Nipals

- *Sicard E, Sabatier R. Theoretical framework for local PLS1 regression, and application to a rainfall data set. Comput Stat Data Anal. 2006;51(2):1393-1410. https://doi.org/10.1016/j.csda.2006.05.002* ➔ LWPLSR univariate

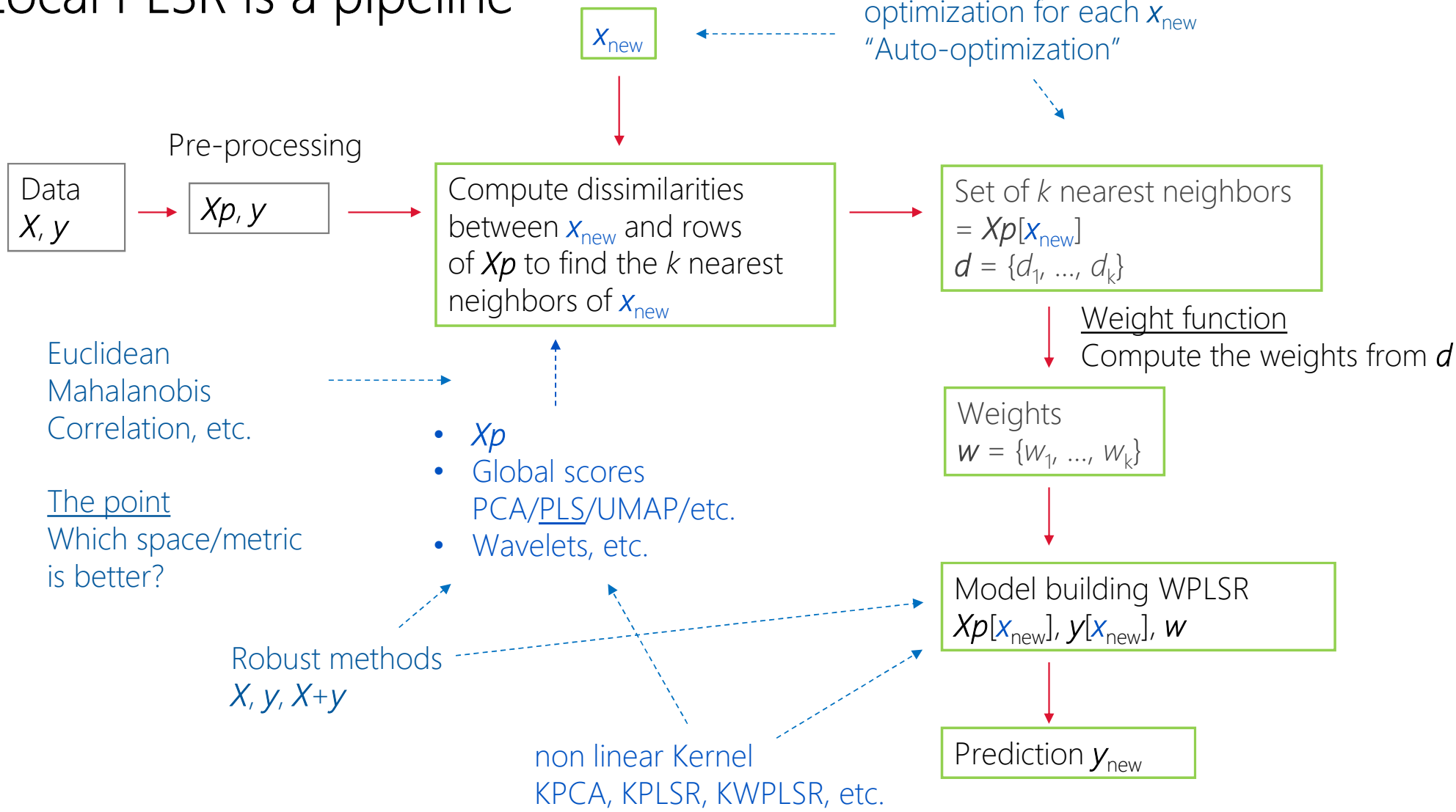
- *Kim S, Kano M, Nakagawa H, Hasebe S. Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection. Int J Pharm. 2011;421(2):269-274. https://doi.org/10.1016/j.ijpharm.2011.10.007* ➔ LWPLSR Nipals, multivariate

- *Bevilacqua, M., Marini, F., 2014. Local classification: Locally weighted–partial least squares-discriminant analysis (LW–PLS-DA). Analytica Chimica Acta 838, 20–30. https://doi.org/10.1016/j.aca.2014.05.057* ➔ LWPLSDA

- *Lesnoff, M., Metz, M., Roger, J.-M., 2020. Comparison of locally weighted PLS strategies for regression and discrimination on agronomic NIR data. Journal of Chemometrics n/a, e3209. https://doi.org/10.1002/cem.3209* ➔ kNN-LWPLSR/DA, Kernel algorithm 1 (Dayal et al.)

Many possible extensions & variants

Local PLSR is a pipeline



Local PLS Discrimination

kNN-LWPLSR-DA

PLS-LDA

PLS-QDA

etc.

Bevilacqua, M., Marini, F., 2014. Local classification: Locally weighted–partial least squares-discriminant analysis (LW–PLS-DA). Analytica Chimica Acta 838, 20–30. https://doi.org/10.1016/j.aca.2014.05.057

Lesnoff, M., Metz, M., Roger, J.-M., 2020. Comparison of locally weighted PLS strategies for regression and discrimination on agronomic NIR data. Journal of Chemometrics n/a, e3209. https://doi.org/10.1002/cem.3209

Model averaging or stacking

$$\hat{y}_{\text{new}} = \theta_0 \hat{y}_{\text{new},nlv=0} + \theta_1 \hat{y}_{\text{new},nlv=1} + \cdots + \theta_a \hat{y}_{\text{new},nlv=a}$$

- averaging: $0 \leq \theta_r \leq 1 \qquad \sum_{r=0}^{a} \theta_r = 1$
- stacking: $\{\theta_0, \theta_1, \ldots, \theta_a\}$ = outputs of a regression model

kNN-LWPLSR-AVG/STACK

Shenk, J., Westerhaus, M., Berzaghi, P., 1997. Investigation of a LOCAL calibration procedure for near infrared instruments. Journal of Near Infrared Spectroscopy 5, 223. https://doi.org/10.1255/jnirs.115 "LOCAL" = kNN-PLSR-AVG

Lesnoff, M., Andueza, D., Barotin, C., Barre, P., Bonnal, L., Fernández Pierna, J.A., Picard, F., Vermeulen, P., Roger, J.-M., 2022. Averaging and Stacking Partial Least Squares Regression Models to Predict the Chemical Compositions and the Nutritive Values of Forages from Spectral Near Infrared Data. Applied Sciences 12, 7850. https://doi.org/10.3390/app12157850

With big data?

Brute-force (=usual) kNN search can be very time consuming

Building indexes with hashing algorithms

- Random projections Ex: Parsketch
- iSaks

etc.

Metz, M., Lesnoff, M., Abdelghafour, F., Akbarinia, R., Masseglia, F., Roger, J.-M., 2020. A "big-data" algorithm for KNN-PLS. Chemometrics and Intelligent Laboratory Systems 203, 104076. https://doi.org/10.1016/j.chemolab.2020.104076

Ryckewaert, M., Metz, M., Héran, D., George, P., Grèzes-Besset, B., Akbarinia, R., Roger, J.-M., Bendoula, R., 2021. Massive spectral data analysis for plant breeding using parSketch-PLSDA method: Discrimination of sunflower genotypes. Biosystems Engineering 210, 69–77. https://doi.org/10.1016/j.biosystemseng.2021.08.005

Zhang, X., Wei, C., Song, Z., 2020. Fast Locally Weighted PLS Modeling for Large-Scale Industrial Processes. Ind. Eng. Chem. Res. 59, 20779–20786. https://doi.org/10.1021/acs.iecr.0c03932

"Hot spots"

- How to define/select the space/neighborhood so that the relation between X and y is as linear as possible?

  This is what we expect when we do the PLSR on the neighbors

- "Relevancy" of a neighbor

Conclusions

Pros

- Simple to understand: uses linear well-know tools (PLSR)
- Simple to optimize (for the simplest pipelines)
- Efficient in many common situations

Cons

- Not one single model
- Predictions can be time consuming when very large set of new observations to predict
- Hazardous for far extrapolations

❑ Not clear how to find versatile and always optimal approach for building neighborhood

❑ Comparative performances of the pipelines are data dependent

➔ My own strategy: to use "omnibus" models, not optimal but never very far …