kNN-LWPLSR

matthieu.lesnoff@cirad.fr

Cirad, UMR Selmet

kNN : k Nearest Neighbors

LWPLSR : Locally Weighted Partial Least Squares Regression

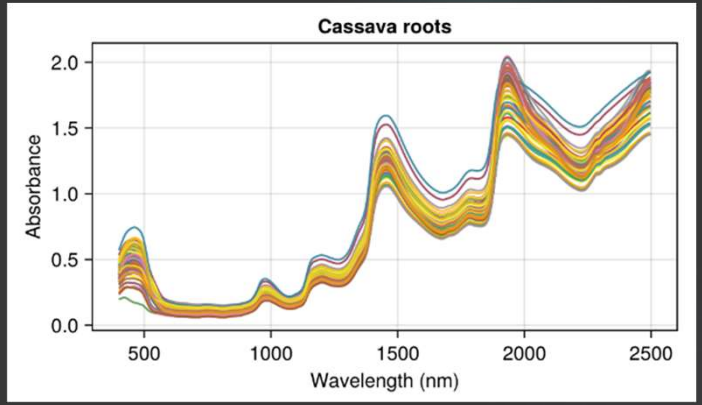
(LWPLSDA ⇒ Discrimination)

Algorithm useful when non-linearity between X and Y (data heterogeneity, etc.)

- Globally non linear
- Locally linear

kNN-LWPLSR

- Very performant for NIR data



Available pipelines

Journal of CHEMOMETRICS

RESEARCH ARTICLE

Comparison of locally weighted PLS strategies for regression and discrimination on agronomic NIR data

Matthieu Lesnoff, Maxime Metz, Jean-Michel Roger

First published: 16 January 2020 | https://doi.org/10.1002/cem.3209 | Citations: 30

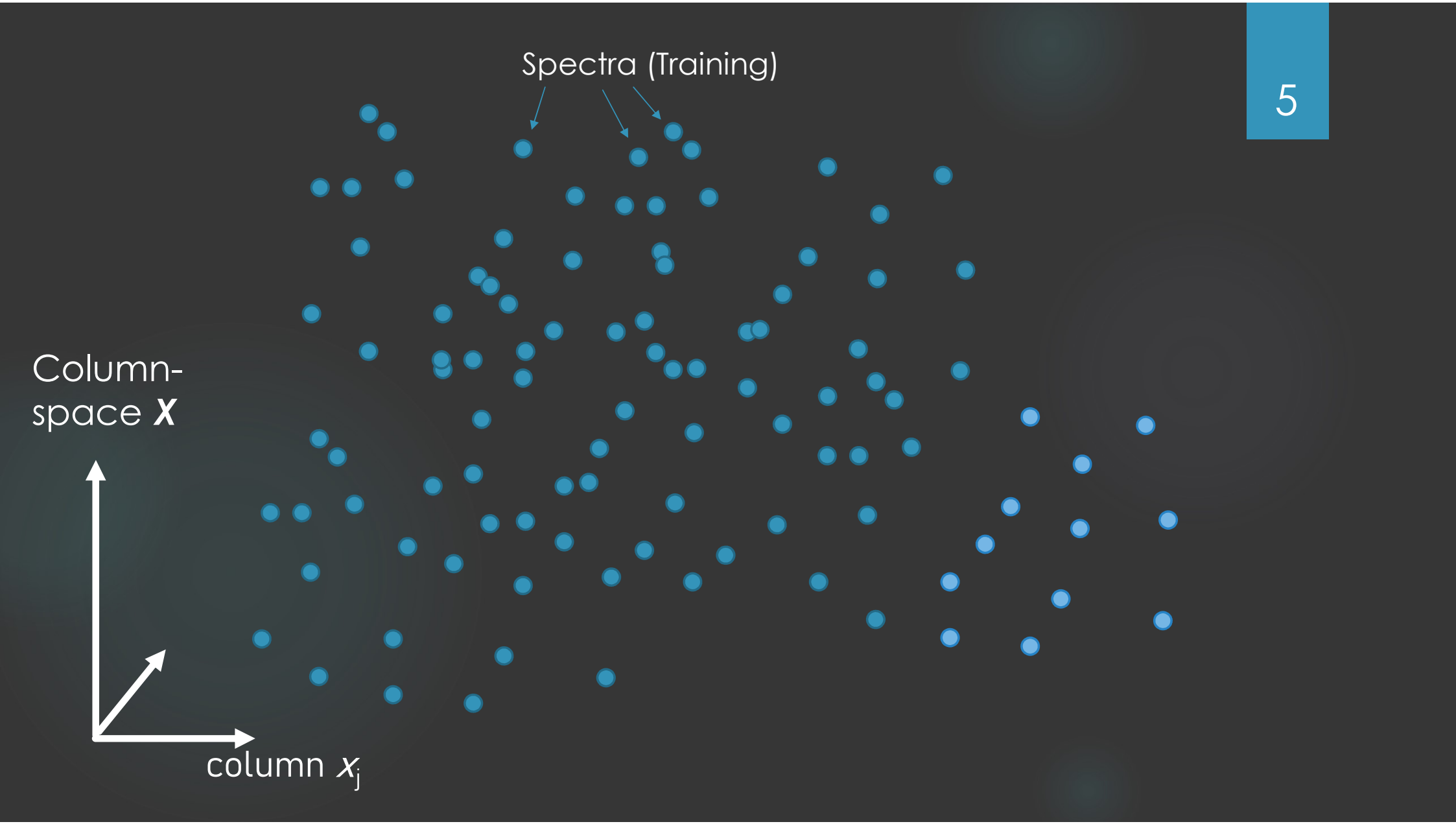
ELSEVIER

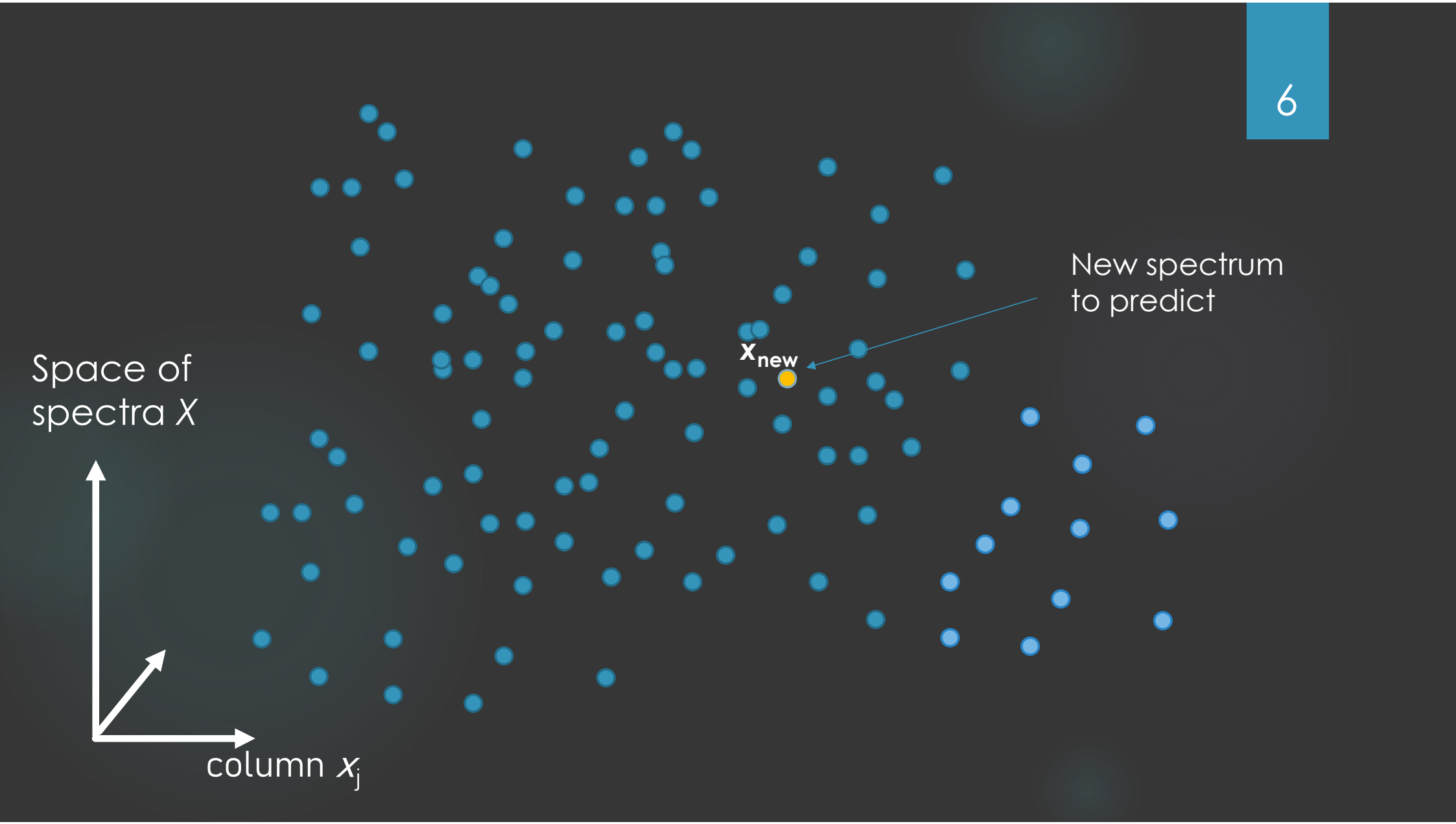
Chemometrics and Intelligent Laboratory Systems

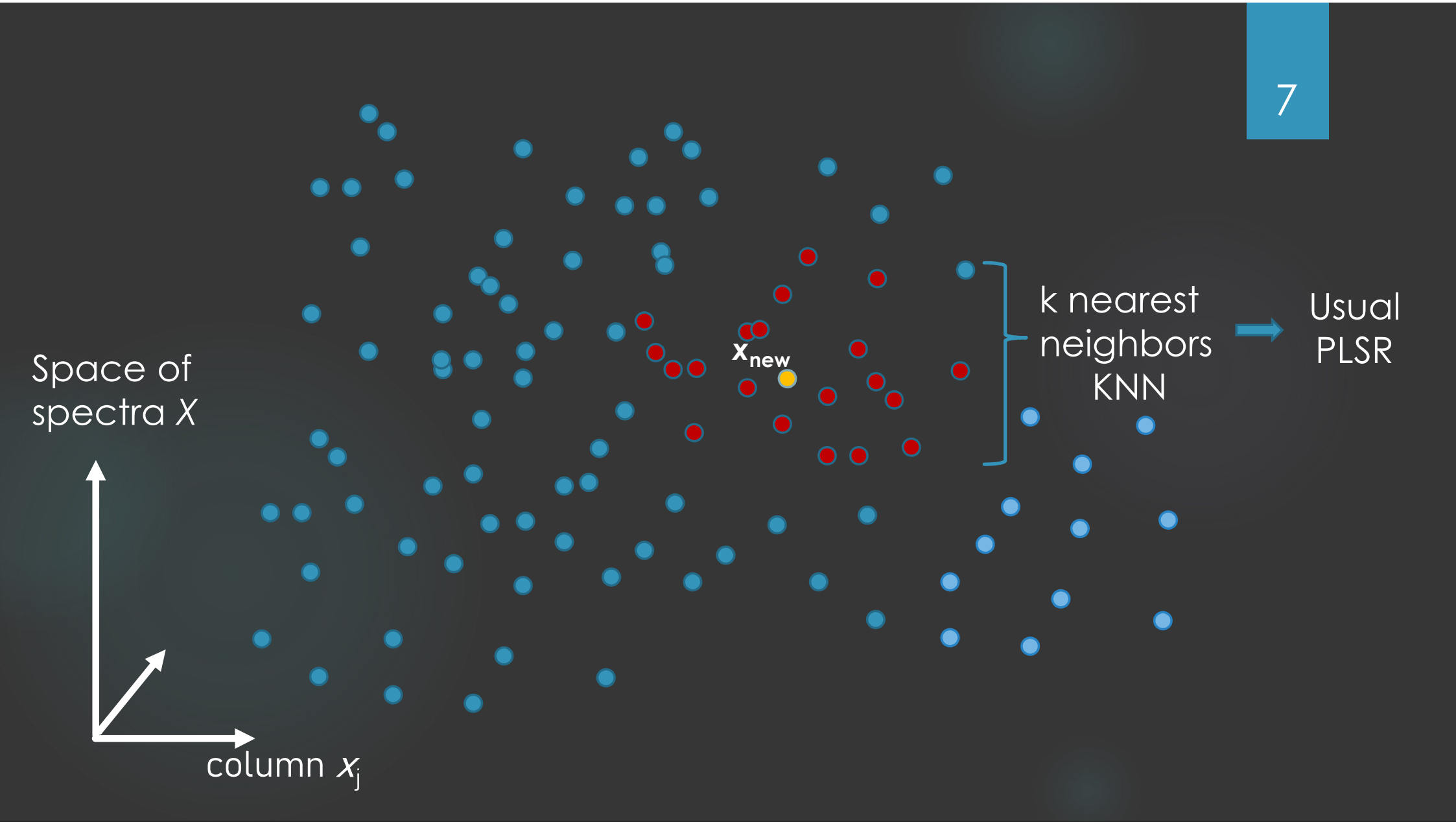
Volume 244, 15 January 2024, 105031

Averaging a local PLSR pipeline to predict chemical compositions and nutritive values of forages and feed from spectral near infrared data

Matthieu Lesnoff [a b c]

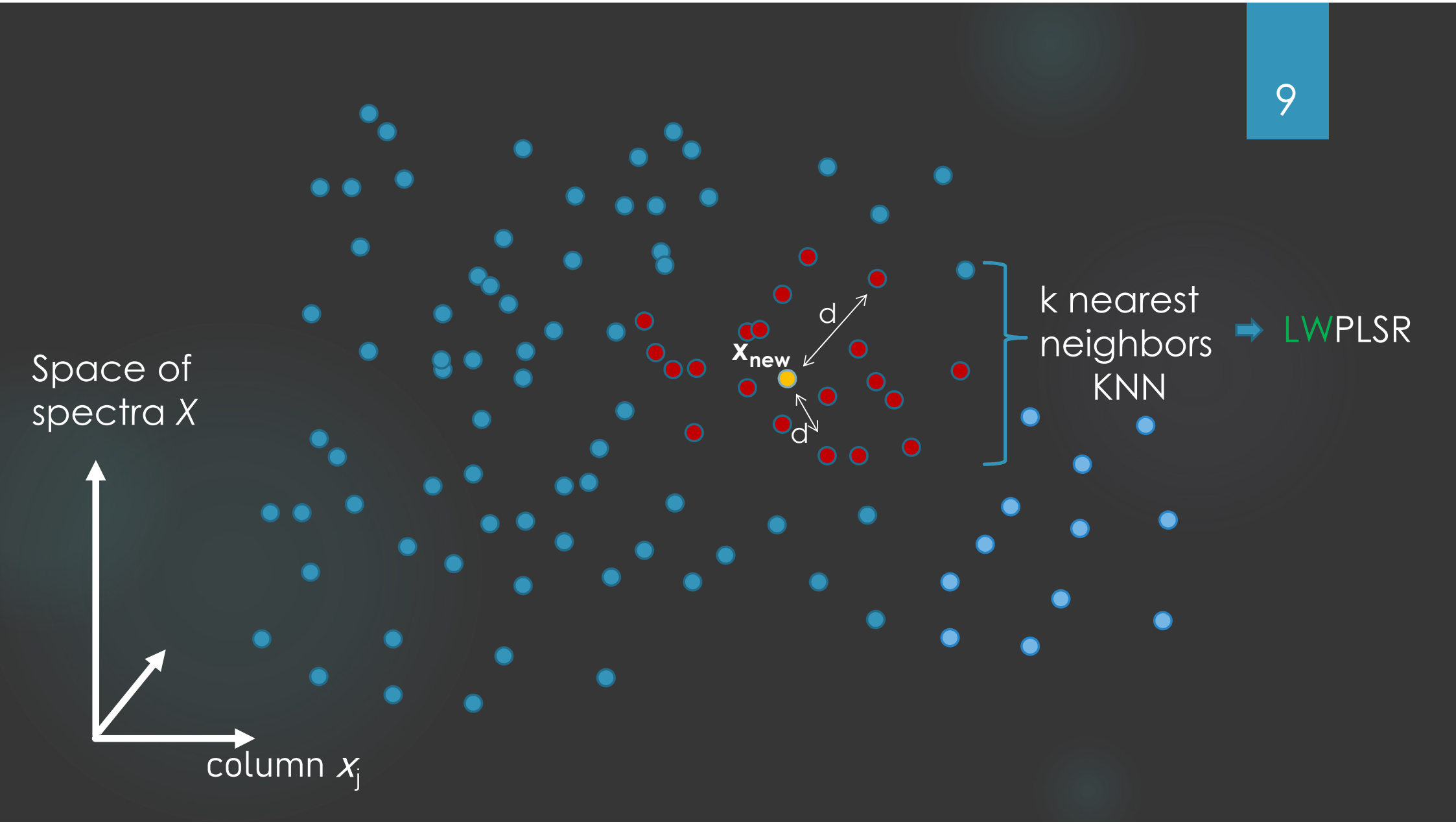

Spectra (Training)
Column-space **X**
column x_j


Space of
spectra X
column x_j
x_{new}
New spectrum
to predict


Space of spectra X
column x_j
x_{new}
k nearest neighbors KNN
Usual PLSR

- Usual PLSR $\max_{\boldsymbol{t}} \mathrm{Cov}(\boldsymbol{t}, \boldsymbol{y})^2 = \sum_{i=1}^{n} \left(\frac{1}{n} t_i y_i\right)^2$

- Weighted PLSR $\max_{\boldsymbol{t}} \mathrm{Cov}_W(\boldsymbol{t}, \boldsymbol{y})^2 = \sum_{i=1}^{n} (w_i t_i y_i)^2$


Space of spectra X
column x_j
x_{new}
d
d
k nearest neighbors KNN
LWPLSR

Weight function used by Jchemo.lwplsr

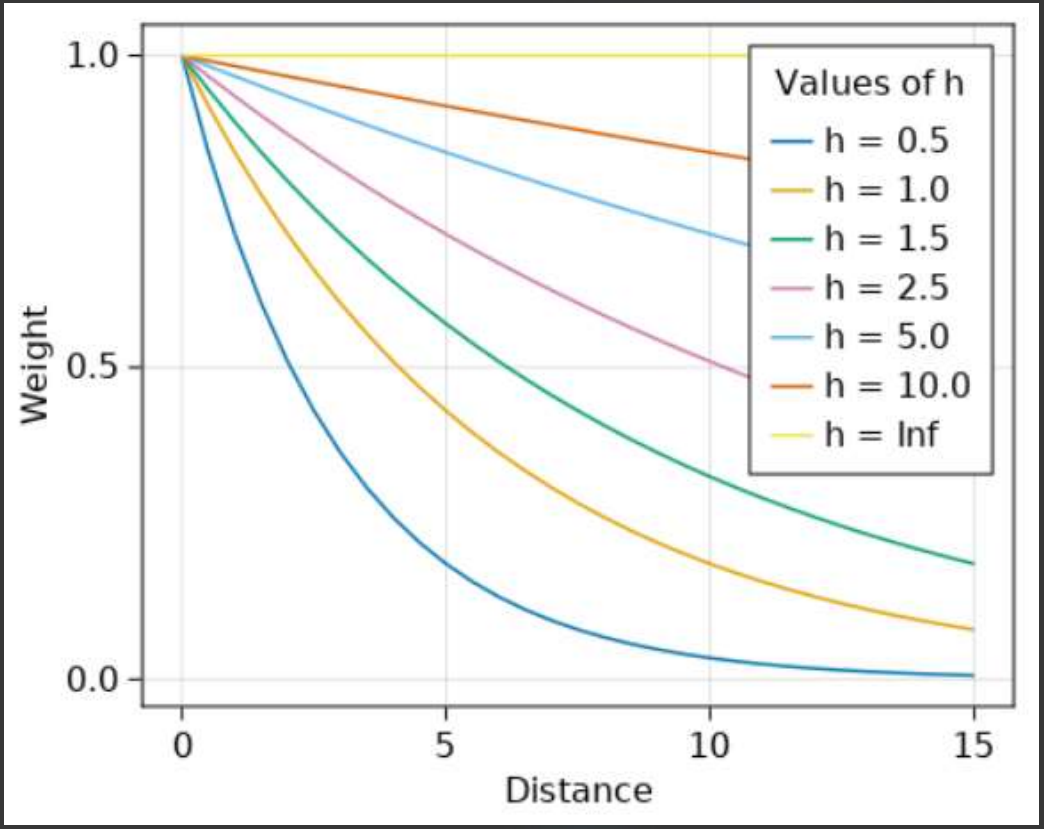
Adaptation from

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

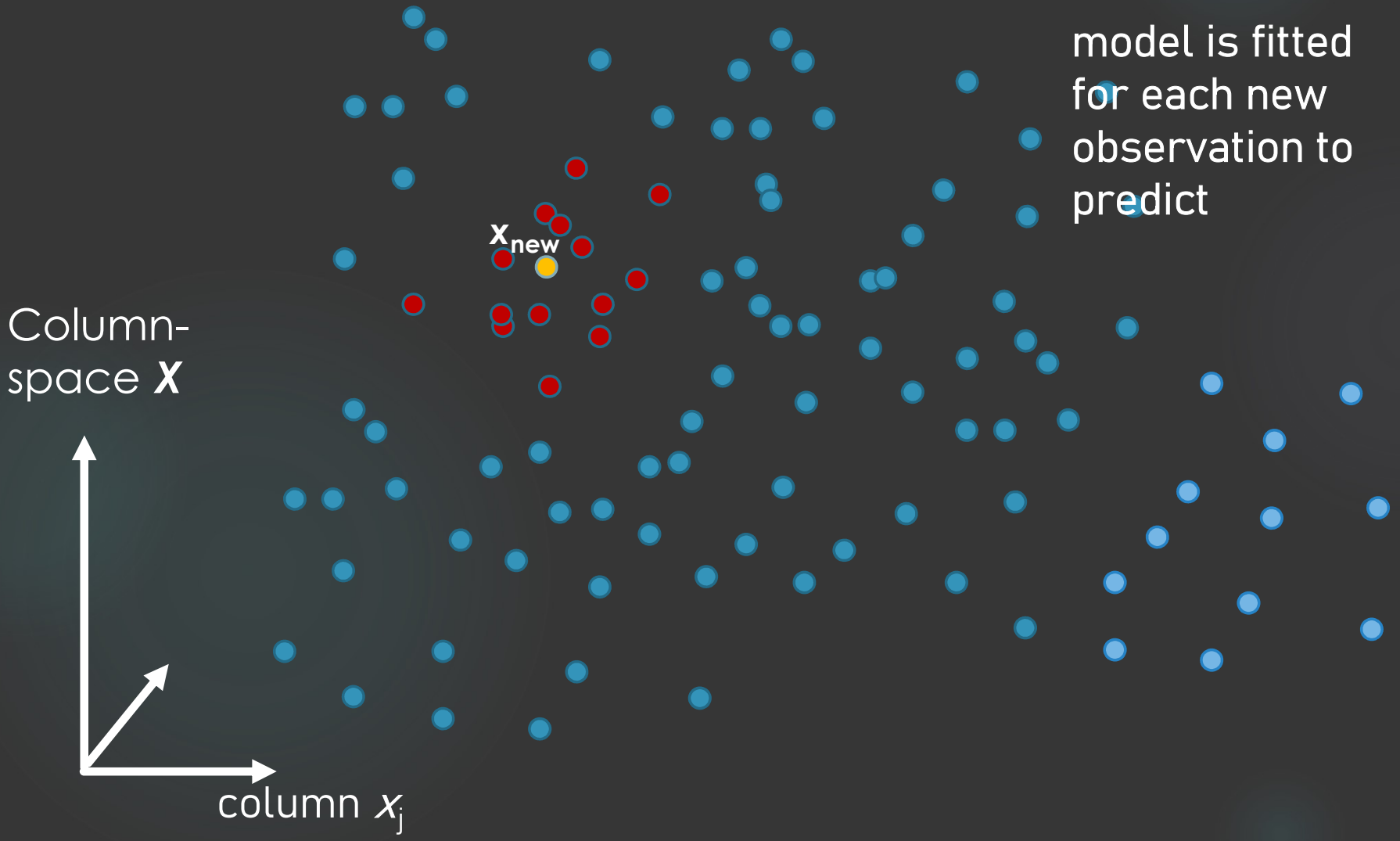
$j = 1, \ldots, k$ neighbors

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

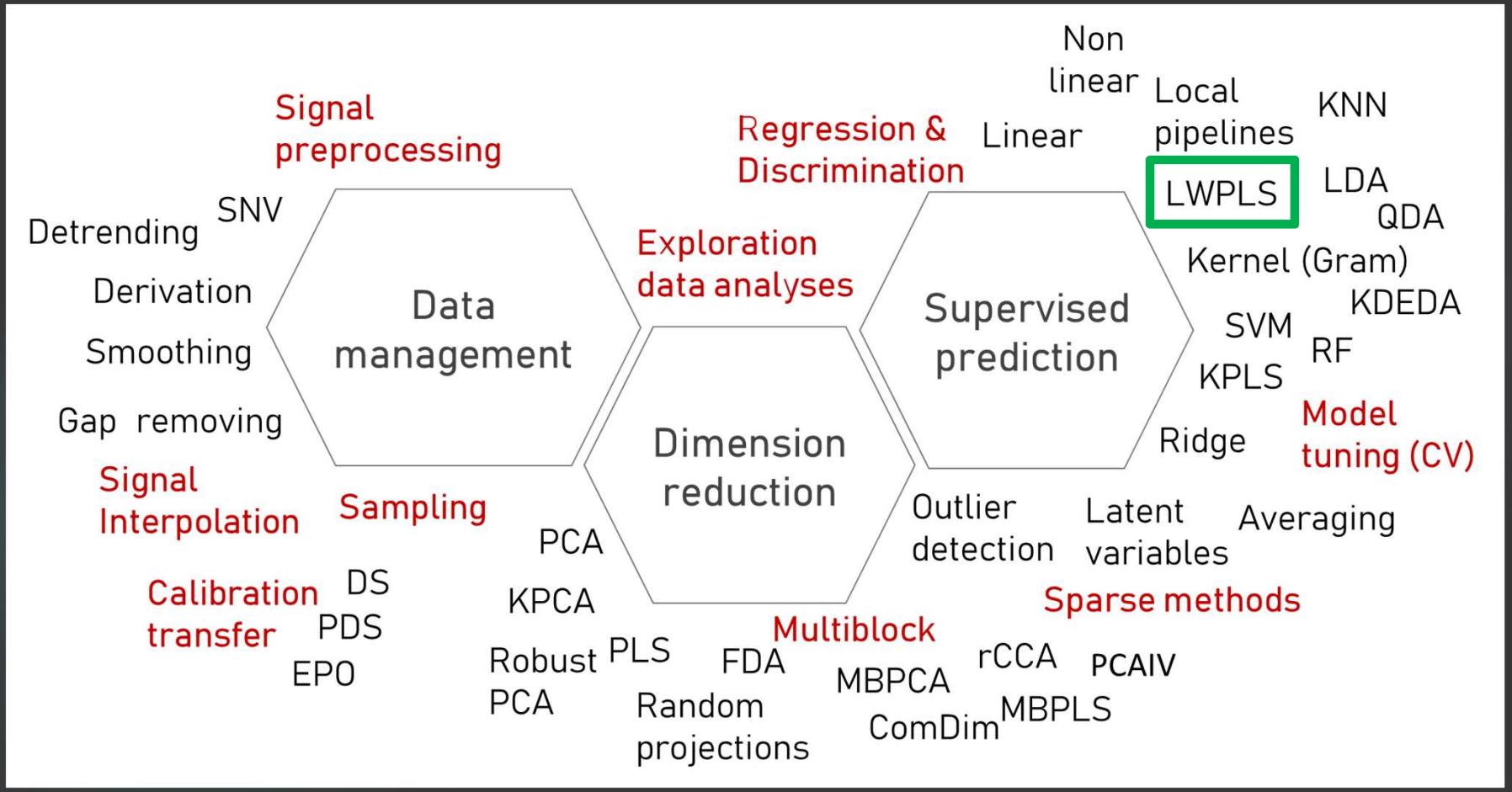
$w_j = w_j$ / maximum$\{w_1, \ldots, w_k\}$



!! A new model is fitted for each new observation to predict

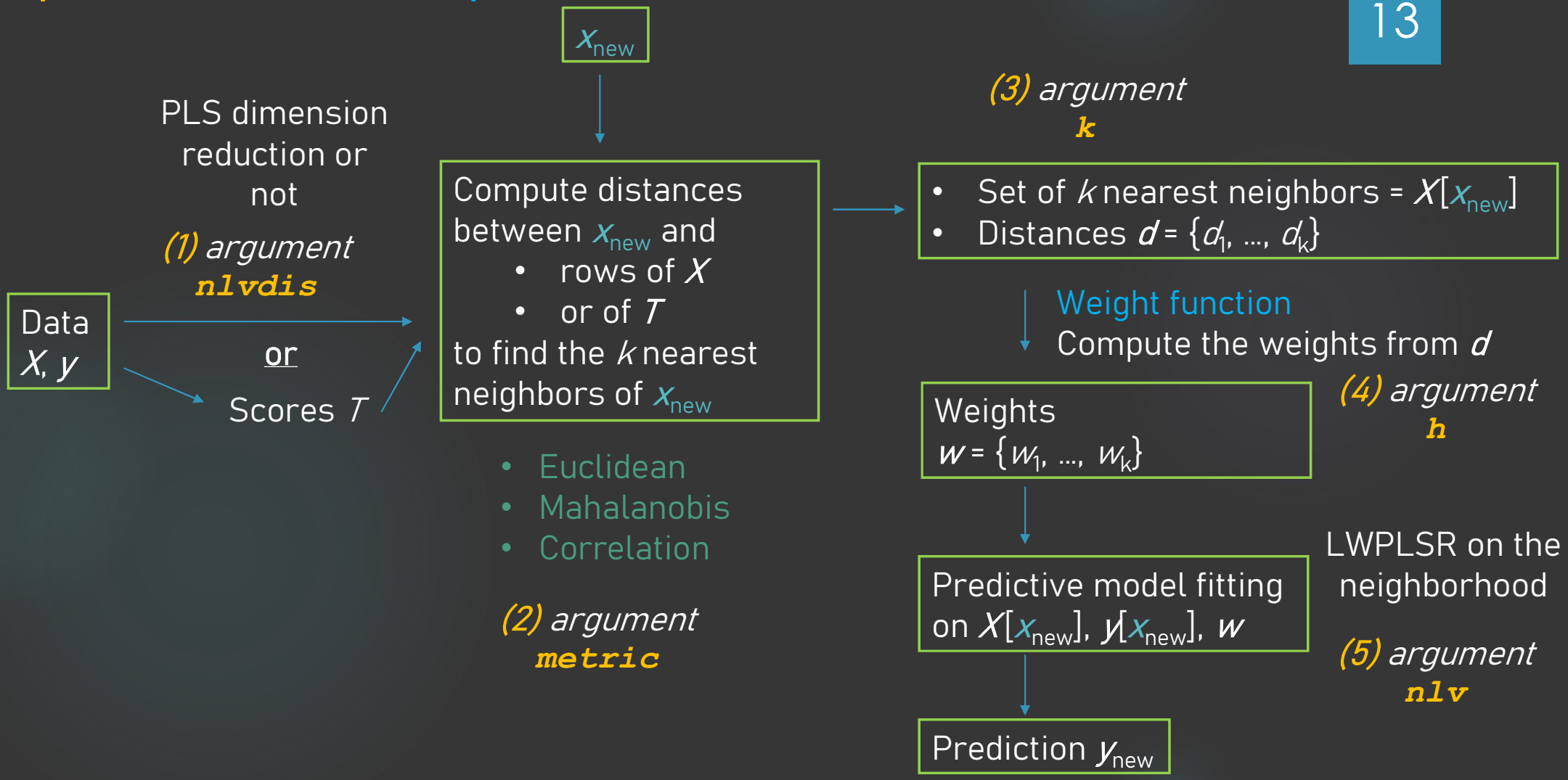


Package Jchemo.jl



https://mlesnoff.github.io/Jchemo.jl/dev/domains

Pipeline of function lwplsr



```
## function lwplsr

nlvdis = 20 ; metric = :mah
h = 1 ; k = 500 ; nlv = 15


mod = model(lwplsr;
                nlvdis, metric, h, k, nlv)
fit!(mod, X, y)


res = predict(mod, Xnew)
```

5 main arguments

Keyword arguments:

- nlvdis : Number of latent variables (LVs) to consider in the global PLS used for the dimension reduction before computing the dissimilarities. If nlvdis = 0, there is no dimension reduction.
- metric : Type of dissimilarity used to select the neighbors and to compute the weights. Possible values are: :eucl (Euclidean distance), :mah (Mahalanobis distance).
- h : A scalar defining the shape of the weight function computed by function wdist. Lower is h, sharper is the function. See function wdist for details (keyword arguments criw and squared of wdist can also be specified here).
- k : The number of nearest neighbors to select for each observation to predict.
- tolw : For stabilization when very close neighbors.
- nlv : Nb. latent variables (LVs) for the local (i.e. inside each neighborhood) models.
- scal : Boolean. If true, each column of X and Y is scaled by its uncorrected standard deviation for the global dimension reduction and the local models.