Spec2Vec

Patrick Michl <patrick.michl@gmail.com>

Energy based non-linear Calibration in NIRS

2022-06-21

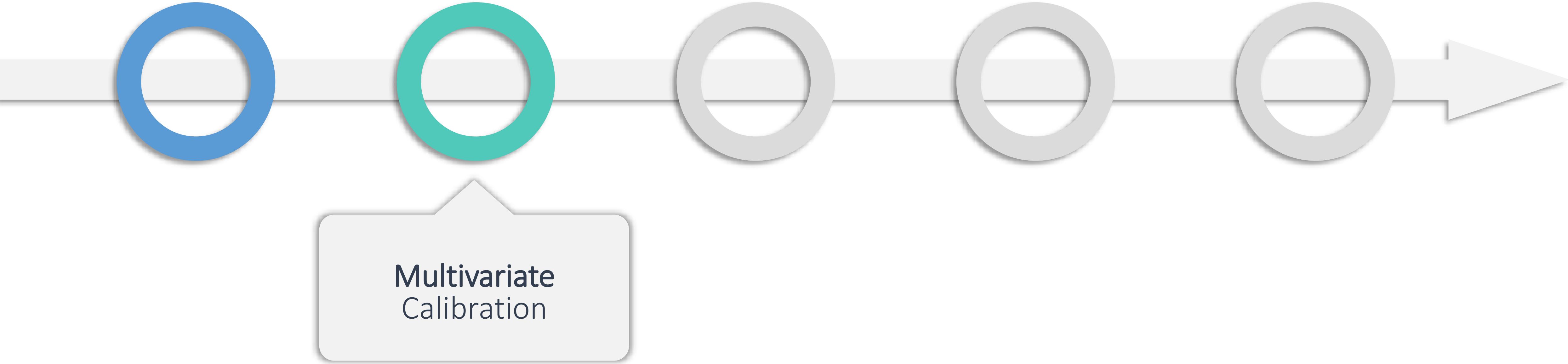
Abstract

The success of multivariate calibration depends on the generalizability of the underlying regression models, for which the spectral features are usually decorrelated in a preliminary step. In order to also allow the application of deep neural networks like transformers, this presentation introduces a powerful approach to highly non-linear decorrelation of spectral features using deep energy based modeling.

Outline

NIR
Spectroscopy

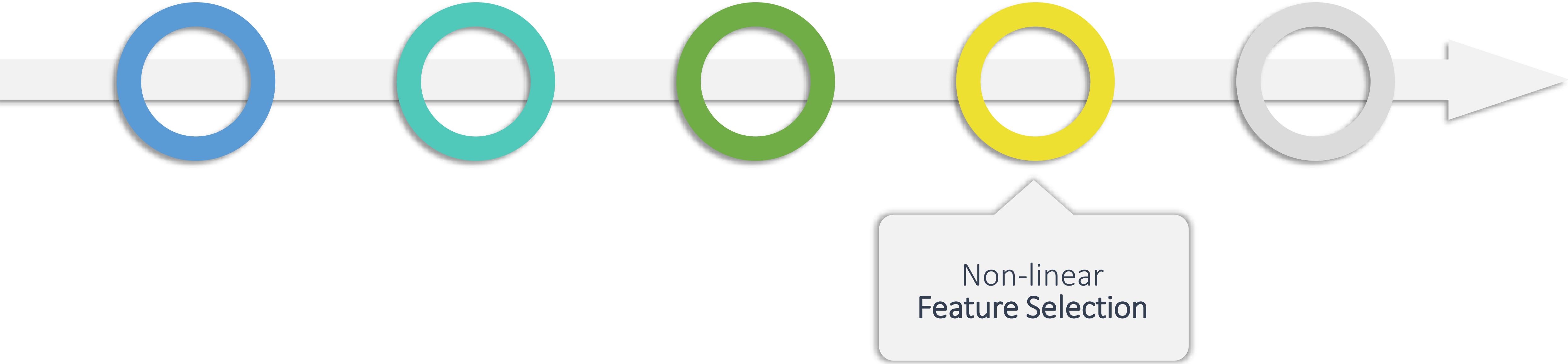
Outline



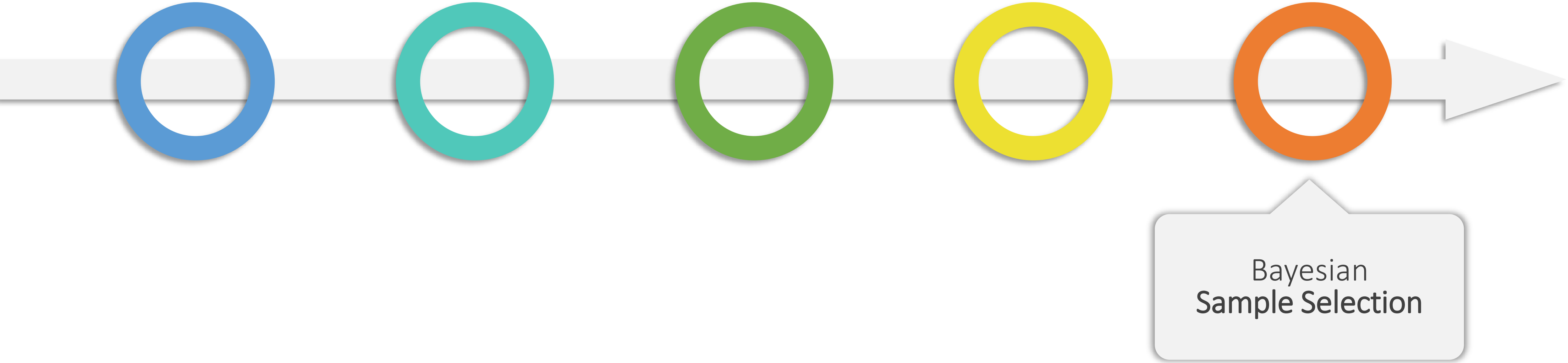
Outline

Energy based
Calibration

Outline



Outline



Part 1

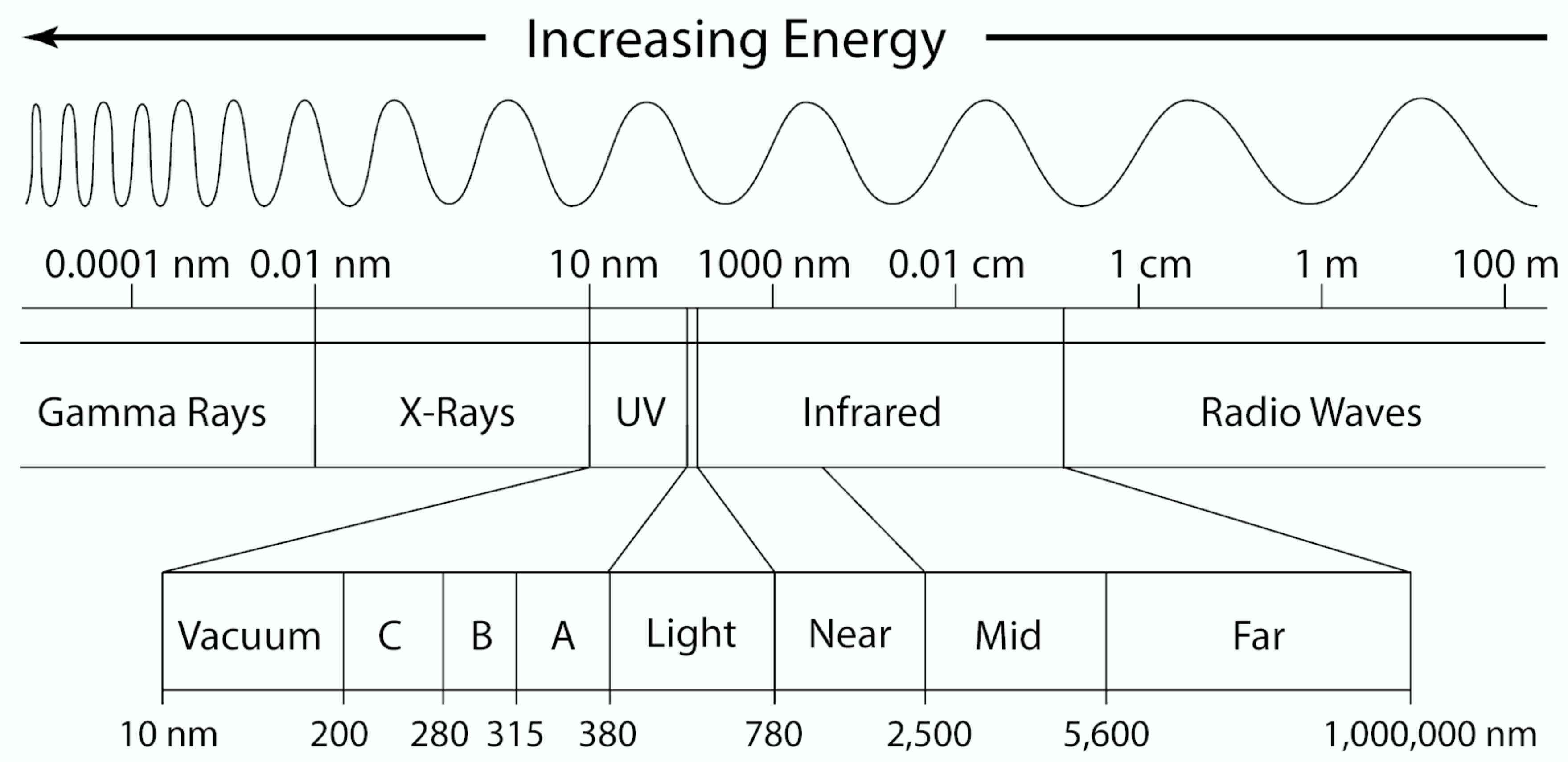
NIR Spectroscopy

What is **NIR**?

NIR Spectroscopy

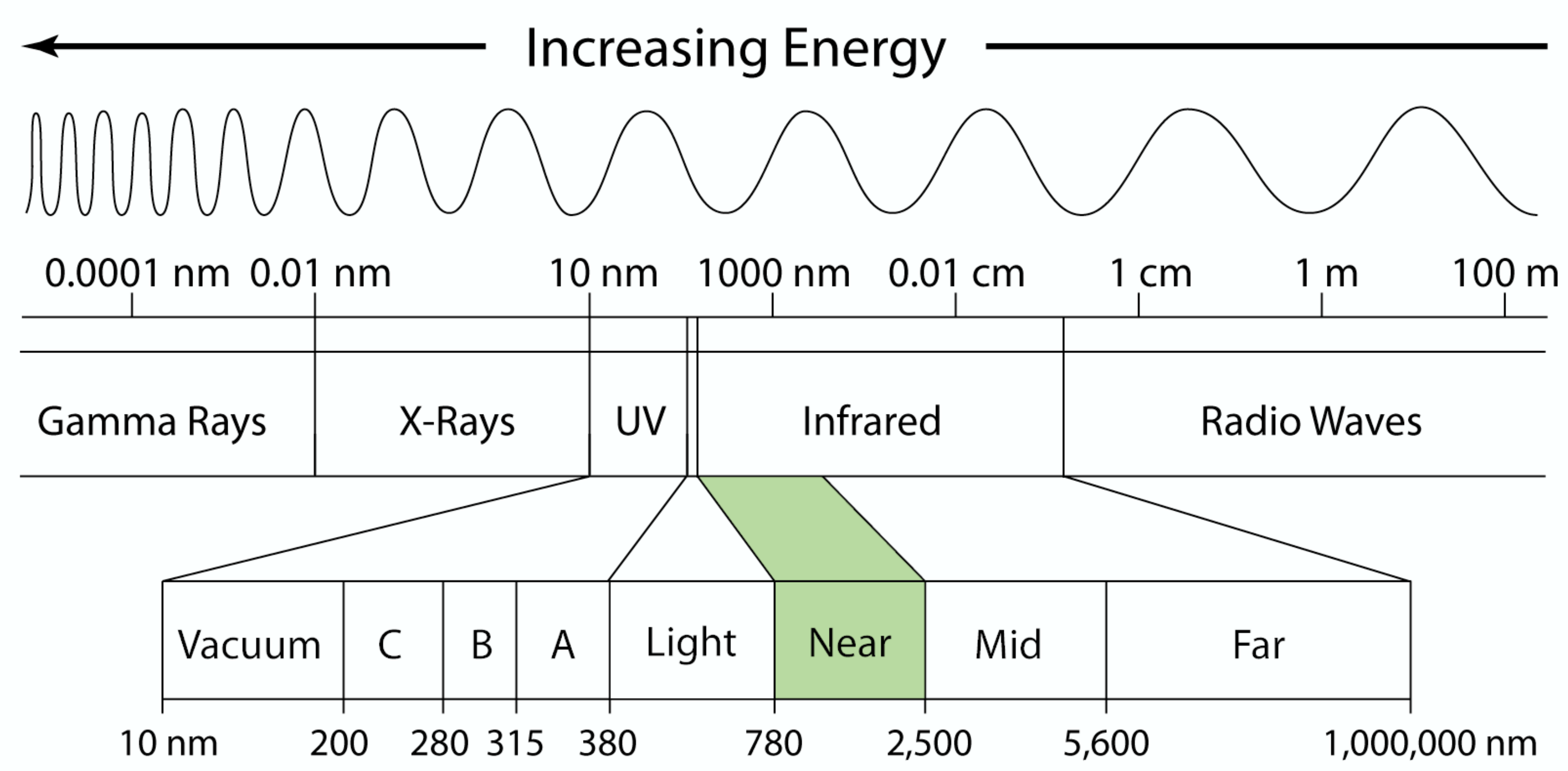
NIR = Near Infrared

Region in the
EM spectrum



Ranging

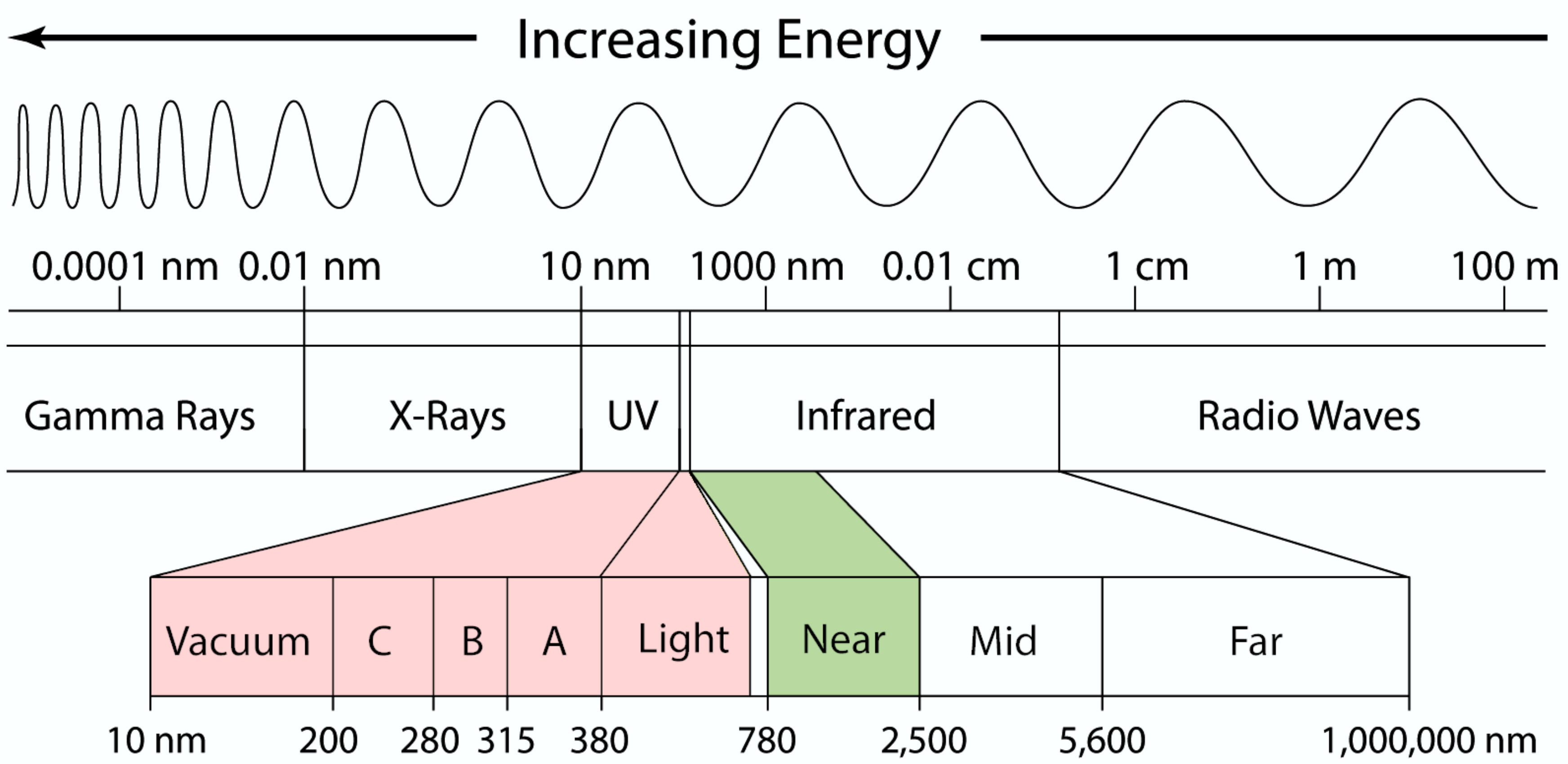
from ~780nm
to ~2500nm



Ranging

No Atomic Spectral Lines

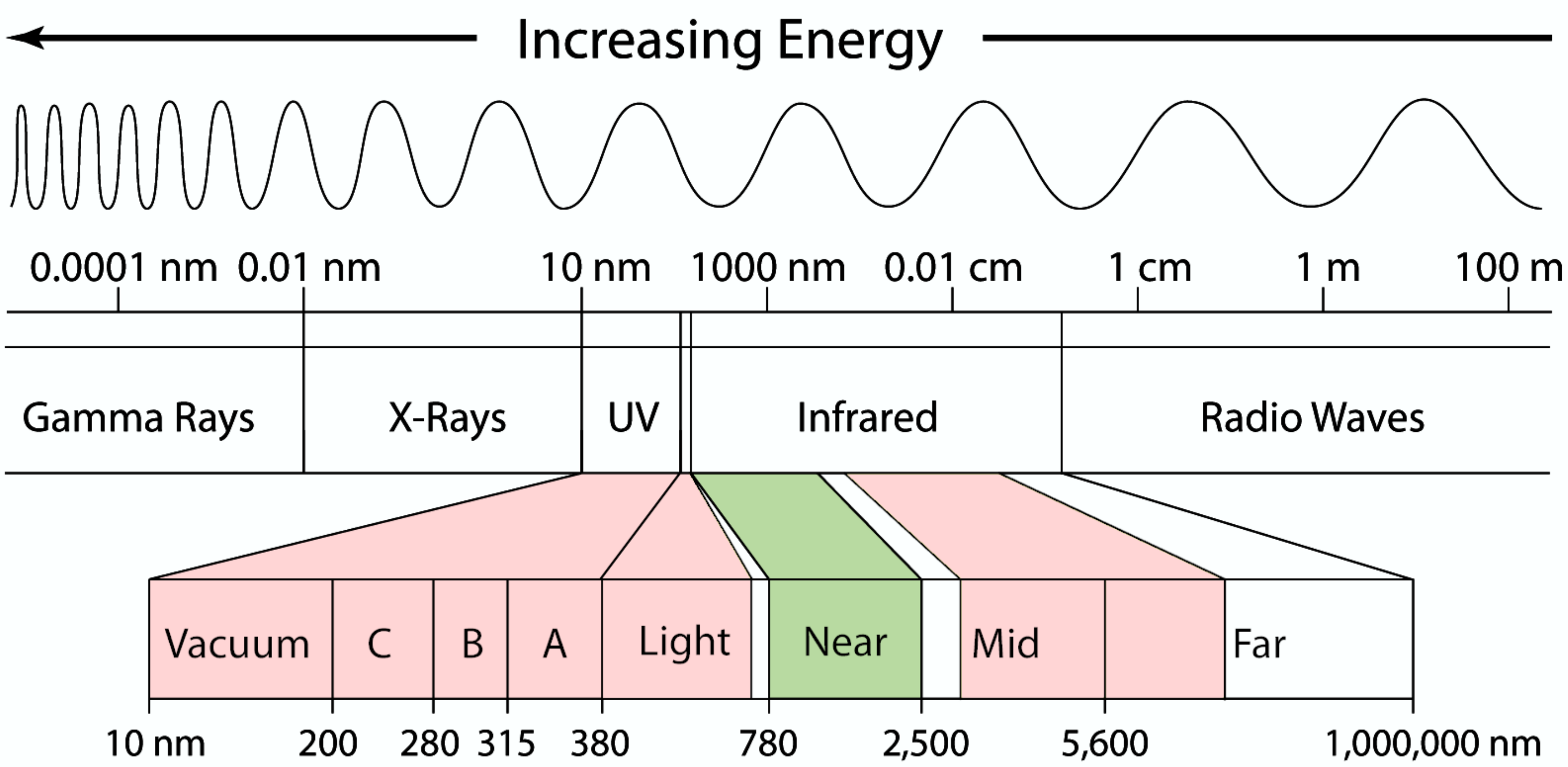
Eigenstates have transitions at typical wavelengths: ~10nm to ~750nm



Ranging

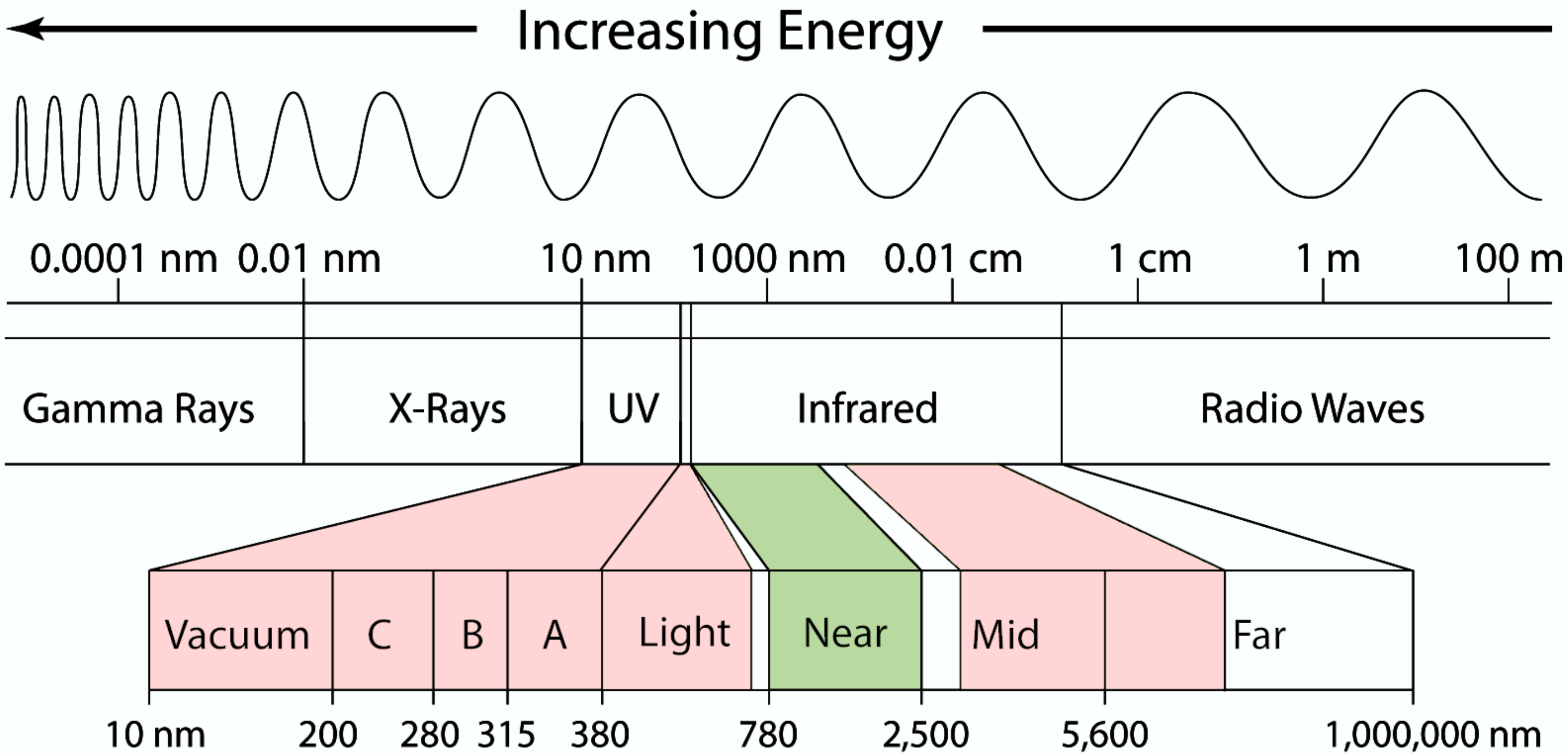
No
Molecule Vibrations

Eigenfrequencies
of bonds are typically at wavelengths:
~3000nm to ~30000nm



Ranging

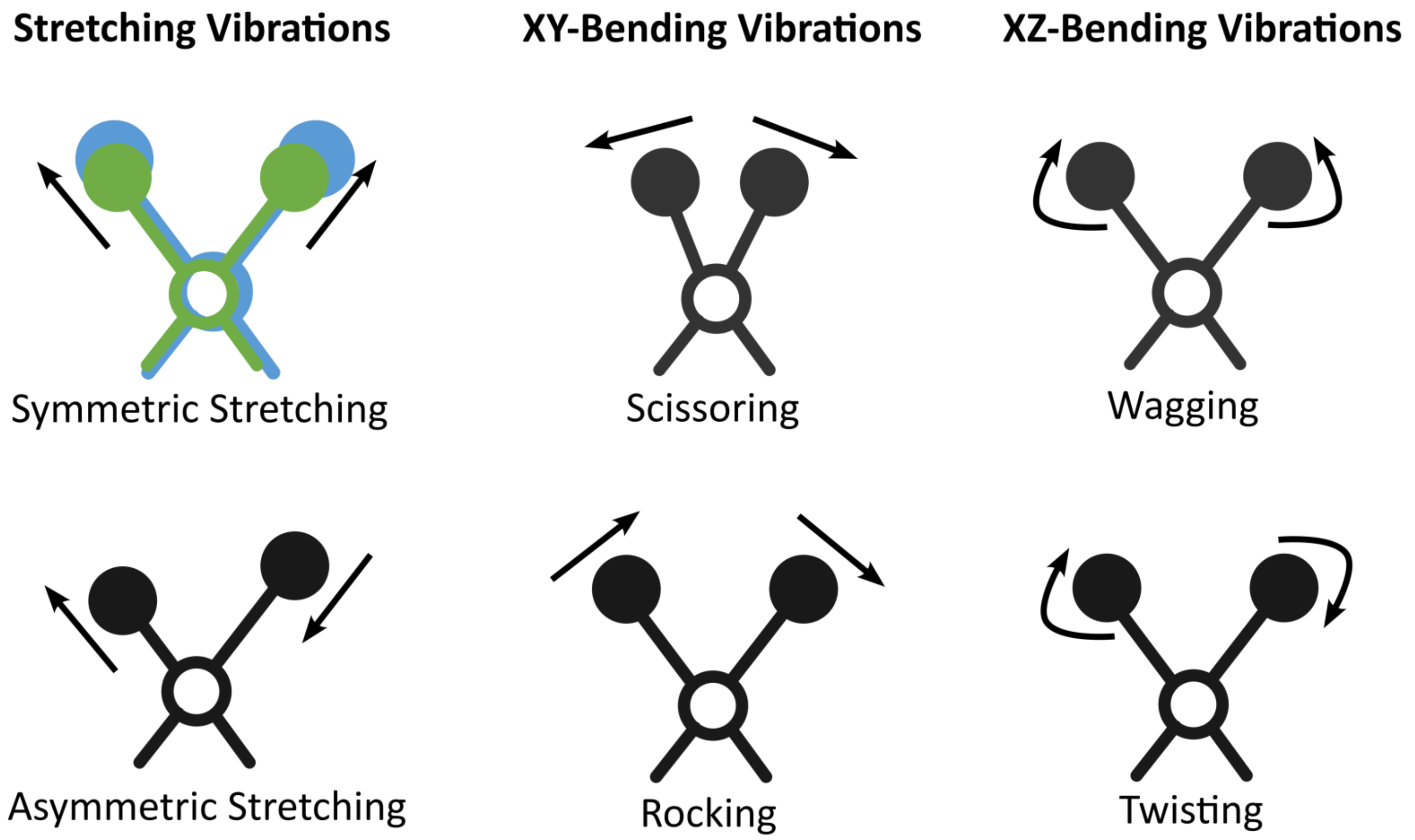
No Fundamental Spectral Lines!



Ranging

Vibrational Overtones

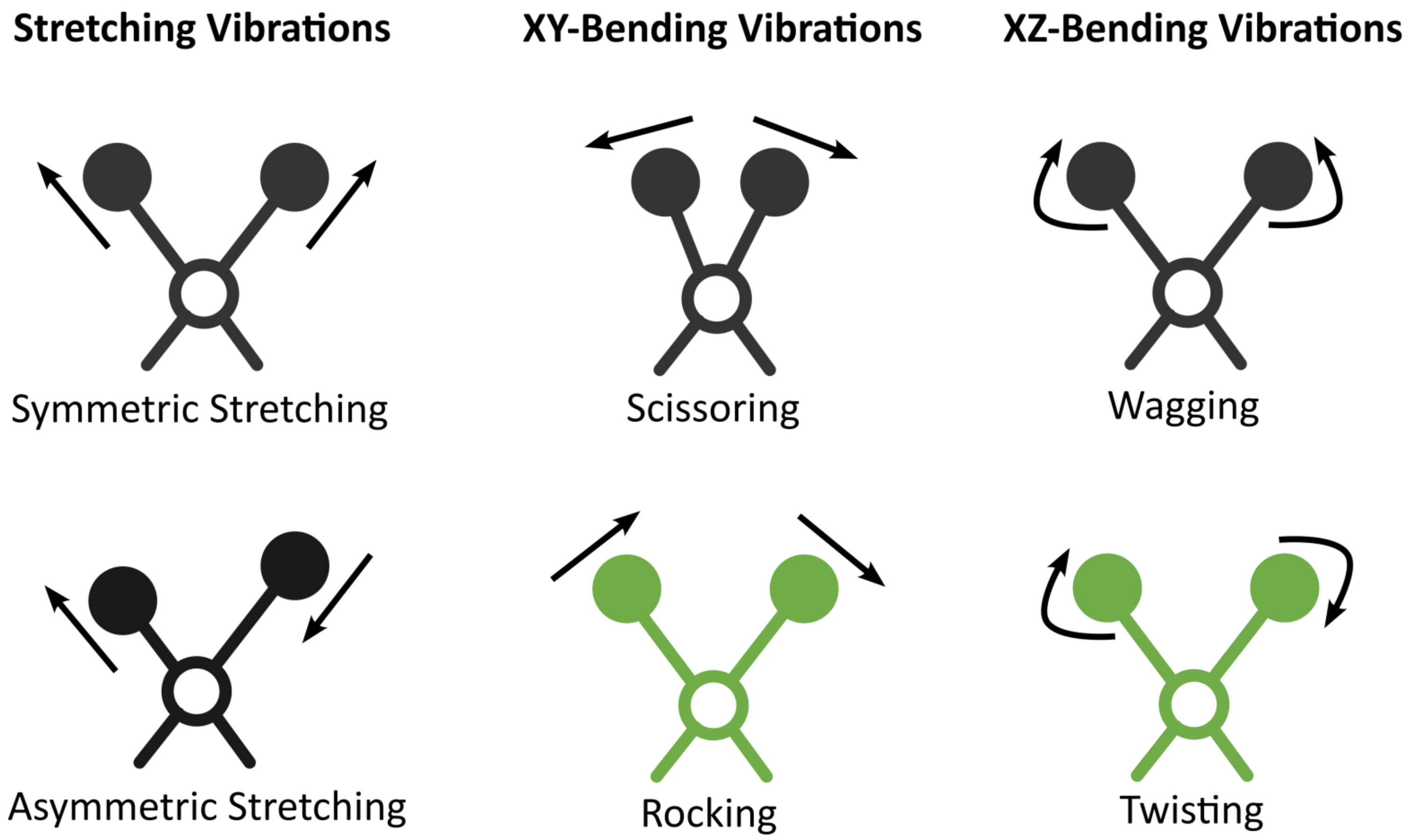
Vibration modes provide higher modes. 2*nd* modes are about 10-100x weaker than fundamental modes



Ranging

Combination Modes

12 possible pairwise combinations of fundamental modes with highly unspecific (broad) bands



What can be concluded about **NIR**?

Deeper Penetration of samples

Weaker absorption allows deeper penetration of samples

$\mu m \rightarrow mm$

Deeper Penetration of samples

Weaker absorption allows deeper penetration of samples

$\mu m \rightarrow mm$

Simpler Preparation

Due to deeper penetration the sample surface is less cruzial

Deeper Penetration of samples

Weaker absorption allows deeper penetration of samples

$\mu m \rightarrow mm$

Simpler Preparation

Due to deeper penetration the sample surface is less cruzial

Condensed Information

Overlapping bands of deeper penetrations provide comprehensive information within the NIR region

Deeper Penetration of samples

Weaker absorption allows deeper penetration of samples

$\mu m \rightarrow mm$

Simpler Preparation

Due to deeper penetration the sample surface is less cruzial

Condensed Information

Overlapping bands of deeper penetrations provide comprehensive information within the NIR region

Unspecific Features

Due to overlapping, individual spectral features are insufficient for the prediction of chemical features

Deeper Penetration of samples

Weaker absorption allows deeper penetration of samples

μm → mm

Simpler Preparation

Due to deeper penetration the sample surface is less cruzial

Condensed Information

Overlapping bands of deeper penetrations provide comprehensive information within the NIR region

Unspecific Features

Due to overlapping, individual spectral features are insufficient for the prediction of chemical features

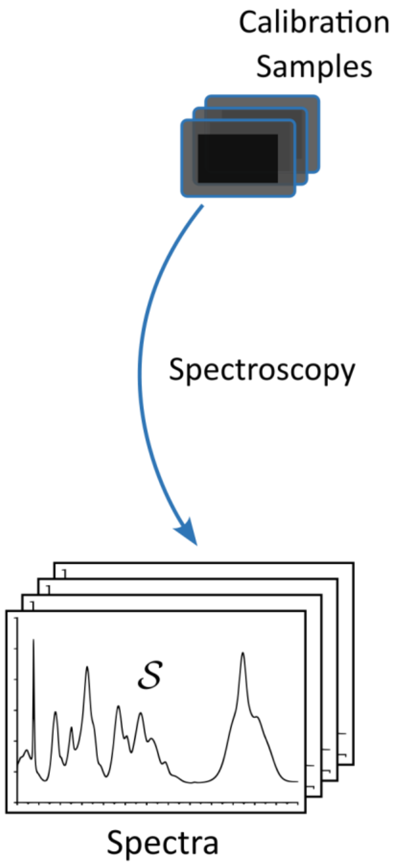
Interpretability of spectra?

Part 2

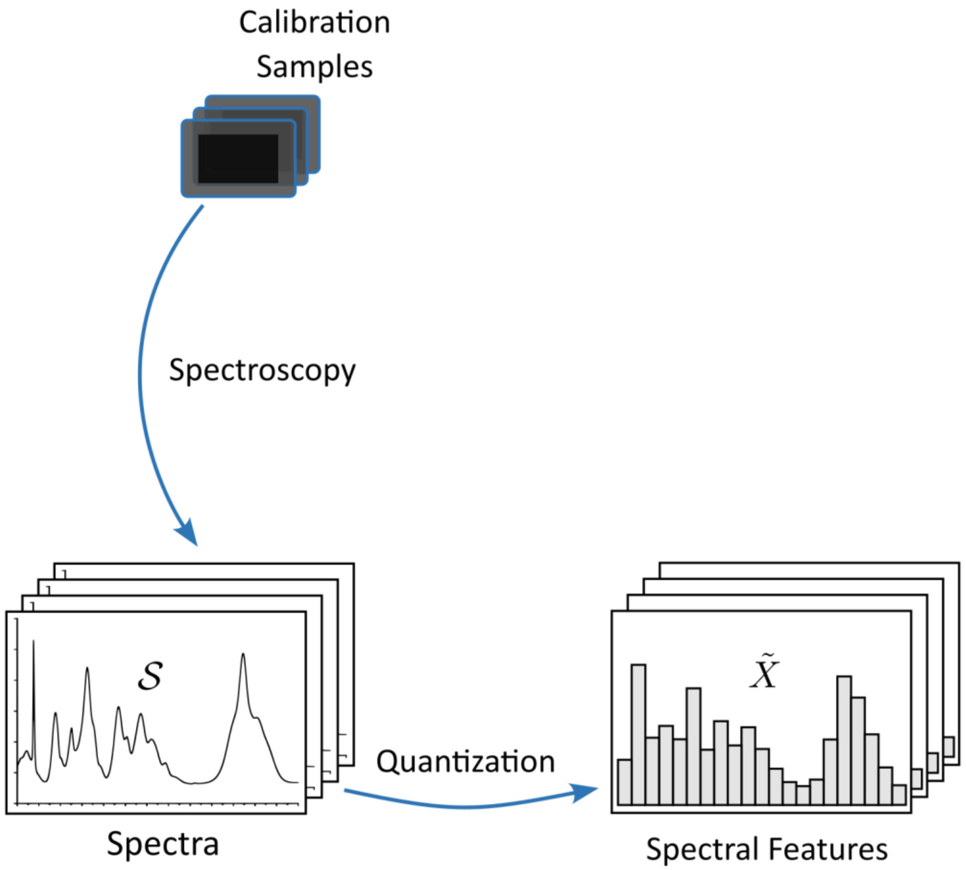
Multivariate Calibration

What is the **General Setup**?

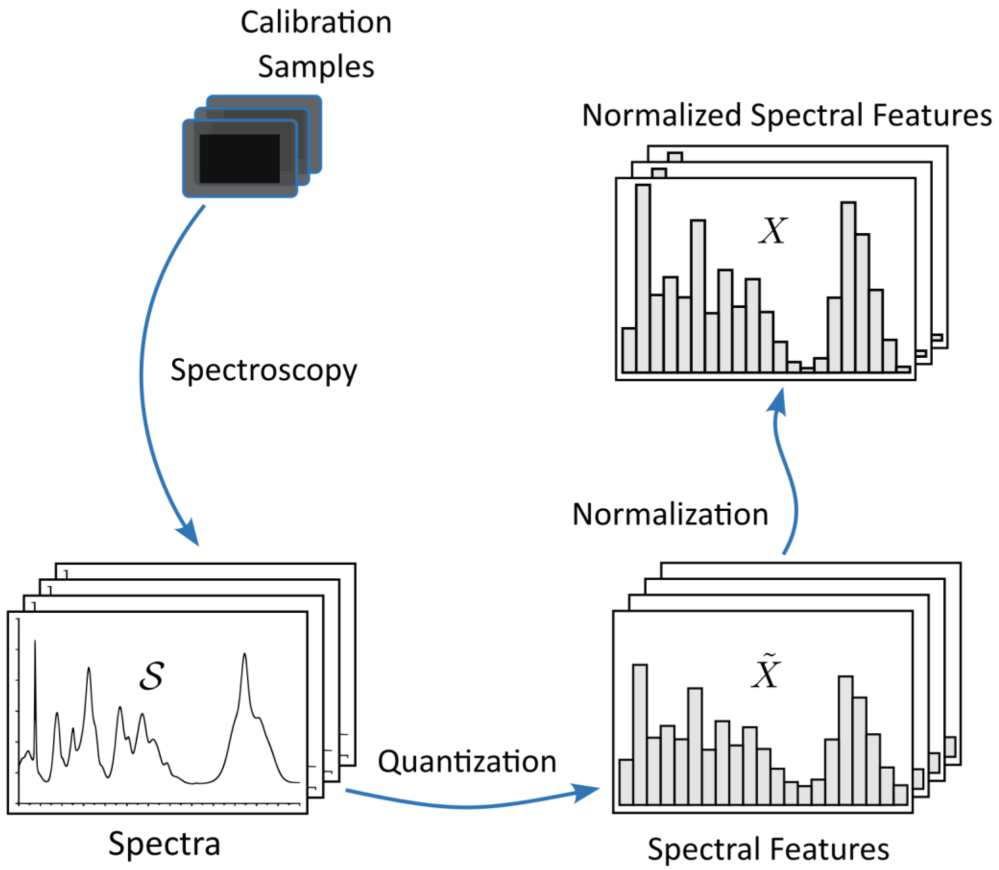
1. **Spectroscopy**



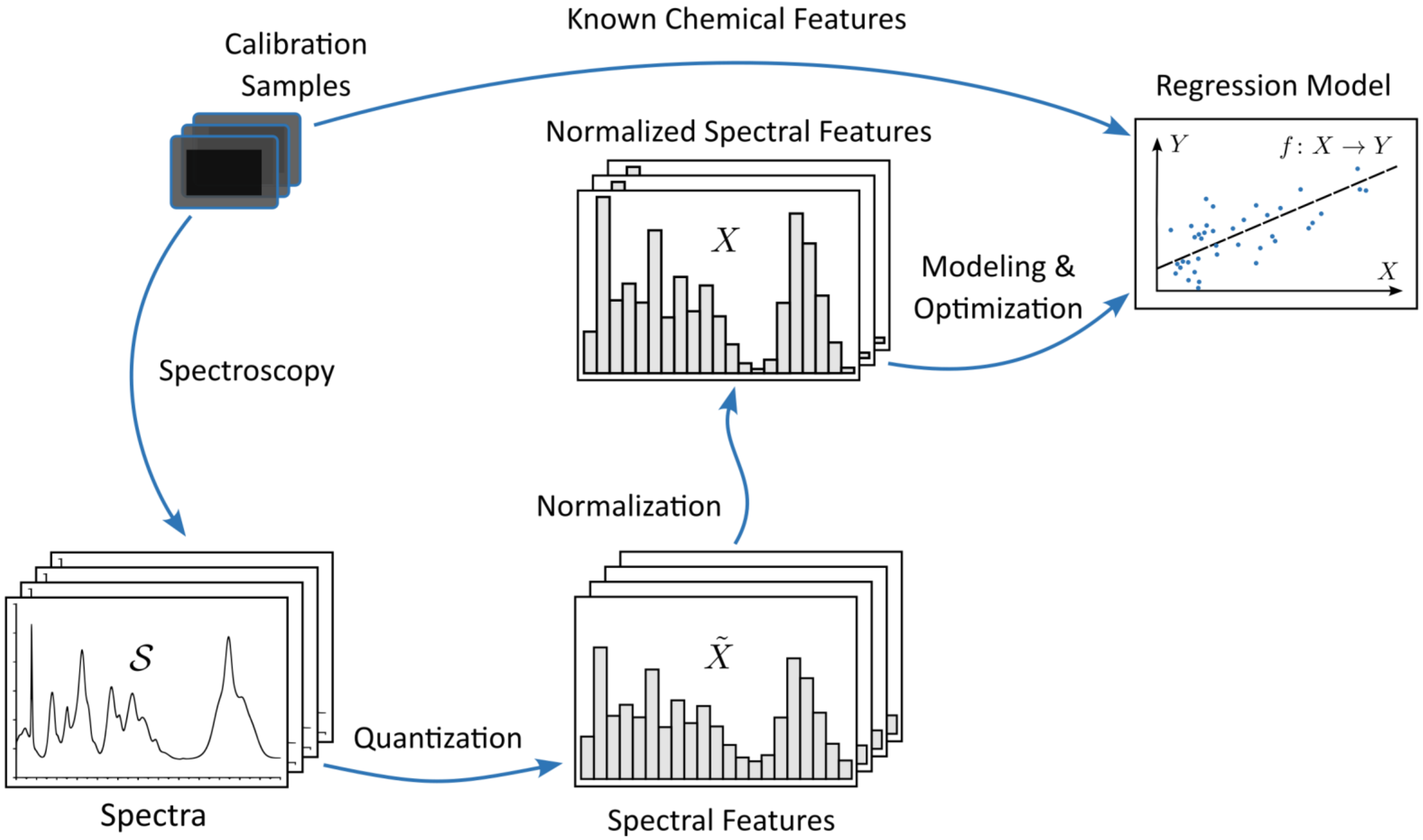
1. **Spectroscopy**
2. **Quantization**



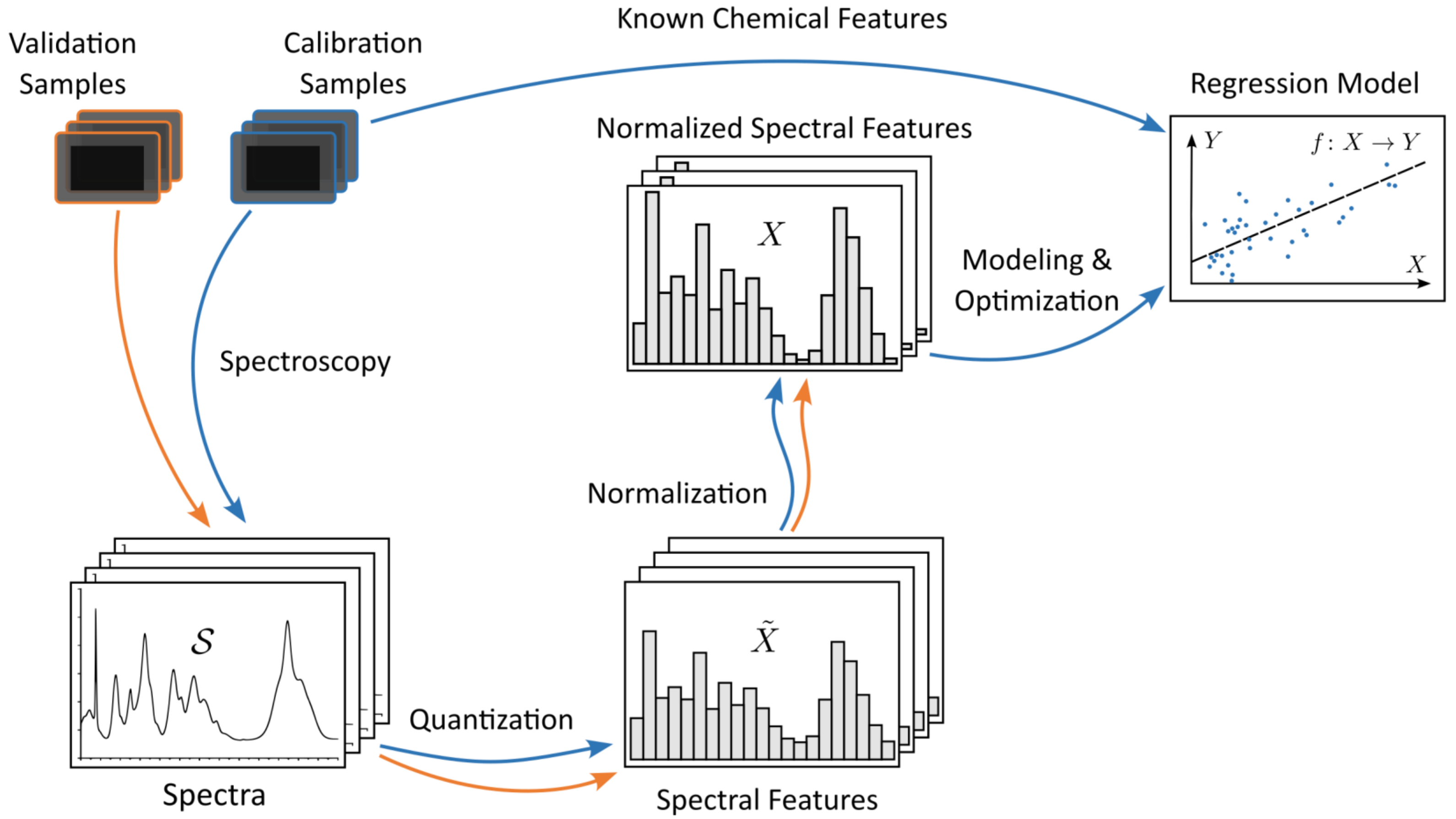
1. **Spectroscopy**
2. **Quantization**
3. **Normalization**



1. **Spectroscopy**
2. **Quantization**
3. **Normalization**
4. **Modeling & Optimization**

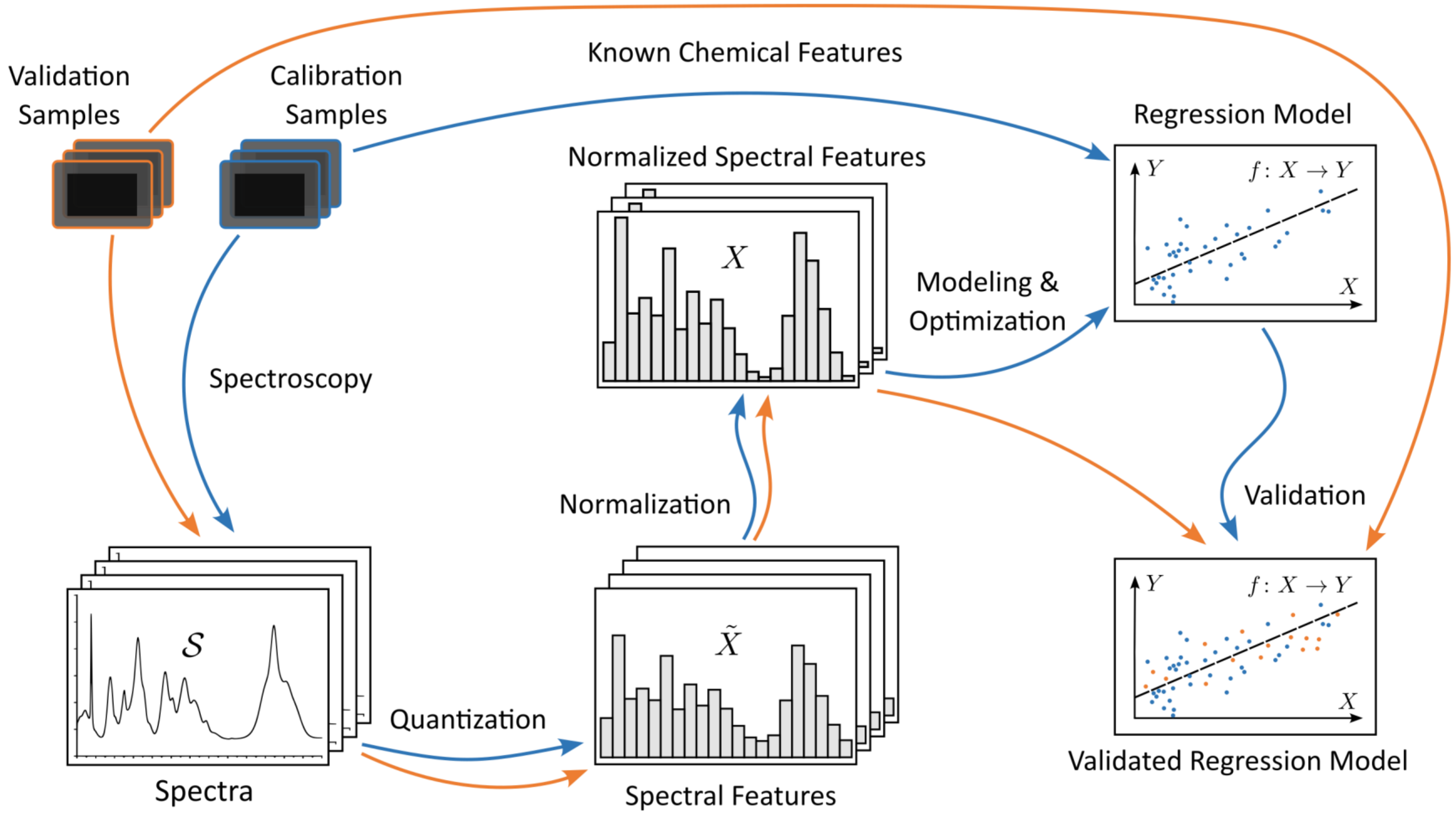


1. **Spectroscopy**
2. **Quantization**
3. **Normalization**
4. **Modeling & Optimization**
5. **Validation**



Multivariate Calibration

1. **Spectroscopy**
2. **Quantization**
3. **Normalization**
4. **Modeling & Optimization**
5. **Validation**



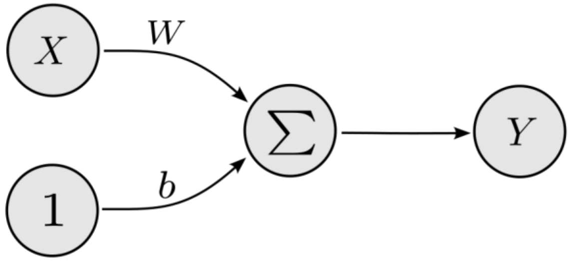
What are the tools in
Statistical Modeling?

Approach 1: *Multiple Linear Regression*

Linear Superposition!

Model chemical features by superpositions of spectral features

Approach 1: *Multiple Linear Regression*



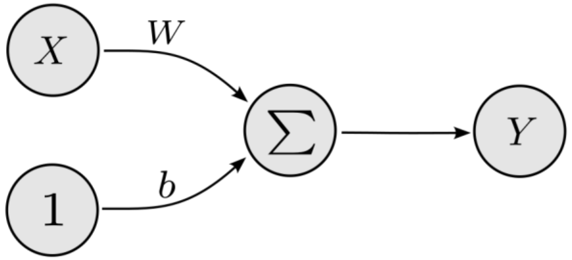
Linear Superposition!

Model chemical features by superpositions of spectral features

Linear Regression

$\boldsymbol{W}$ Weights
$\boldsymbol{b}$ Bias

Approach 1: *Multiple Linear Regression*



Linear Superposition!

Model chemical features by superpositions of spectral features

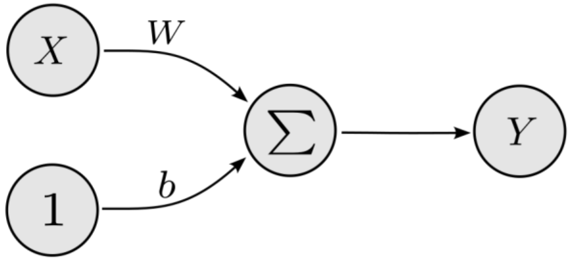
Linear Regression

$\boldsymbol{W}$ Weights
$\boldsymbol{b}$ Bias

Not Generalizable!

Due to multicollinearity

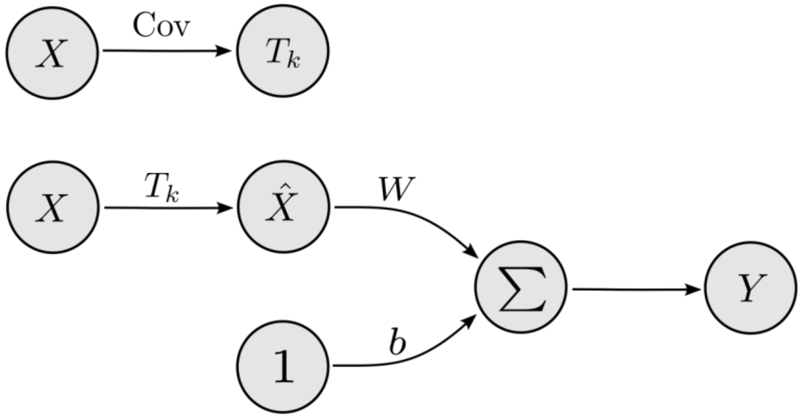
Approach 1: *Multiple Linear Regression*



Linear Decorrelation!

Decorrelate spectral features by orthogonal components

Approach 2: *Principal Component (PC) Regression*



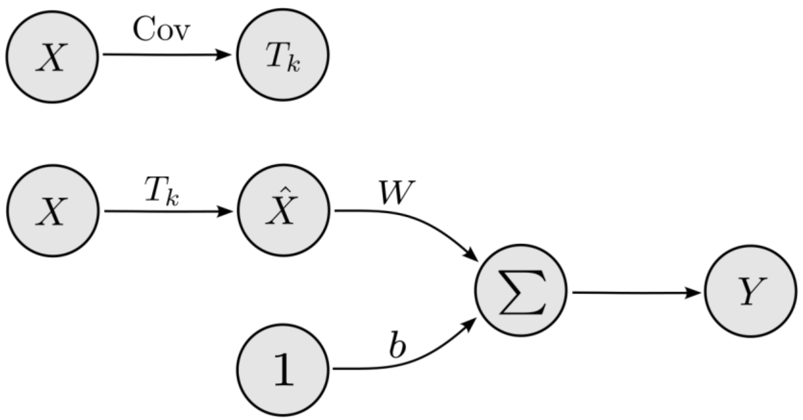
Linear Decorrelation!

Decorrelate spectral features by orthogonal components

PC Regression

$\boldsymbol{W}$ Weights
$\boldsymbol{b}$ Bias
$\boldsymbol{T_k}$ Projection of X to the first k Principal Components of X

Approach 2: *Principal Component (PC) Regression*



Linear Decorrelation!

Decorrelate spectral features by orthogonal components

PC Regression

$\boldsymbol{W}$ Weights
$\boldsymbol{b}$ Bias
$\boldsymbol{T_k}$ Projection of X to the first k Principal Components of X

Not Efficient!

Only accounts for covariances of spectral features

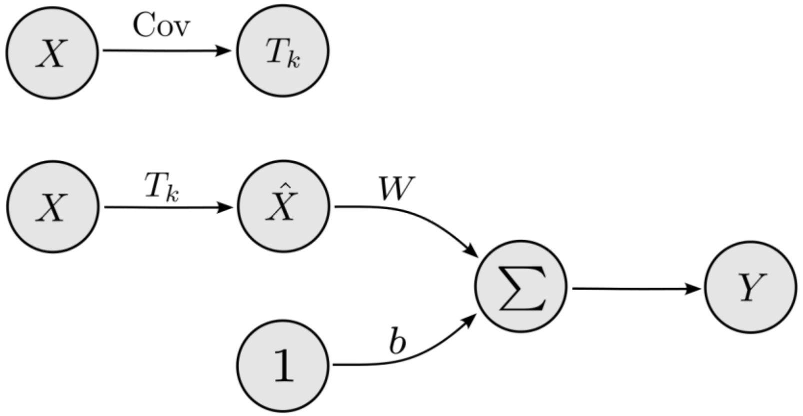
patrick.michl@gmail.com

Multivariate Calibration

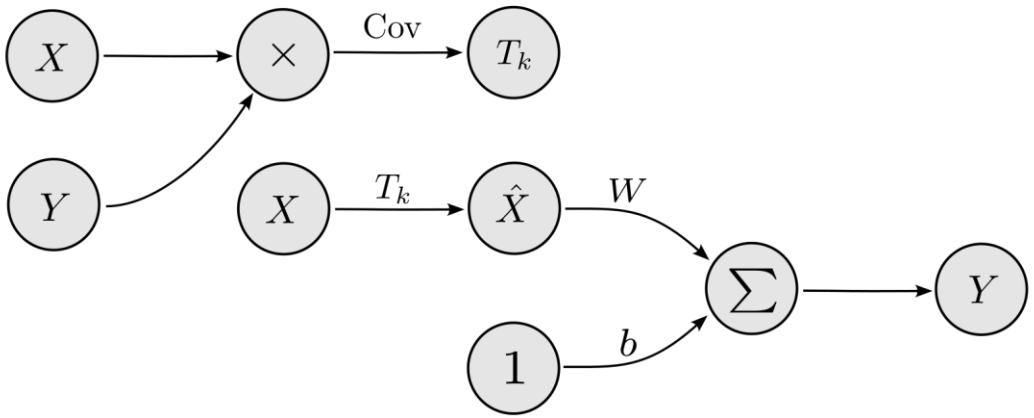
Approach 2: *Principal Component (PC) Regression*

Incorporate Chemical Features!

Decorrelate the combined vector of spectral features and chemical features



Approach 3: *Partial Least Squares (PLS) Regression*



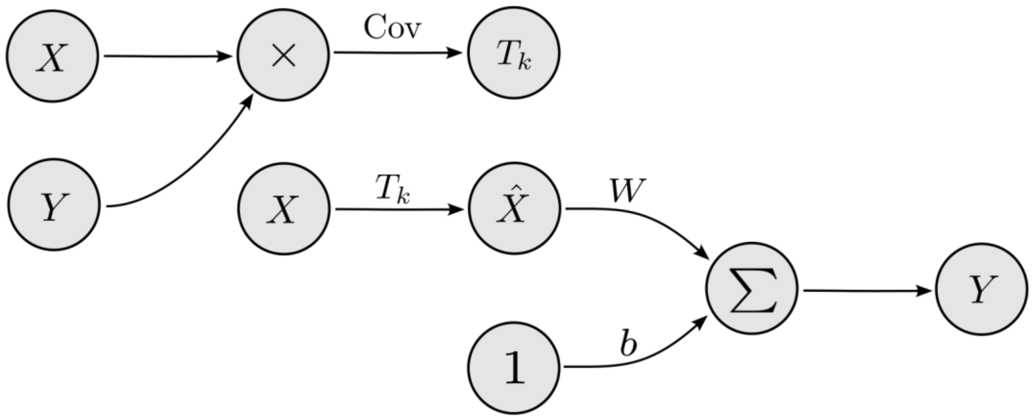
Incorporate Chemical Features!

Decorrelate the combined vector of spectral features and chemical features

PLS Regression

$\boldsymbol{W}$ Weights
$\boldsymbol{b}$ Bias
$\boldsymbol{T_k}$ Projection of X to the first k Principal Components of $X \times Y$

Approach 3: *Partial Least Squares (PLS) Regression*



Incorporate Chemical Features!

Decorrelate the combined vector of spectral features and chemical features

PLS Regression

$\boldsymbol{W}$ Weights
$\boldsymbol{b}$ Bias
$\boldsymbol{T_k}$ Projection of X to the first k Principal Components of $X \times Y$

Sufficient ?

What can be concluded
about **Multivariate Calibration**?

Samples are rare!

The requirement to extract chemical features by **wet-lab** experiments makes reference samples rare

Samples are rare!

The requirement to extract chemical features by **wet-lab experiments** makes reference samples rare

Direct Regression is not possible!

Overlapping and scattered bands create **multicollinearity** in spectral features

Samples are rare!

The requirement to extract chemical features by **wet-lab experiments** makes reference samples rare

Direct Regression is not possible!

Overlapping and scattered bands create **multicollinearity** in spectral features

Decorrelation in PLS allows efficient linear Regression

A preceeded **decorrelation step** in PLS allows allows linear regression

Non-Linear Spectral Dependencies in NIR

Combined vibrational modes in NIR create highly **non-linear conditional dependencies** between spectral features

Samples are rare!

The requirement to extract chemical features by **wet-lab experiments** makes reference samples rare

Direct Regression is not possible!

Overlapping and scattered bands create **multicollinearity** in spectral features

Decorrelation in PLS allows efficient linear Regression

A preceeded **decorrelation step** in PLS allows allows linear regression

Samples are rare!

The requirement to extract chemical features by **wet-lab experiments** makes reference samples rare

Direct Regression is not possible!

Overlapping and scattered bands create **multicollinearity** in spectral features

Non-Linear Spectral Dependencies in NIR

Combined vibrational modes in NIR create highly **non-linear conditional dependencies** between spectral features

Decorrelation in PLS allows efficient linear Regression

A preceeded **decorrelation step** in PLS allows allows linear regression

PLS based non-linear Regression is not possible!

The **linear decorrelation** step in PLS destroys non-linear relationships

Samples are rare!

The requirement to extract chemical features by **wet-lab experiments** makes reference samples rare

Non-Linear Spectral Dependencies in NIR

Combined vibrational modes in NIR create highly **non-linear conditional dependencies** between spectral features

Direct Regression is not possible!

Overlapping and scattered bands create **multicollinearity** in spectral features

Decorrelation in PLS allows efficient linear Regression

A preceeded **decorrelation step** in PLS allows allows linear regression

PLS based non-linear Regression is not possible!

The **linear decorrelation** step in PLS destroys non-linear relationships

Non-Linear Decorrelation?

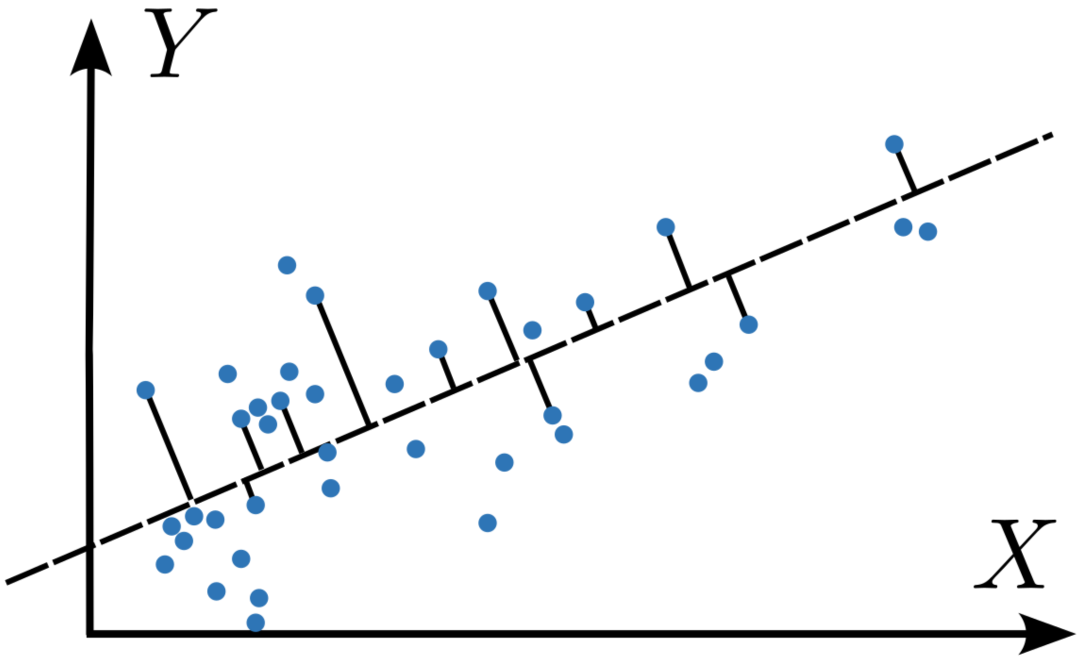
Part 3

Energy based Calibration

What is
non-linear Decorrelation?

Linear Correlations

„Multicollinearity“

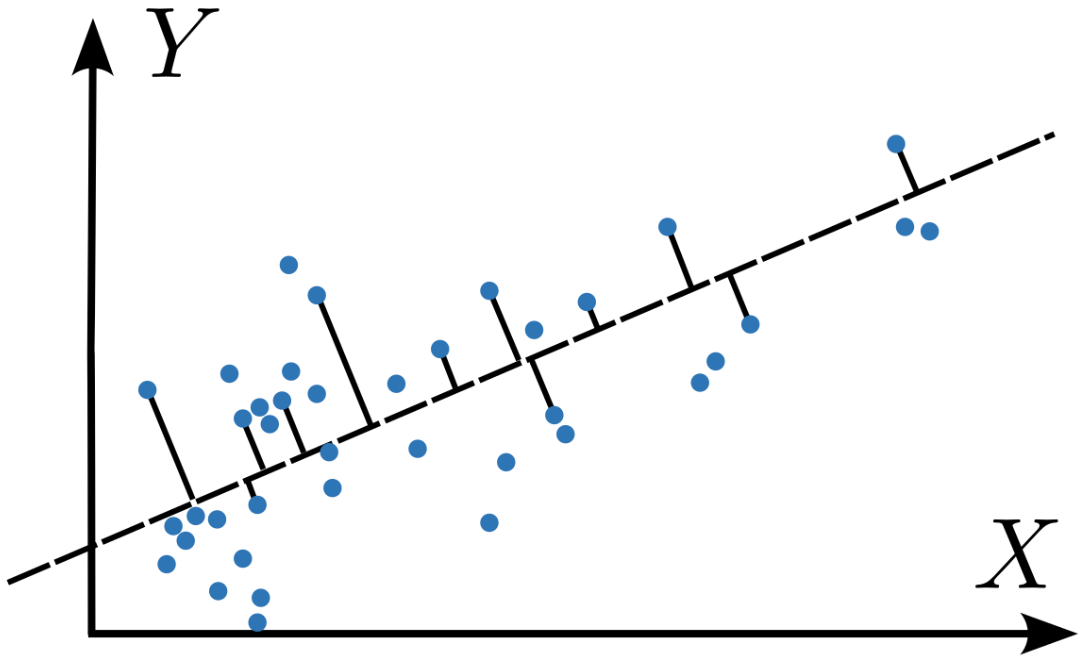


Linear Correlations

„Multicollinearity“

Linear Decorrelation

Orthogonal Projection
to first k
Principal Components

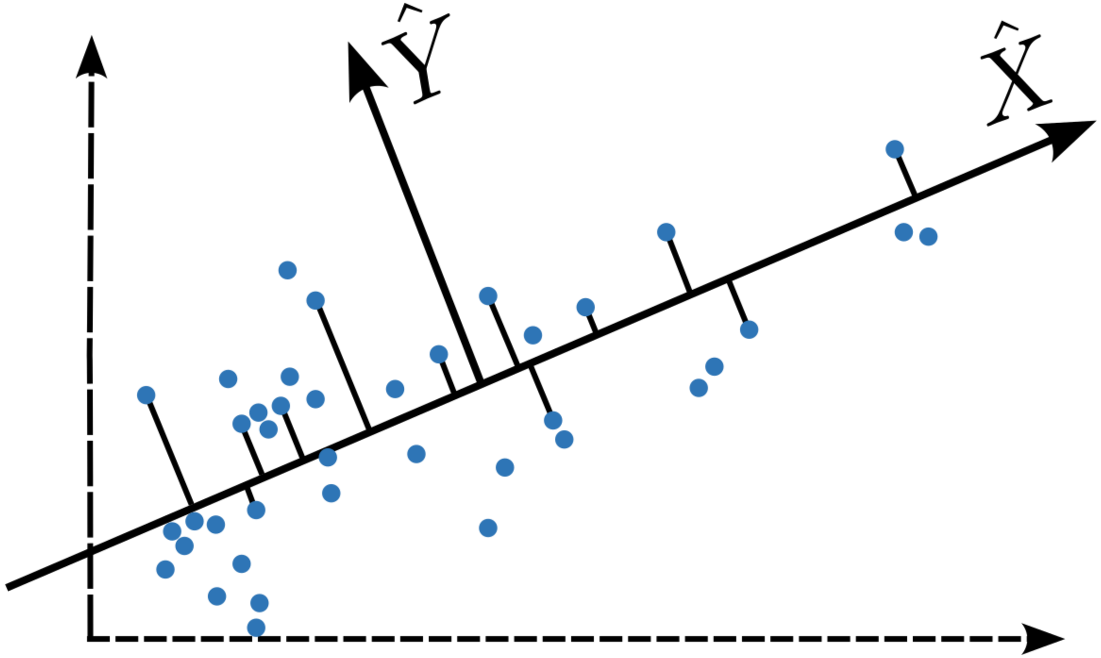


Linear Correlations

„Multicollinearity“

Linear Decorrelation

Orthogonal Projection
to first k
Principal Components



Linear Correlations

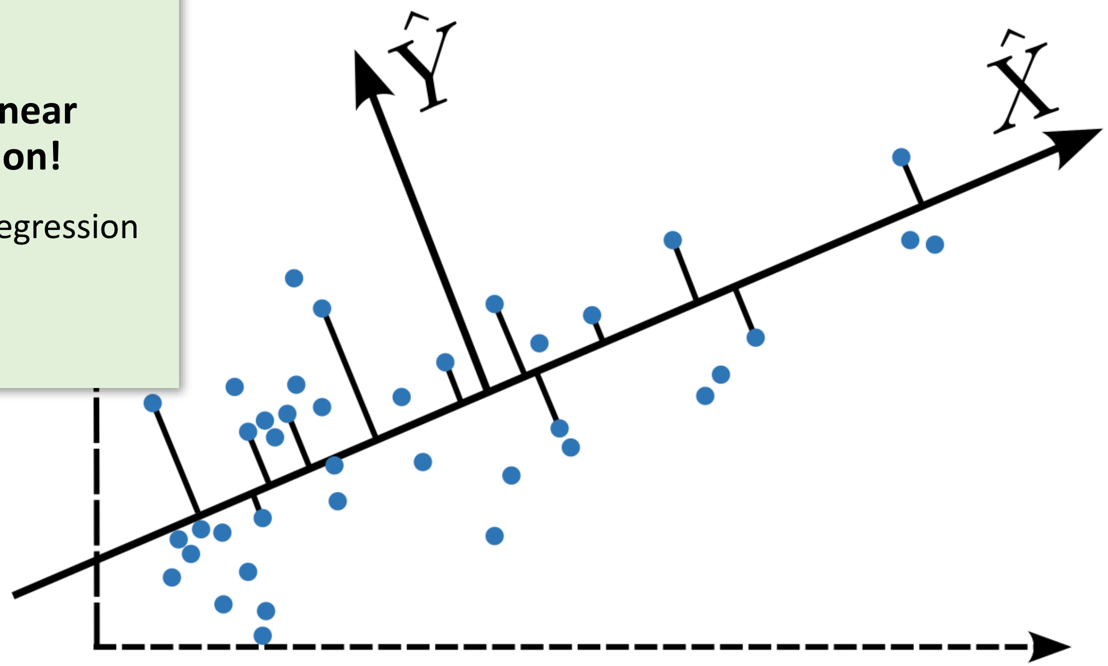
„Multicollinearity“

Allows Linear Regression!

e.g. PC-, PLS-Regression

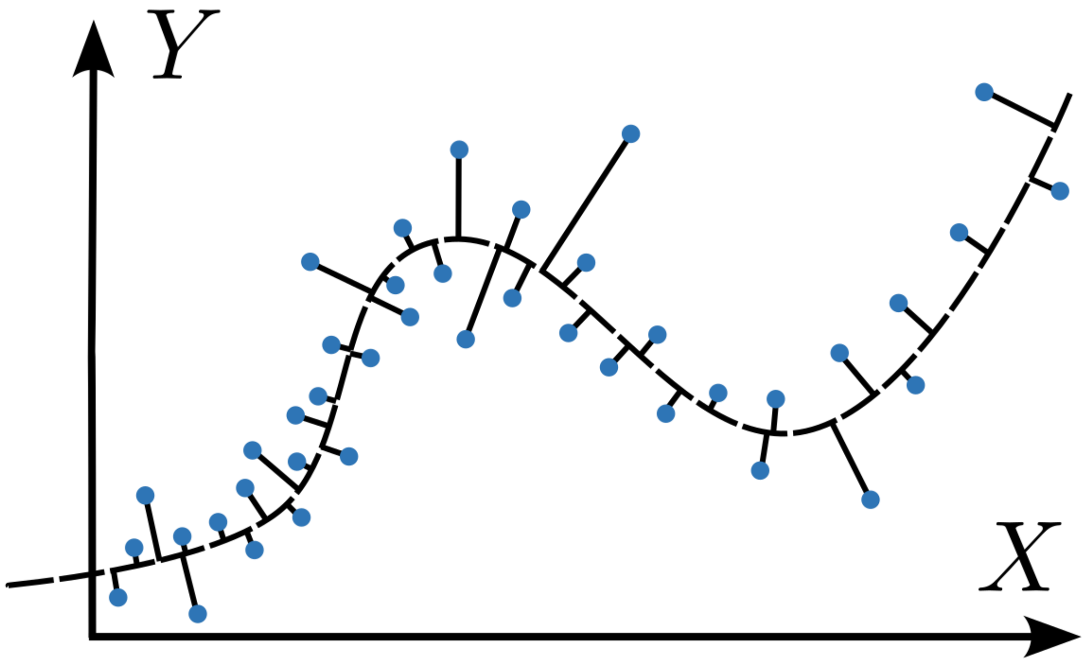
Linear Decorrelation

Orthogonal Projection
to first k
Principal Components



Generic Correlations

Linear or non-linear

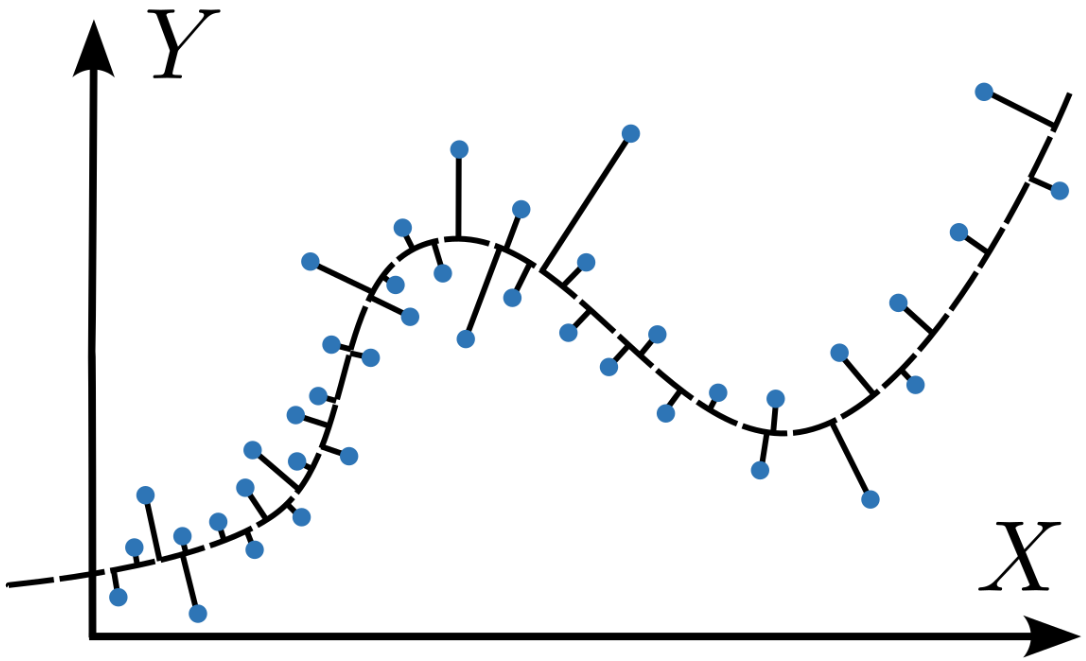


Generic Correlations

Linear or non-linear

Non-Linear Decorrelation

Orthogonal Projection to k dimensional **Principal Manifold**

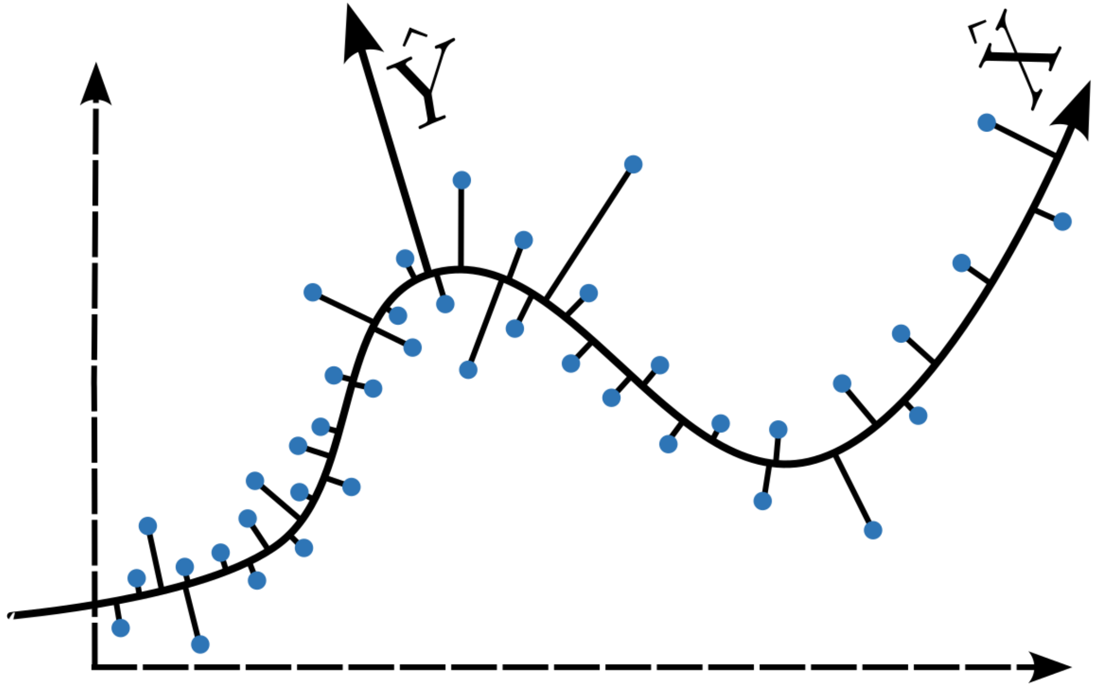


Generic Correlations

Linear or non-linear

Non-Linear Decorrelation

Orthogonal Projection
to k dimensional
Principal Manifold



Generic Correlations

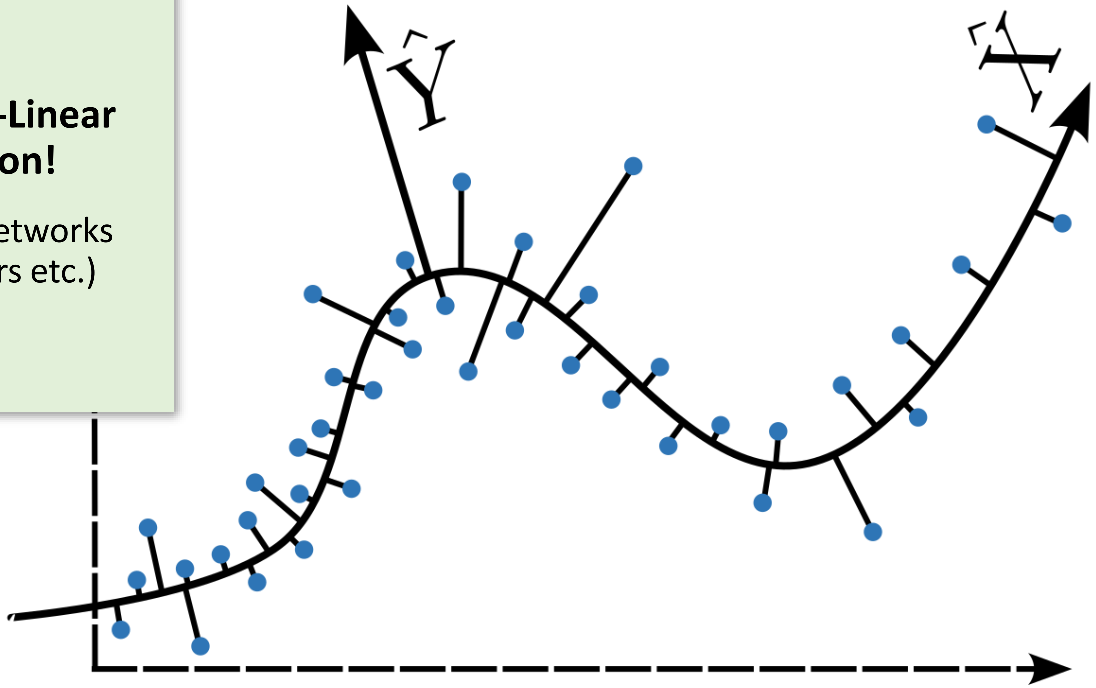
Linear or non-linear

Non-Linear Decorrelation

Orthogonal Projection to k dimensional **Principal Manifold**

Allows Non-Linear Regression!

e.g. Neural Networks (Transformers etc.)



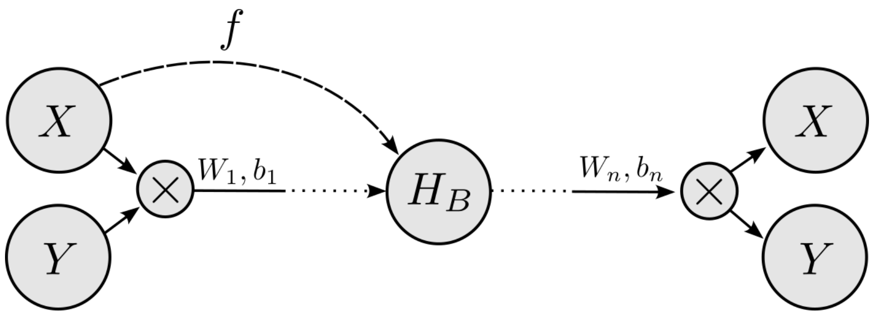
How to model
non-linear Decorrelation?

Approach 1: *Directed Graphical Model*

Directed mapping to latent space

Map features onto itself through a latent bottleneck layer and **minimize Error**

Approach 1: *Directed Graphical Model*



Directed mapping to latent space

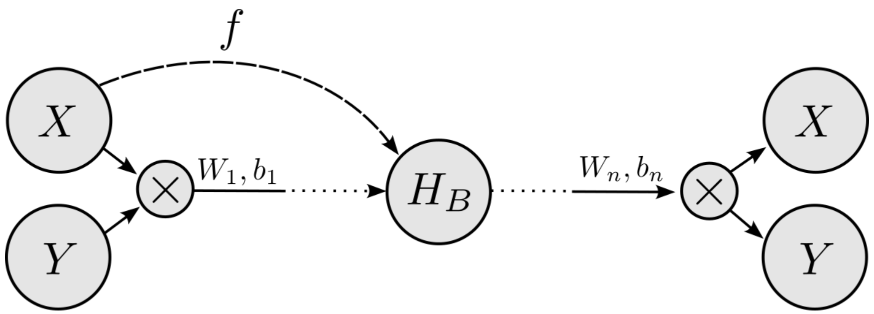
Map features onto itself through a latent bottleneck layer and **minimize Error**

Autoencoder

$\boldsymbol{W_i}$ Weights
$\boldsymbol{b_i}$ Bias

+ *Hyperparameters*

Approach 1: *Directed Graphical Model*



Directed mapping to latent space

Map features onto itself through a latent bottleneck layer and **minimize Error**

Autoencoder

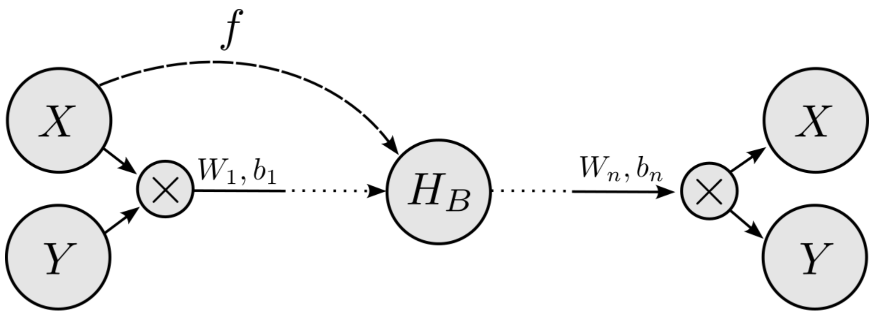
$\boldsymbol{W_i}$ Weights
$\boldsymbol{b_i}$ Bias

+ *Hyperparameters*

Not Generalizable!

does not equally consider information from different features due to a vanishing error gradient

Approach 1: *Directed Graphical Model*



Directed mapping to latent space

Map features onto itself through a latent bottleneck layer and **minimize Error**

Autoencoder

$\boldsymbol{W_i}$ Weights
$\boldsymbol{b_i}$ Bias

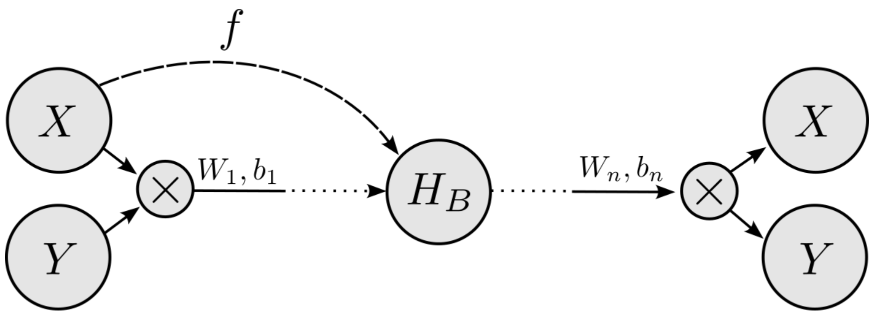
+ *Hyperparameters*

?

Not Generalizable!

does not equally consider information from different features due to a vanishing error gradient

Approach 1: *Directed Graphical Model*



t

Example

Let X_1 and X_2 be highly similar

Autoencoder

$\boldsymbol{W_i}$ Weights
$\boldsymbol{b_i}$ Bias

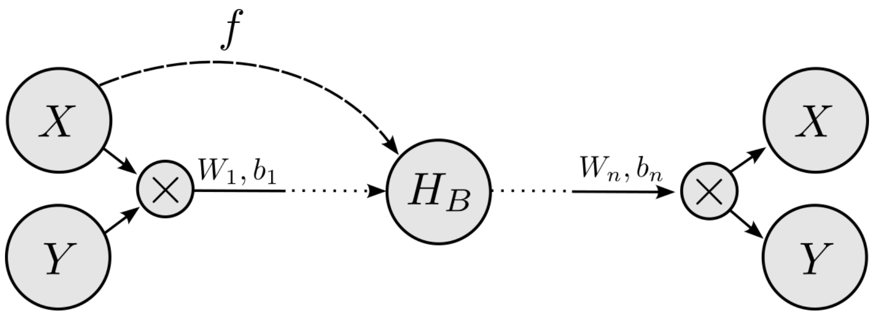
+ *Hyperparameters*

?

Not Generalizable!

does not equally consider information from different features due to a vanishing error gradient

Approach 1: *Directed Graphical Model*



Example

Let X_1 and X_2 be highly similar

Optimization

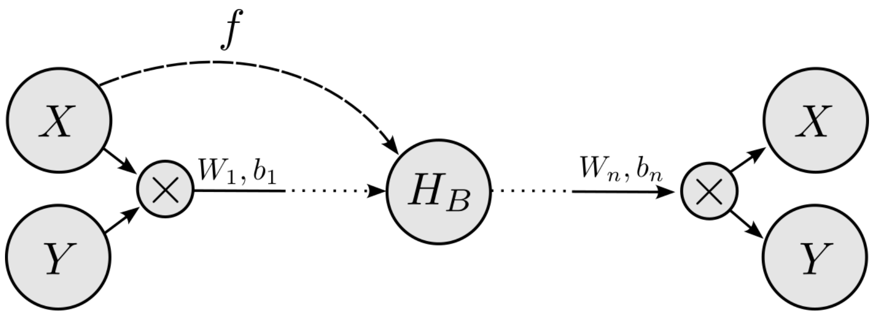
The error is minimized if f depends on X_1 **or** X_2 **or** any combination of them

?

Not Generalizable!

does not equally consider information from different features due to a vanishing error gradient

Approach 1: *Directed Graphical Model*



Example

Let X_1 and X_2 be highly similar

Optimization

The error is minimized if f depends on X_1 **or** X_2 **or** any combination of them

!

Out-of-Calibration Sample

If X_1 and X_2 are not highly similar then f only by chance creates an adequate latent presentation!

Approach 1: *Directed Graphical Model*

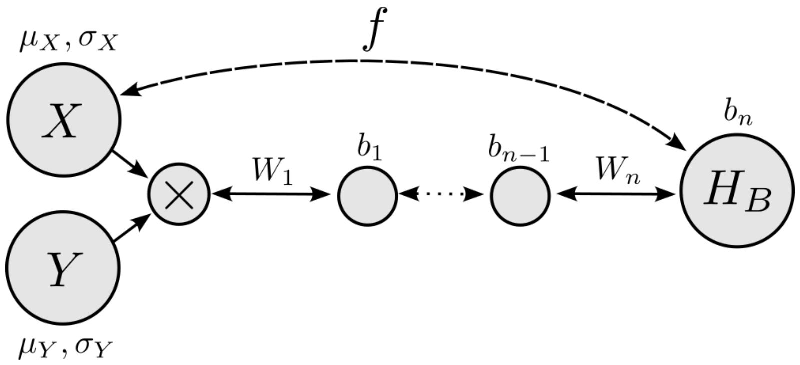
Approach 2: *Undirected Graphical Model*

Undirected mapping to latent space

Regard the mapping between features and the bottleneck layer mutually and **minimize Energy**

patrick.michl@gmail.com

Approach 2: *Undirected Graphical Model*



Undirected mapping to latent space

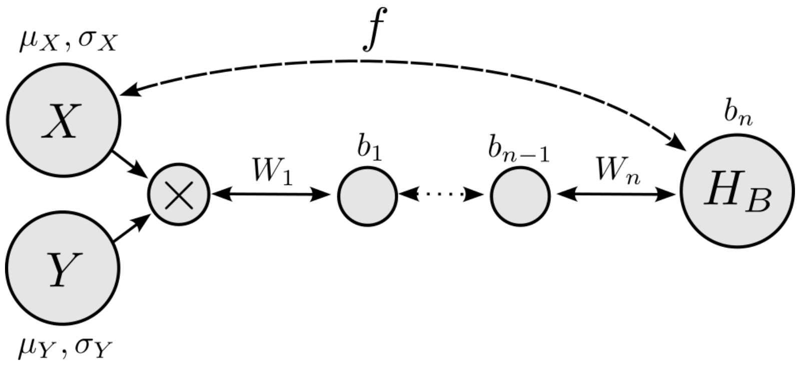
Regard the mapping between features and the bottleneck layer mutually and **minimize Energy**

Deep Boltzmann Machine

$\boldsymbol{W_i}$ Weights
$\boldsymbol{b_i}$ Bias
$\boldsymbol{\mu_X}, \boldsymbol{\sigma_X}$ Dist. of X
$\boldsymbol{\mu_X}, \boldsymbol{\sigma_X}$ Dist. of Y

\+ *Hyperparameters*

Approach 2: *Undirected Graphical Model*



Undirected mapping to latent space

Regard the mapping between features and the bottleneck layer mutually and **minimize Energy**

Deep Boltzmann Machine

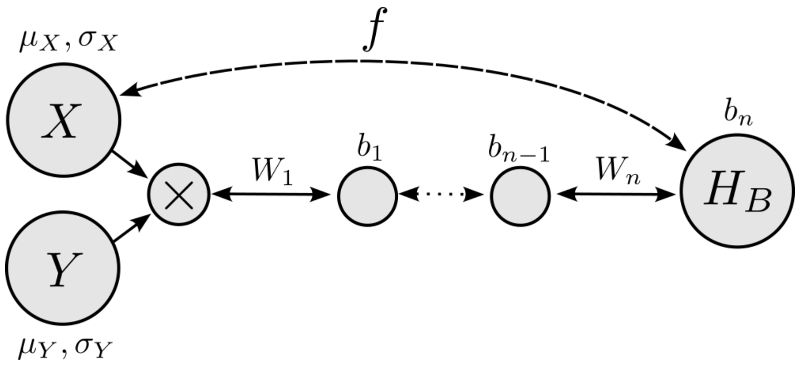
$\boldsymbol{W_i}$ Weights
$\boldsymbol{b_i}$ Bias
$\boldsymbol{\mu_X}, \boldsymbol{\sigma_X}$ Dist. of X
$\boldsymbol{\mu_X}, \boldsymbol{\sigma_X}$ Dist. of Y

\+ *Hyperparameters*

Generalizable!

equally considers information from different features

Approach 2: *Undirected Graphical Model*



Undirected mapping to latent space

Regard the mapping between features and the bottleneck layer mutually and **minimize Energy**

Deep Boltzmann Machine

$\boldsymbol{W_i}$ Weights
$\boldsymbol{b_i}$ Bias
$\boldsymbol{\mu_X}, \boldsymbol{\sigma_X}$ Dist. of X
$\boldsymbol{\mu_X}, \boldsymbol{\sigma_X}$ Dist. of Y

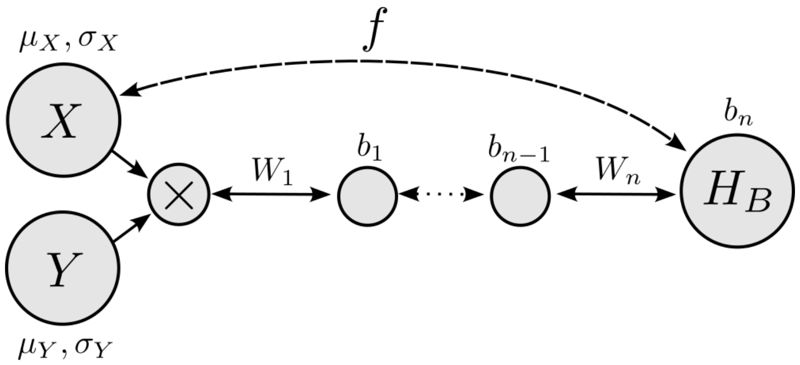
+ *Hyperparameters*

?

Generalizable!

equally considers information from different features

Approach 2: *Undirected Graphical Model*



Example

Let X_1 and X_2 be highly similar

Deep Boltzmann Machine

$\boldsymbol{W_i}$ Weights
$\boldsymbol{b_i}$ Bias
$\boldsymbol{\mu_X}, \boldsymbol{\sigma_X}$ Dist. of X
$\boldsymbol{\mu_X}, \boldsymbol{\sigma_X}$ Dist. of Y

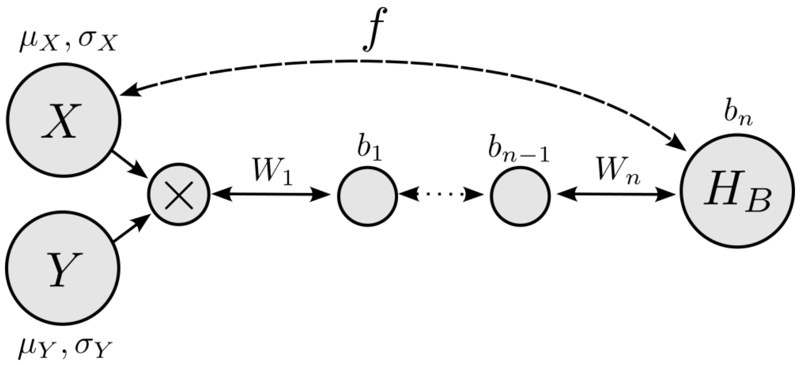
+ *Hyperparameters*

?

Generalizable!

equally considers information from different features

Approach 2: *Undirected Graphical Model*



Example

Let X_1 and X_2 be highly similar

Optimization

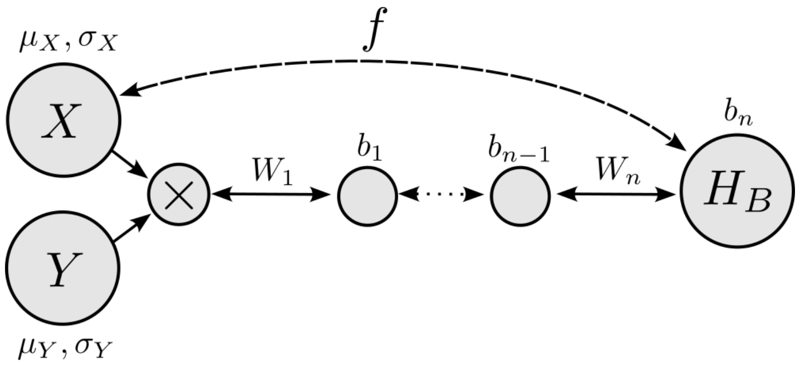
The energy is only minimized if f depends **equally** on X_1 and X_2

?

Generalizable!

equally considers information from different features

Approach 2: *Undirected Graphical Model*



Example

Let X_1 and X_2 be highly similar

Optimization

The energy is only minimized if f depends equally on X_1 **and** X_2

!

Out-of-Calibration Sample

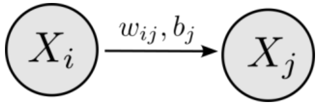
If X_1 and X_2 are not highly similar then f still creates an adequate latent presentation!

What is

Energy based Modeling?

Energy based Calibration

Regression Model

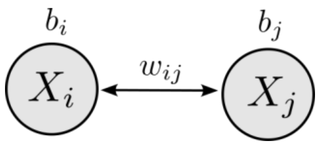


Directed Connections

Weights model one direction:

w_{ij}

Energy based Model



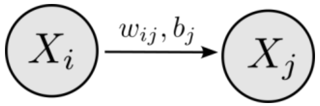
Undirected Connections

Weights model both directions:

$w_{ij} = w_{ji}$

Energy based Calibration

Regression Model



Directed Connections

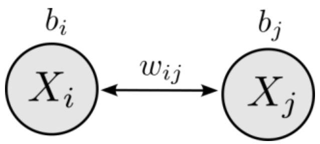
Weights model one direction:

w_{ij}

Minimizes Error

Parameters minimize the error between targets and predictions

Energy based Model



Undirected Connections

Weights model both directions:

$w_{ij} = w_{ji}$

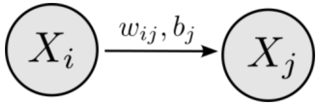
Minimizes Energy

Parameters minimize the pairwise energies of connected vertices:

Pos. Corr.: $w_{ij} > 0$
Neg. Corr.: $w_{ij} < 0$

Energy based Calibration

Regression Model



Directed Connections

Weights model one direction:

w_{ij}

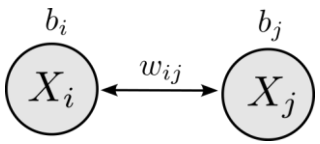
Minimizes Error

Parameters minimize the error between targets and predictions

Calculates Predictions

Predictions **of targets** are directly calculated

Energy based Model



Undirected Connections

Weights model both directions:

$w_{ij} = w_{ji}$

Minimizes Energy

Parameters minimize the pairwise energies of connected vertices:

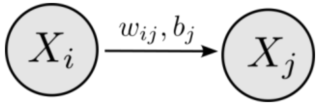
Pos. Corr.: $w_{ij} > 0$
Neg. Corr.: $w_{ij} < 0$

Approximates Predictions

Predictions of **all variables** are approximated by a Markov Chain using Gibbs Sampling

Energy based Calibration

Regression Model



Directed Connections

Weights model one direction:

w_{ij}

Minimizes Error

Parameters minimize the error between targets and predictions

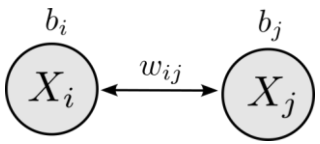
Calculates Predictions

Predictions **of targets** are directly calculated

Optimizes Prediction Function

Gradient Descent in the Errorlandscape approximates the prediction function

Energy based Model



Undirected Connections

Weights model both directions:

$w_{ij} = w_{ji}$

Minimizes Energy

Parameters minimize the pairwise energies of connected vertices:

Pos. Corr.: $w_{ij} > 0$
Neg. Corr.: $w_{ij} < 0$

Approximates Predictions

Predictions of **all variables** are approximated by a Markov Chain using Gibbs Sampling

Optimizes Joint Distribution

Gradient Descent in the Energylandscape approximates the joint distributon

Non-linear **Regression** in Energy based Calibration

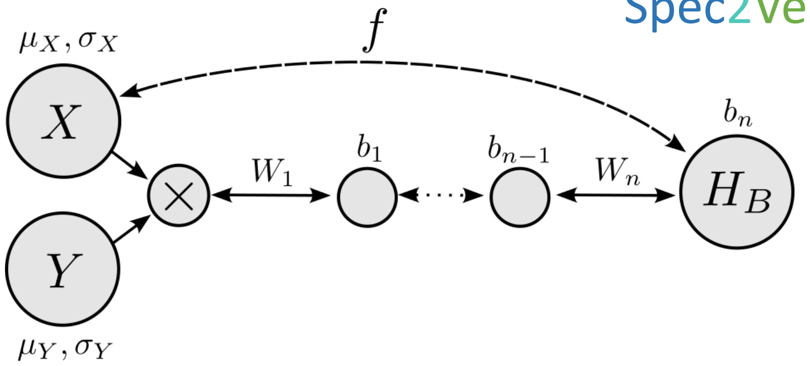
Consists of two Parts

Consists of two Parts

Part 1: Embedding

Non-linear decorrelation by Energy based model

Spec2Vec



Consists of two Parts

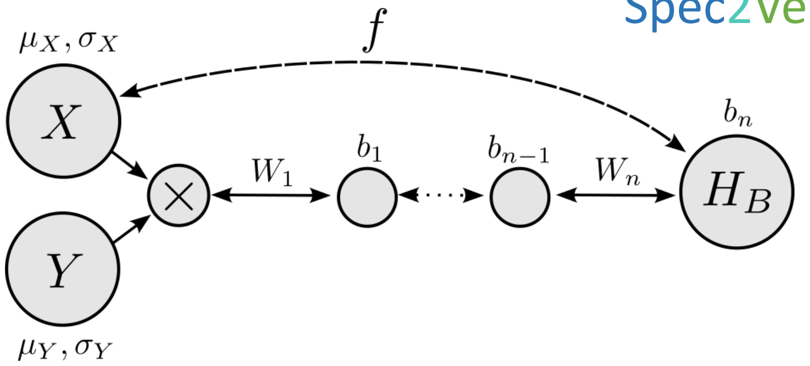
Part 1: Embedding

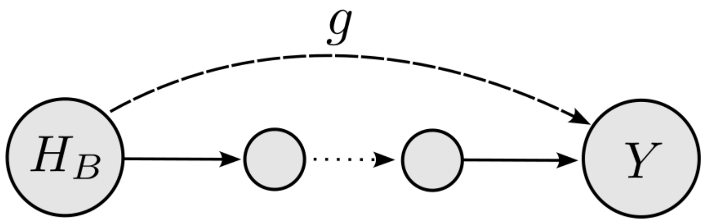
Non-linear decorrelation by Energy based model

Part 2: Regression

Non-linear regression by deep neural network (e.g. transformer etc.)

Spec2Vec





What can be concluded
about **Energy based Calibration**?

Non-Linear Decorrelation

Energy based modeling allows non-linear decorrelation of spectral features

Non-Linear Decorrelation

Energy based modeling allows non-linear decorrelation of spectral features

Increases Signal-to-Noise ratio

Non-linear decorrelation allows the separation of skewed noise (non-Gaussian)

patrick.michl@gmail.com

Energy Based Calibration

Non-Linear Decorrelation

Energy based modeling allows non-linear decorrelation of spectral features

Non-Linear Regression

Non-linear decorrelation allows non-linear regression of chemical features, e.g. by Neural Networks (Transformers, etc.)

Increases Signal-to-Noise ratio

Non-linear decorrelation allows the separation of skewed noise (non-Gaussian)

Energy Based Calibration

Non-Linear Decorrelation

Energy based modeling allows non-linear decorrelation of spectral features

Non-Linear Regression

Non-linear decorrelation allows non-linear regression of chemical features, e.g. by Neural Networks (Transformers, etc.)

Intricates Feature Selection

The non-linearity intricates the selection of non-linear spectral features

Increases Signal-to-Noise ratio

Non-linear decorrelation allows the separation of skewed noise (non-Gaussian)

Non-Linear Decorrelation

Energy based modeling allows non-linear decorrelation of spectral features

Non-Linear Regression

Non-linear decorrelation allows non-linear regression of chemical features, e.g. by Neural Networks (Transformers, etc.)

Increases Signal-to-Noise ratio

Non-linear decorrelation allows the separation of skewed noise (non-Gaussian)

Intricates Feature Selection

The non-linearity intricates the selection of non-linear spectral features

Non-linear Feature Selection?

patrick.michl@gmail.com

Non-Linear Decorrelation

Energy based modeling allows non-linear decorrelation of spectral features

Non-Linear Regression

Non-linear decorrelation allows non-linear regression of chemical features, e.g. by Neural Networks (Transformers, etc.)

Increases Signal-to-Noise ratio

Non-linear decorrelation allows the separation of skewed noise (non-Gaussian)

Intricates Feature Selection

The non-linearity intricates the selection of non-linear spectral features

Intricates Sample Selection

The non-linearity intricates the selection of calibration samples and thus the **experimental design**

Non-linear Feature Selection?

Energy Based Calibration

Non-Linear Decorrelation

Energy based modeling allows non-linear decorrelation of spectral features

Non-Linear Regression

Non-linear decorrelation allows non-linear regression of chemical features, e.g. by Neural Networks (Transformers, etc.)

Increases Signal-to-Noise ratio

Non-linear decorrelation allows the separation of skewed noise (non-Gaussian)

Intricates Feature Selection

The non-linearity intricates the selection of non-linear spectral features

Intricates Sample Selection

The non-linearity intricates the selection of calibration samples and thus the **experimental design**

Non-linear Feature Selection?

Bayesian Sample Selection?

Part 4

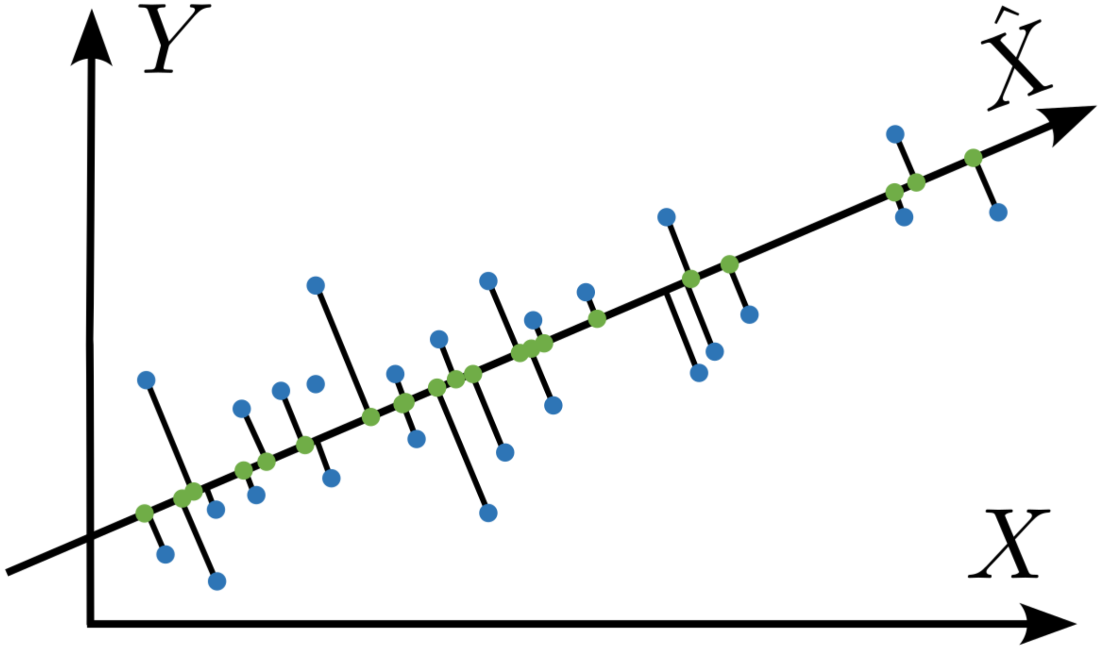
Non-linear feature selection

Differential Geometric
feature selection
for *Principal Components*

patrick.michl@gmail.com

Example (Line)

Feature Selection for
Principal Component

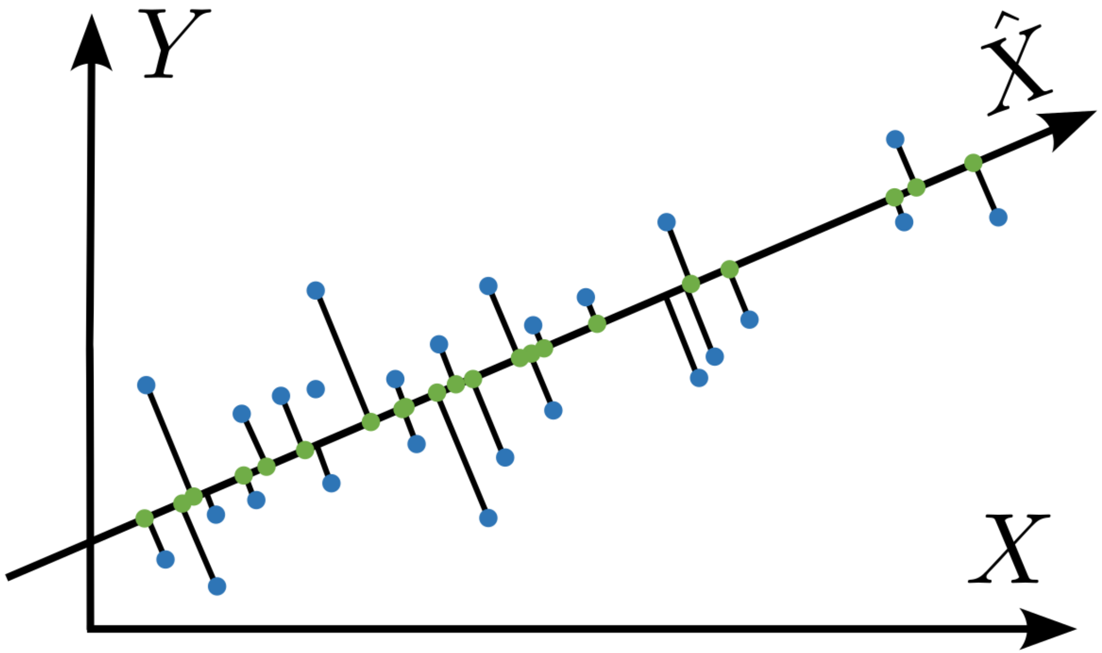


Example (Line)

Feature Selection for **Principal Component**

Approach 1: Statistical

Calculate empirical correlations:



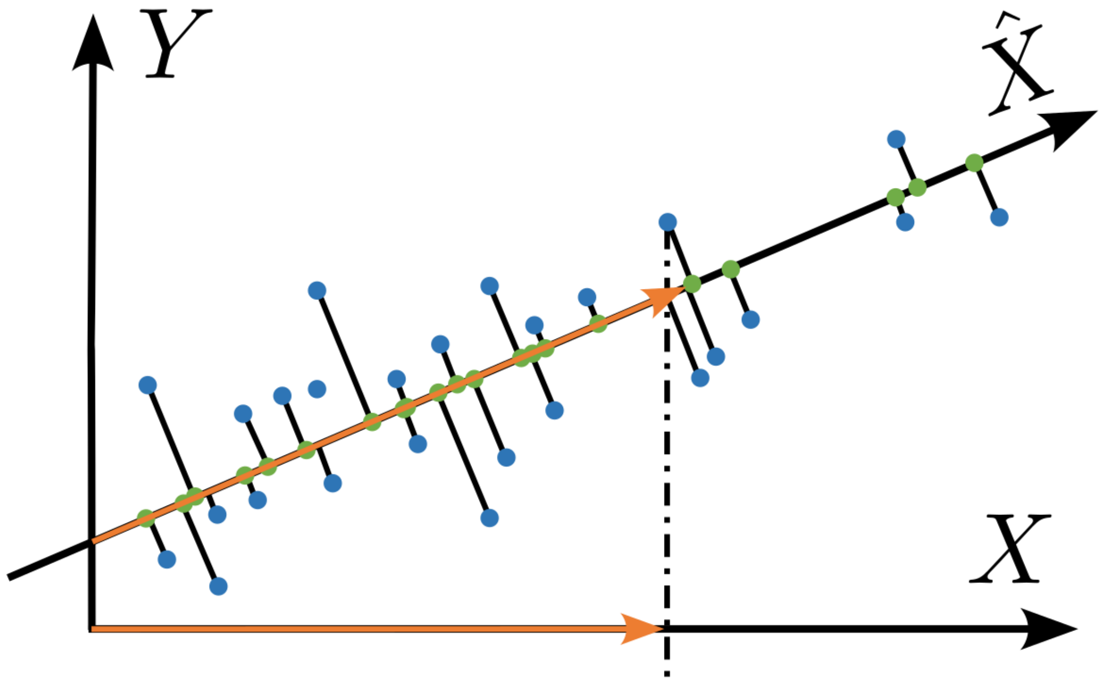
Example (Line)

Feature Selection for **Principal Component**

Approach 1: Statistical

Calculate empirical correlations:

$\mathrm{corr}(\hat{X}, X)$



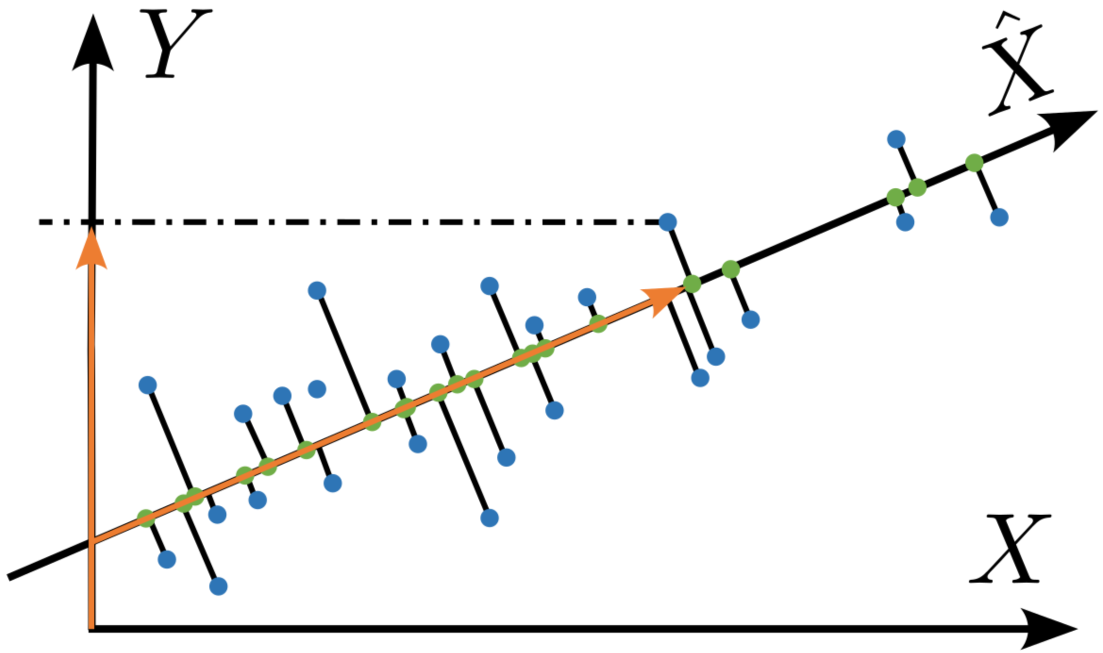
Example (Line)

Feature Selection for
Principal Component

Approach 1: Statistical

Calculate empirical correlations:

$$\mathrm{corr}(\hat{X}, X) > \mathrm{corr}(\hat{X}, Y)$$



Example (Line)

Feature Selection for **Principal Component**

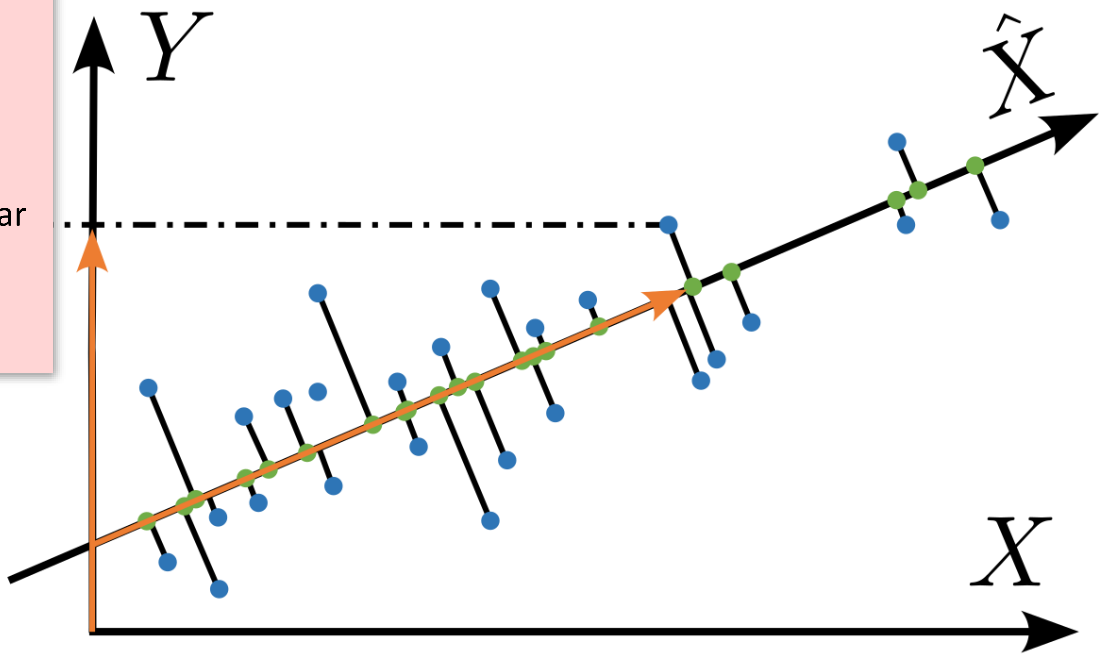
Approach 1 Statistical

Calculate empirical correlations:

$$\mathrm{corr}(\hat{X}, X) > \mathrm{corr}(\hat{X}, Y)$$

Does not allow generalization to non-linearity!

Due to missing non-linear empirical correlation

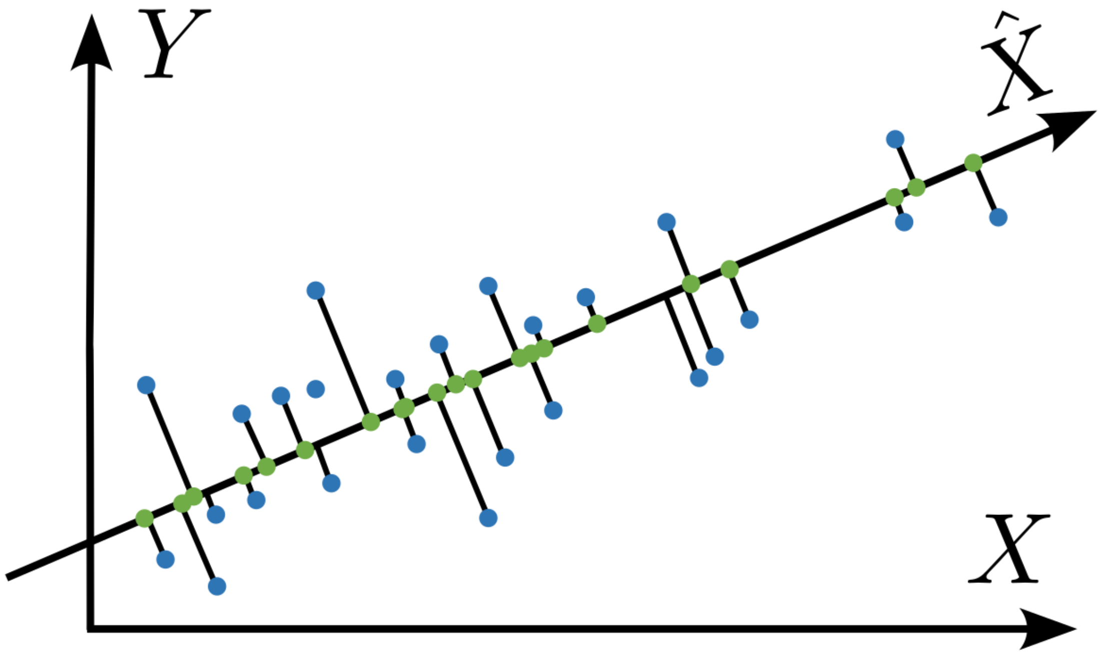


Example (Line)

Feature Selection for **Principal Component**

Approach 2: Geometrical

Calculate partial derivatives:



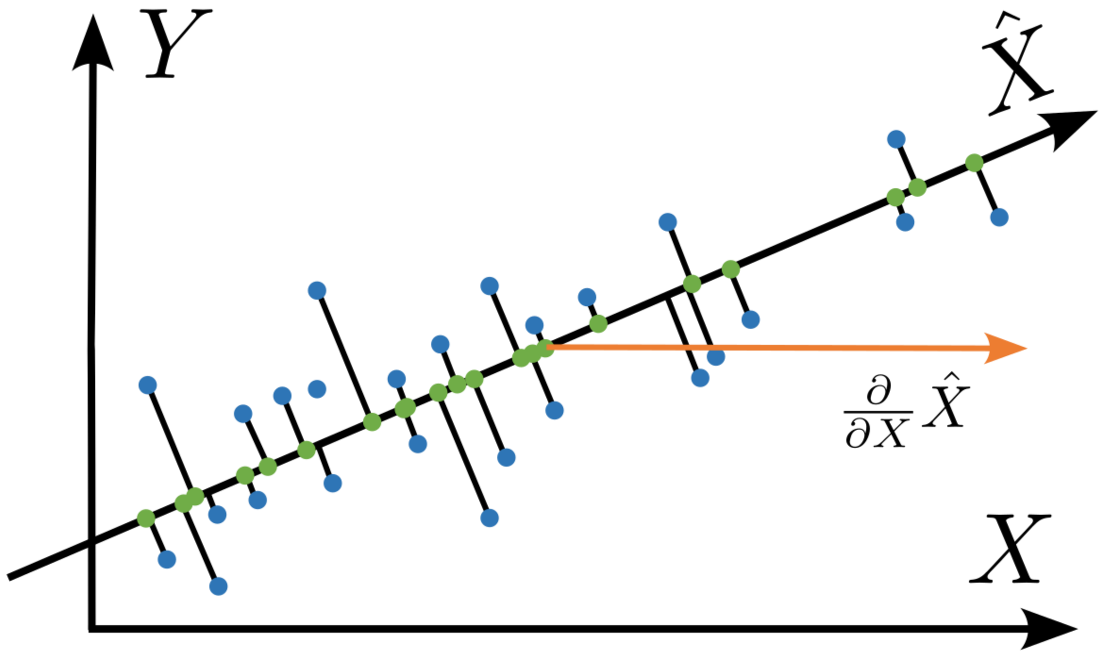
Example (Line)

Feature Selection for
Principal Component

Approach 2: Geometrical

Calculate partial derivatives:

$$\left|\frac{\partial}{\partial x}\hat{X}\right|$$



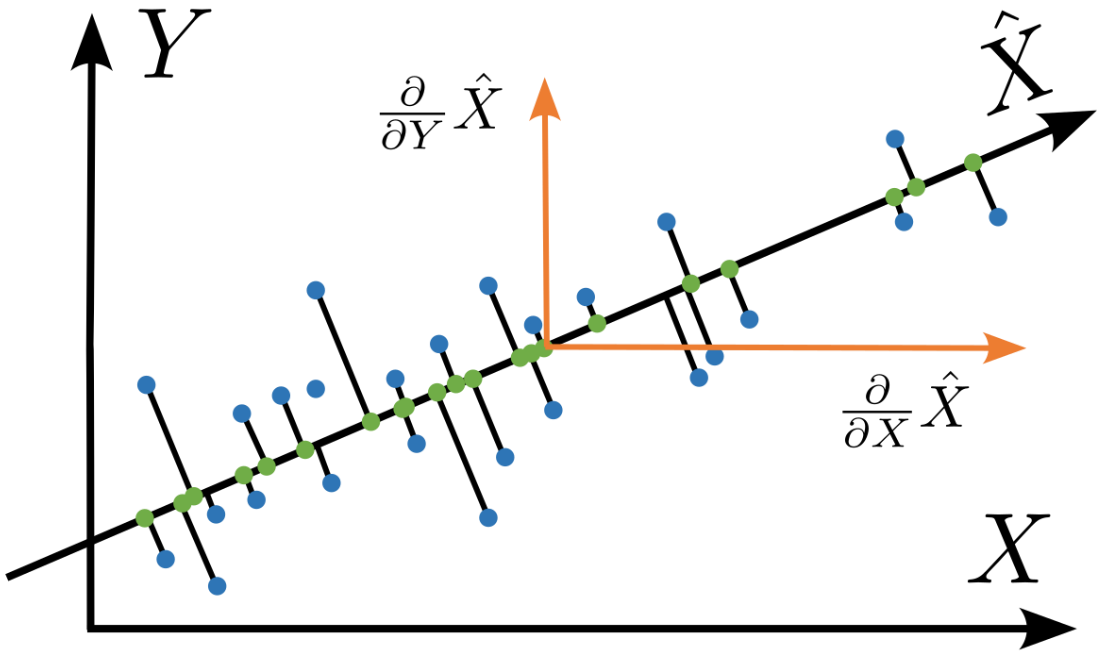
Example (Line)

Feature Selection for **Principal Component**

Approach 2: Geometrical

Calculate partial derivatives:

$$\left|\frac{\partial}{\partial X}\hat{X}\right| > \left|\frac{\partial}{\partial Y}\hat{X}\right|$$



Non-linear feature selection

Example (Line)

Feature Selection for **Principal Component**

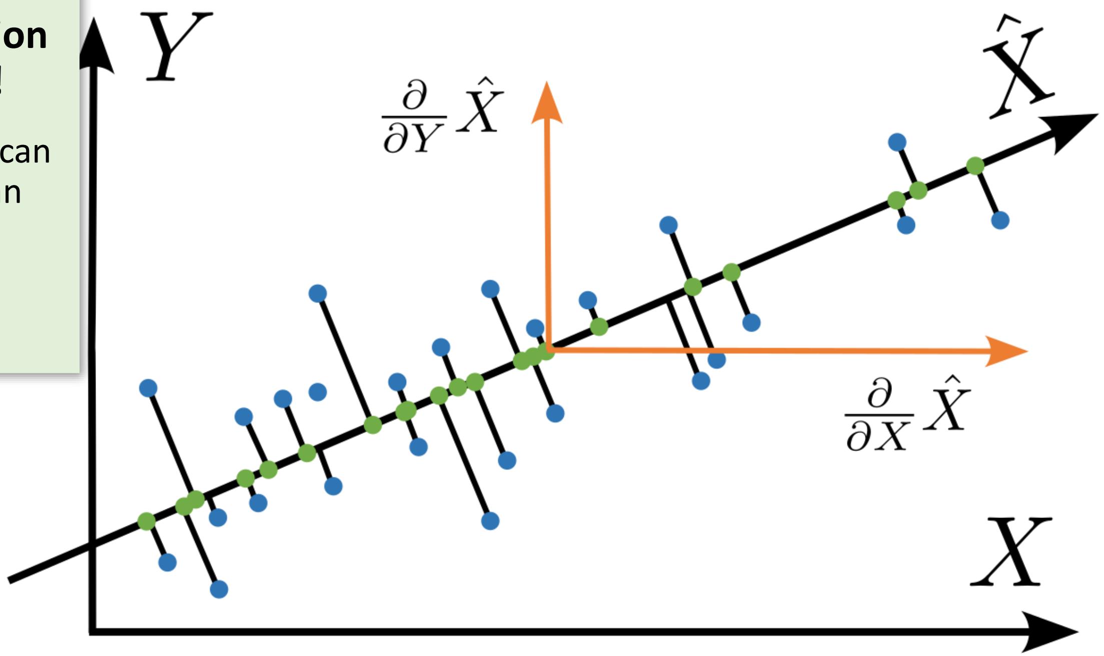
Approach 2: Geometrical

Calculate partial derivatives:

$$\left|\frac{\partial}{\partial X}\hat{X}\right| > \left|\frac{\partial}{\partial Y}\hat{X}\right|$$

Allows generalization to non-linearity!

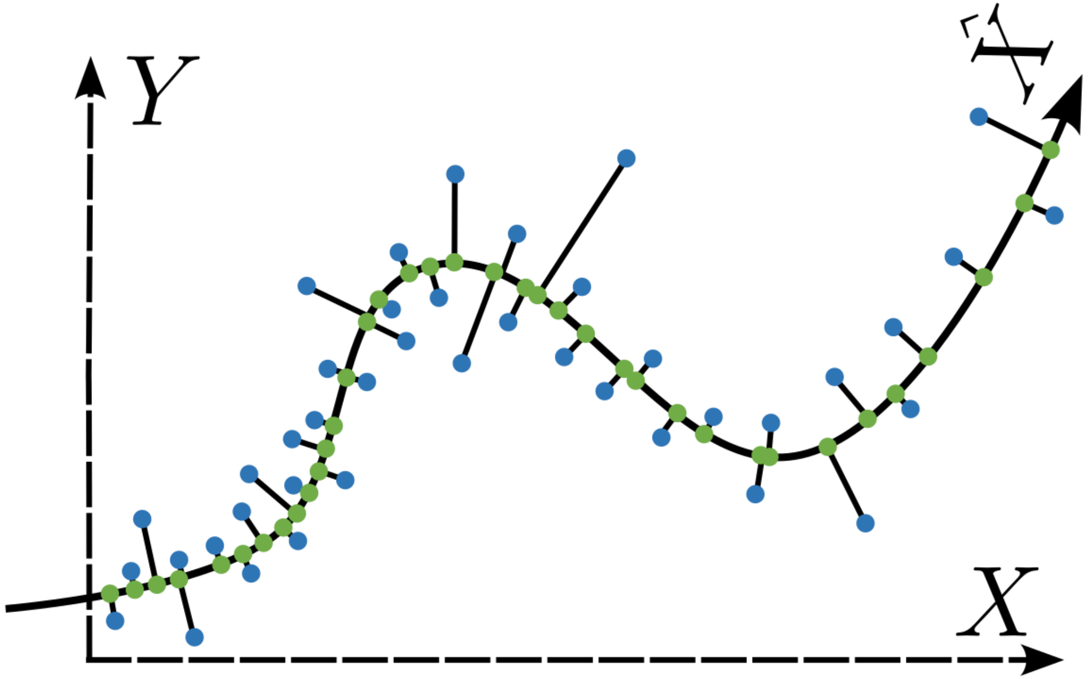
The partial derivative can be integrated over an underlying **principal manifold**



Differential Geometric
feature selection
for *Principal Manifolds*

Example (Curve)

Feature Selection for **Principal Curve**



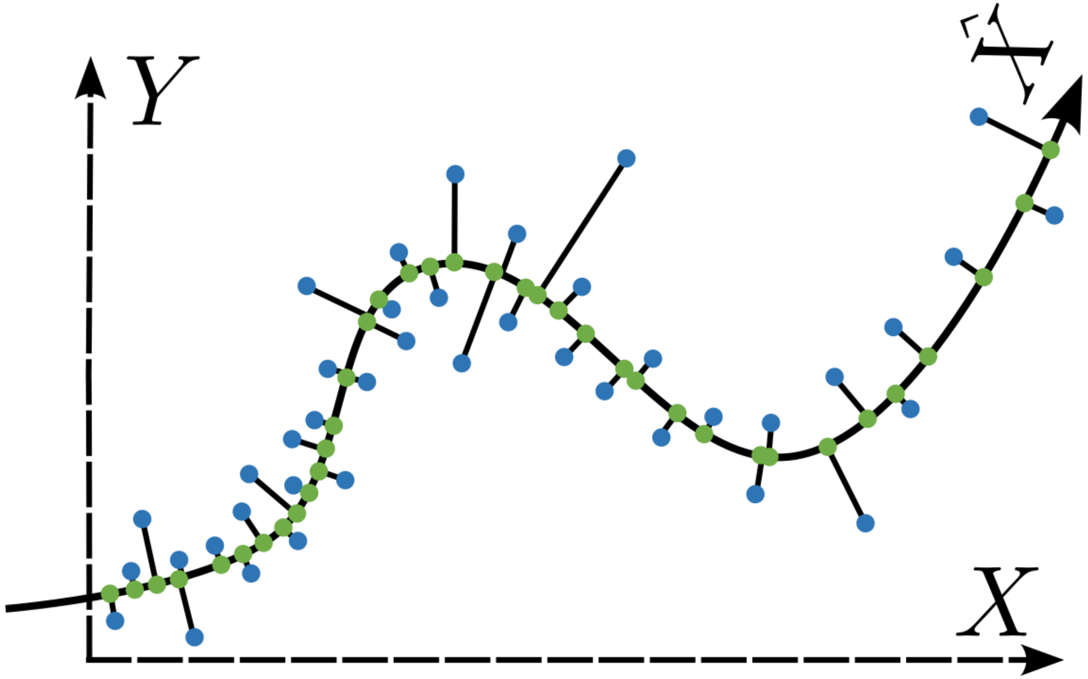
Non-linear feature selection

Example (Curve)

Feature Selection for **Principal Curve**

Approach 1: Analytical Integration

Calculate integrals of partial derivatives along curve

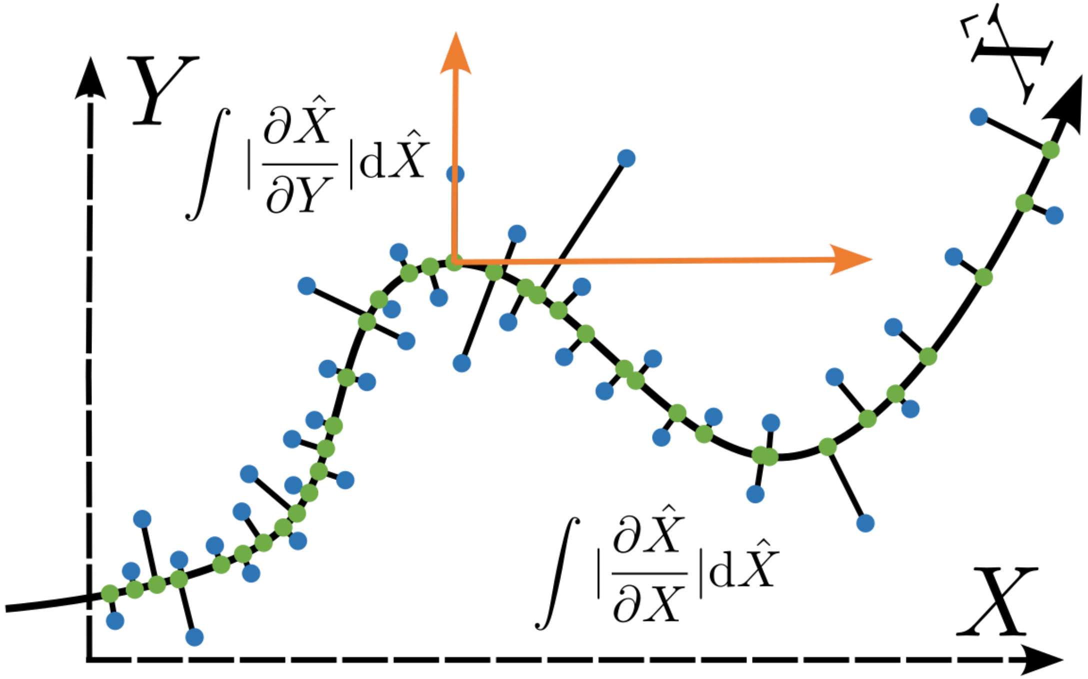


Example (Curve)

Feature Selection for
Principal Curve

Approach 1:
Analytical Integration

Calculate integrals of
partial derivatives
along curve



Example (Curve)

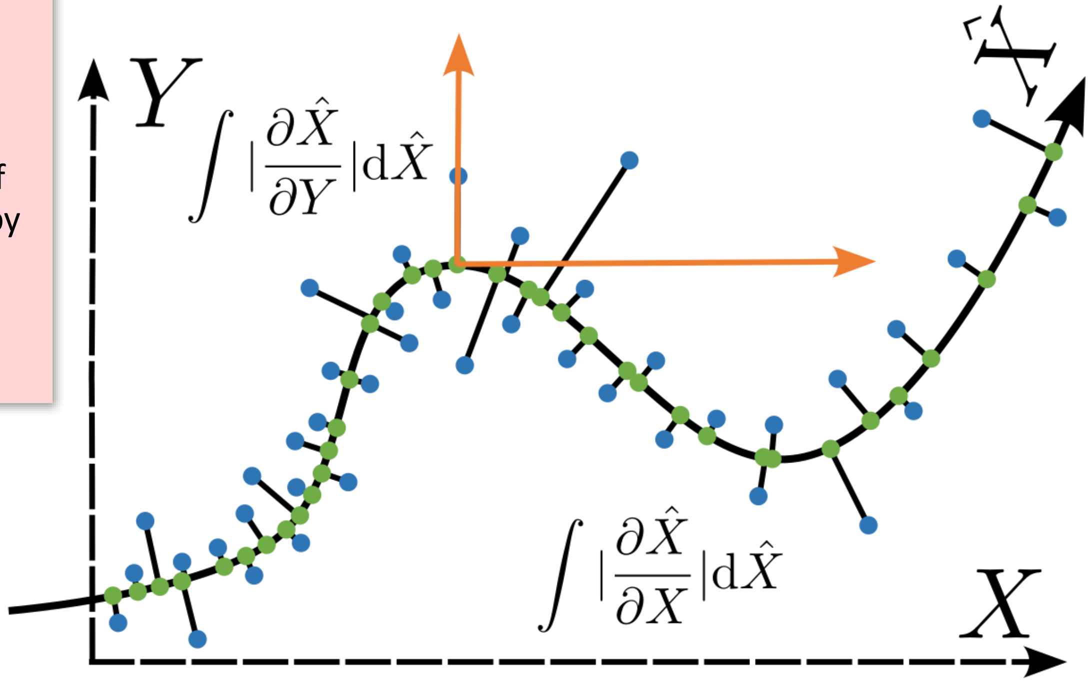
Feature Selection for
Principal Curve

Approach
Analytical Integration

Calculate integrals of
partial derivatives
along curve

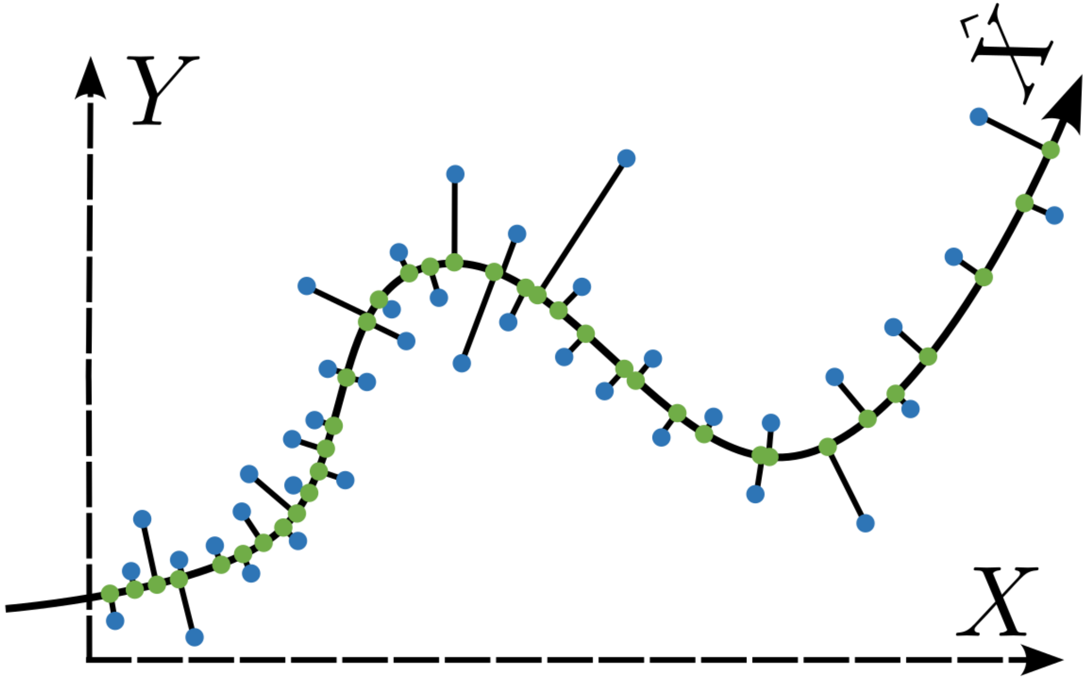
Not Computable!

Individual outcomes of
$\hat{X}$ are to be computed by
Gibbs Sampling



Example (Curve)

Feature Selection for
Principal Curve

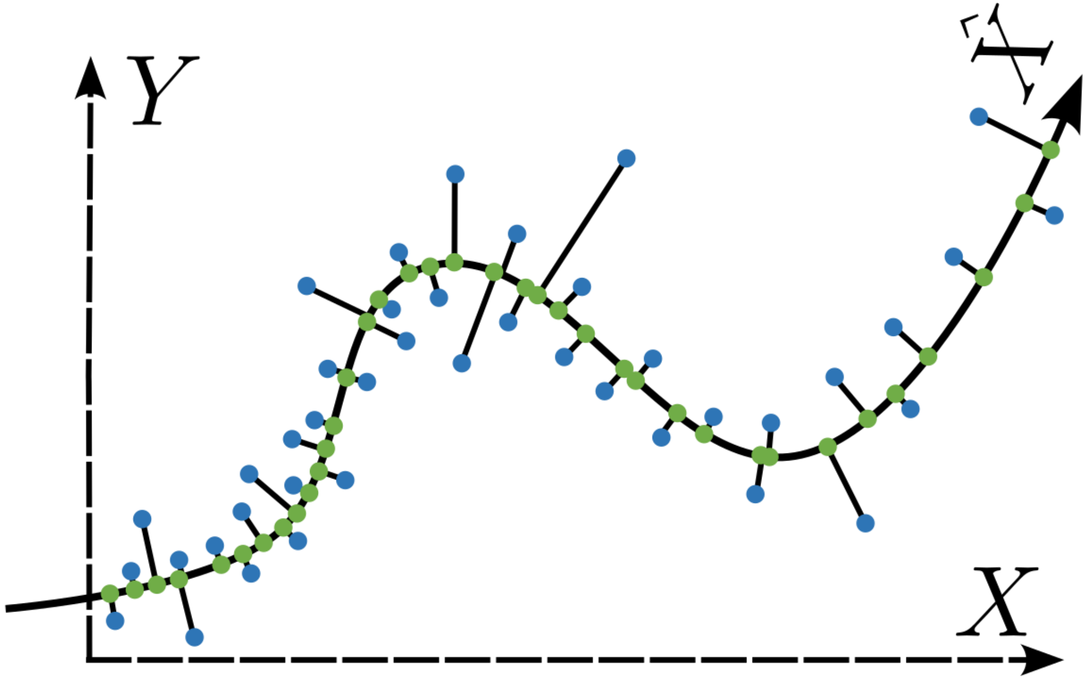


Example (Curve)

Feature Selection for
Principal Curve

**Approach 2:
Monte Carlo
Integration**

Approximate integrals
of partial derivatives
along curve

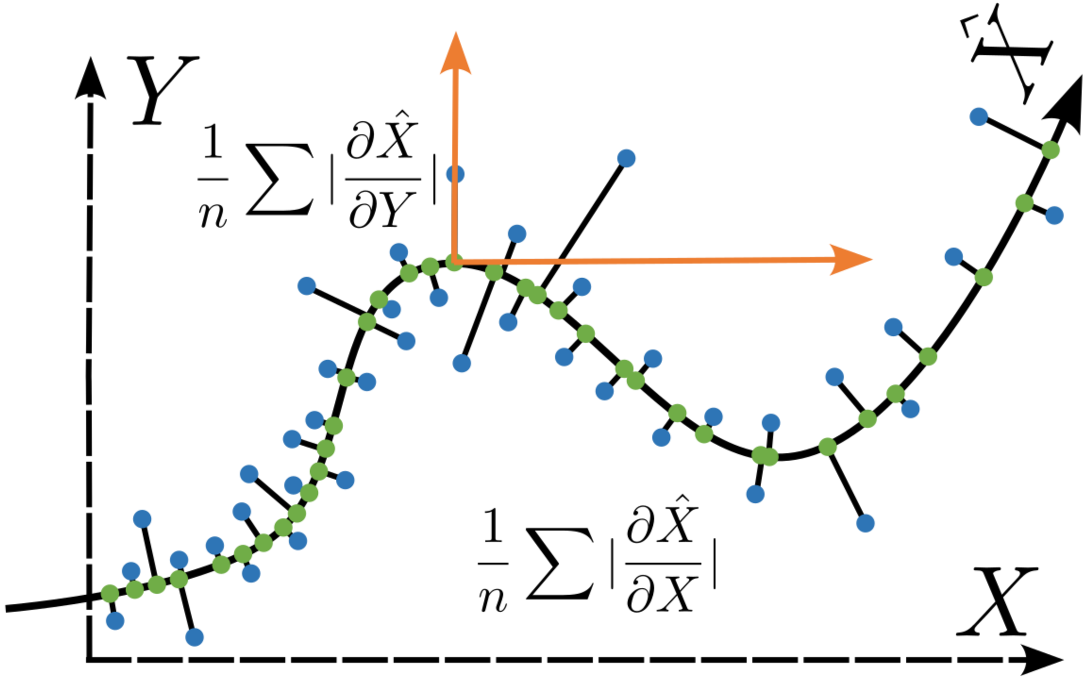


Example (Curve)

Feature Selection for **Principal Curve**

Approach 2: Monte Carlo Integration

Approximate integrals of partial derivatives along curve



Example (Curve)

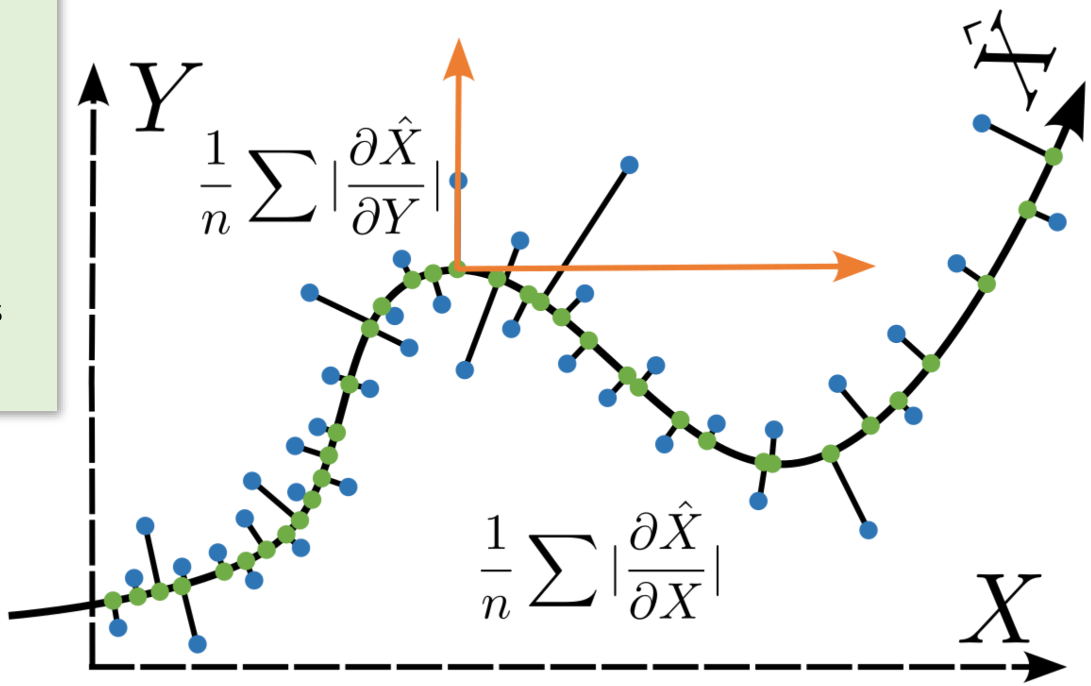
Feature Selection for **Principal Curve**

Approach 2 Monte Carlo Integration

Approximate integrals of partial derivatives along curve

Allows non-linear feature selection!

The empirical means converge towards the integrals under mild conditions



What can be concluded
about **non-linear feature selection**?

Feature Selection

for non-linear
models works!

Non-linear feature selection

Feature Selection

for non-linear
models works!

Sufficiency

The non-linearity
allows **higher accuracy**
at same number of spectral
features

Non-linear feature selection

Feature Selection

for non-linear
models works!

Sufficiency

The non-linearity
allows **higher accuracy**
at same number of spectral
features

**Improves
Parallel Measures**

e.g. by diffraction grating

Non-linear feature selection

Feature Selection

for non-linear
models works!

Sufficiency

The non-linearity
allows **higher accuracy**
at same number of spectral

Efficiency

The non-linearity
allows **fewer spectral features**
at same accuracy

**Improves
Parallel Measures**

e.g. by diffraction grating

patrick.michl@gmail.com

Non-linear feature selection

Feature Selection

for non-linear
models works!

Sufficiency

The non-linearity
allows **higher accuracy**
at same number of spectral

Efficiency

The non-linearity
allows **fewer spectral features**
at same accuracy

**Improves
Parallel Measures**

e.g. by diffraction grating

**Accelerates
Sequential Measures**

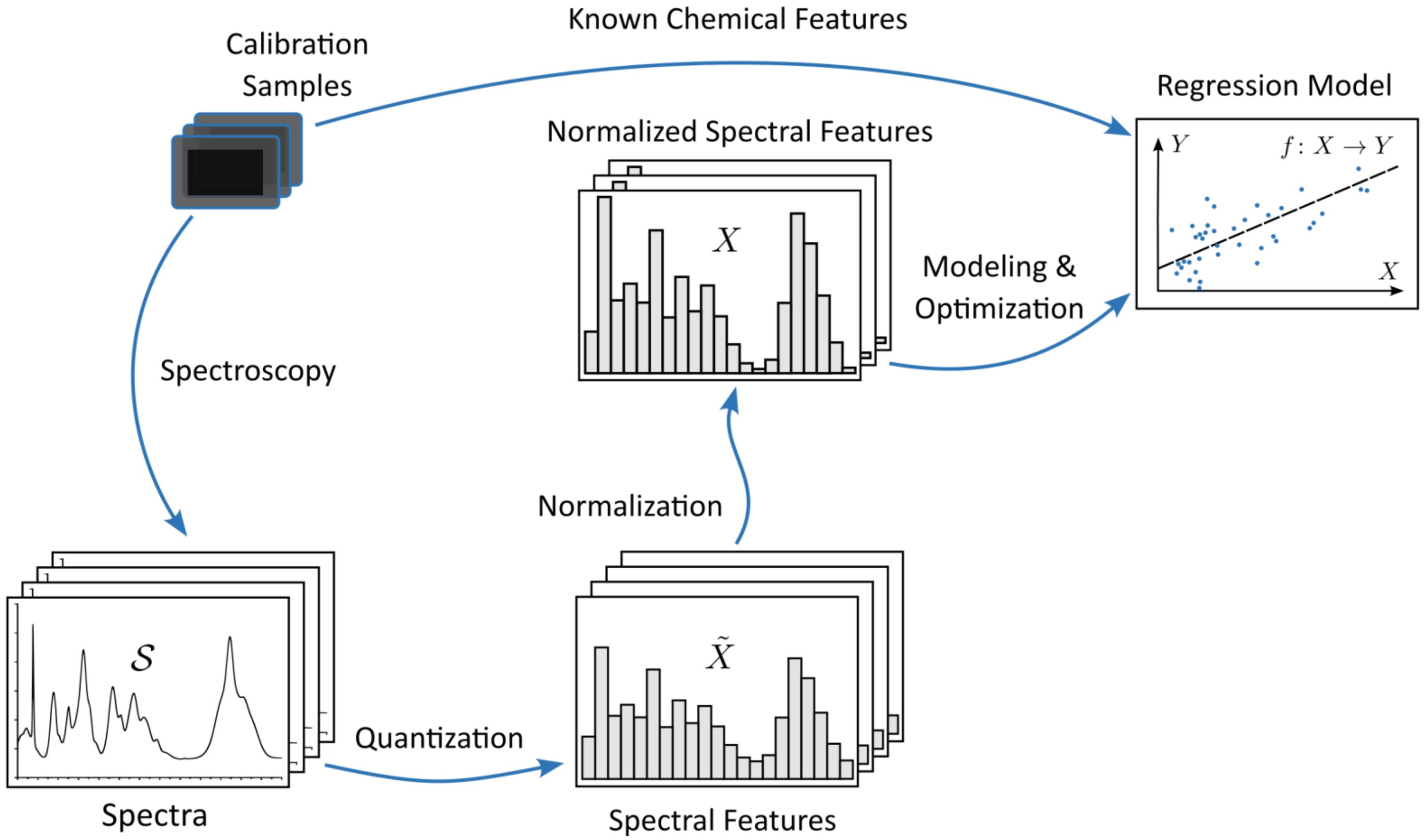
Part 5

Bayesian sample selection

Bayesian **Experimental Design**

Bayesian sample selection

Initial Selection of calibration samples

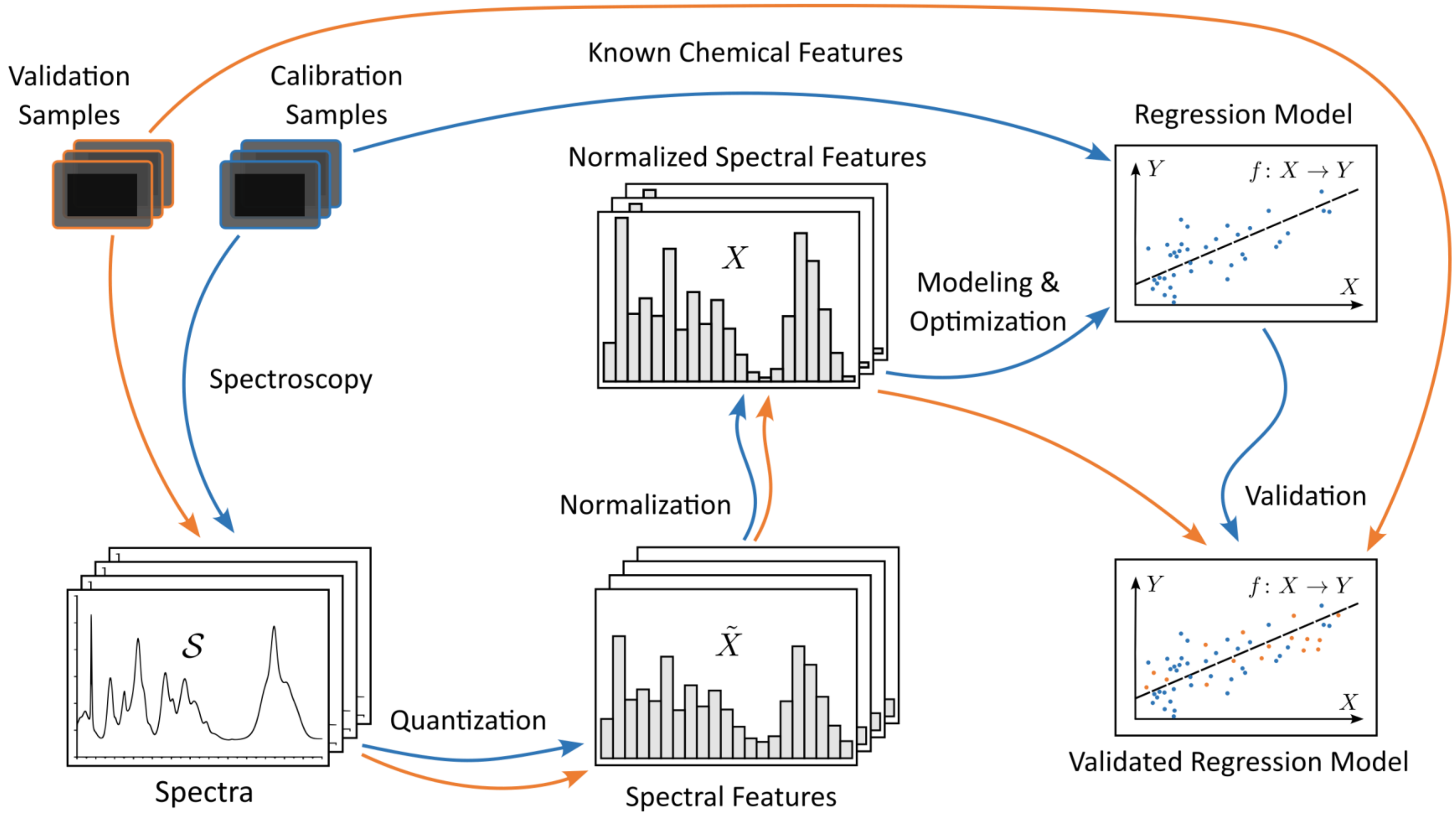


Bayesian sample selection

Ir
ca

Determine confidence within validation samples

using quantile loss

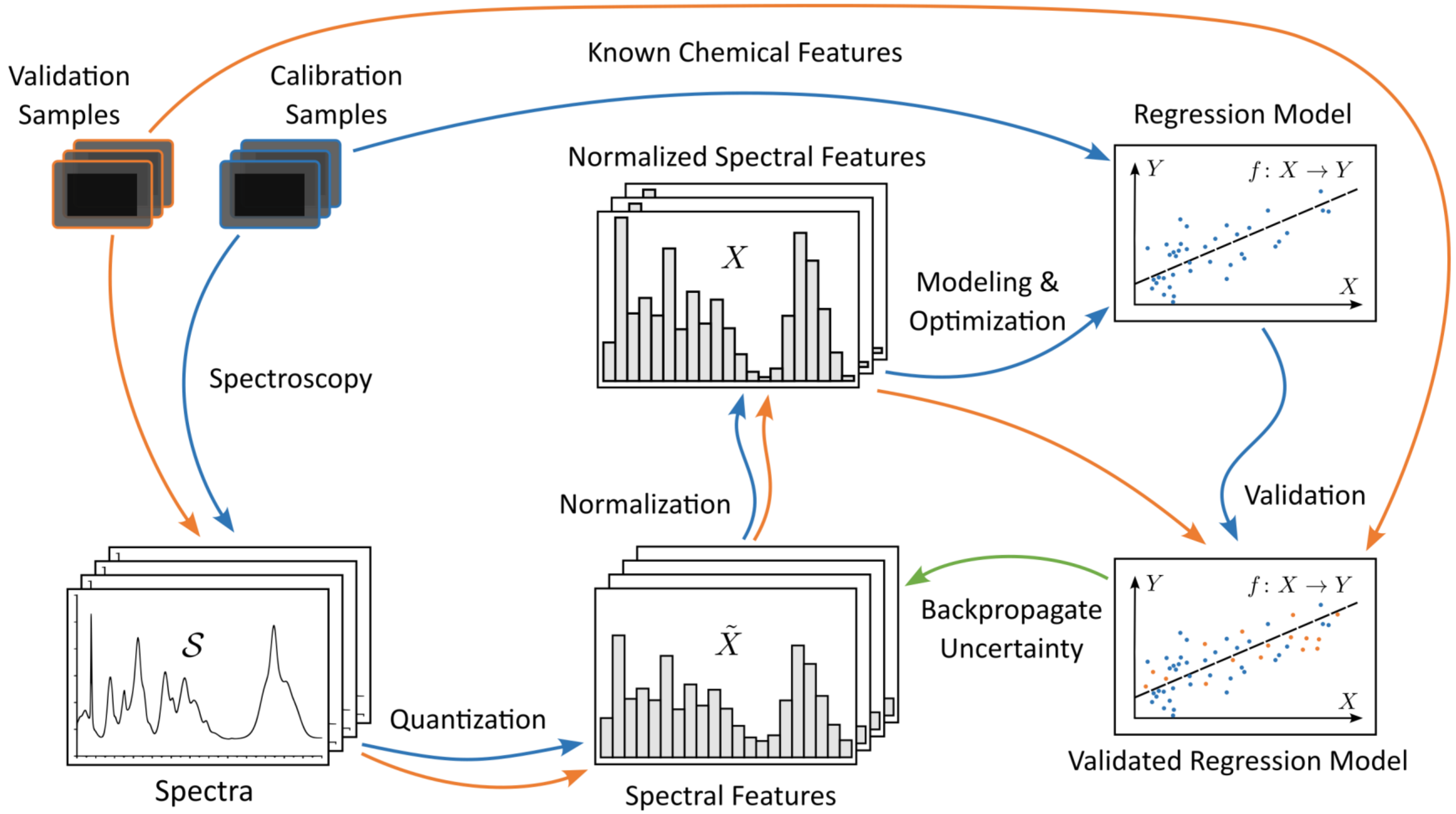


Bayesian sample selection

Ir
ca
e

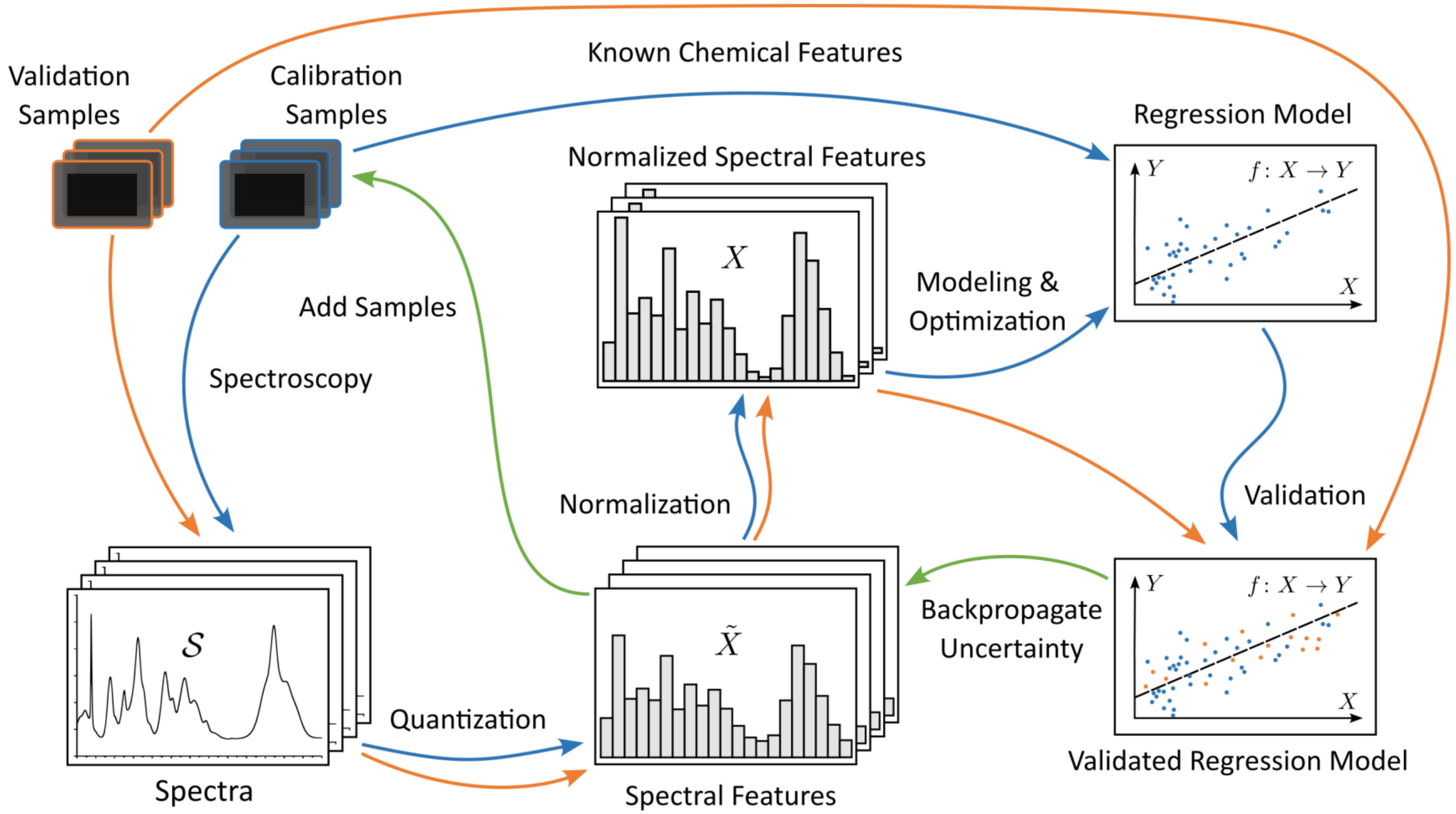
Backpropagate uncertainty to spectral features

using quantile loss and non-linear feature selection



Bayesian sample selection

Add calibration samples with distinct features



Conclusions about **Bayesian** sample selection

Iteratively increases Confidence in Validation Predictions

by Bayesian Optimization

Iteratively increases Confidence in Validation Predictions

by Bayesian Optimization

Can lead to Overfitting of Validation samples

Bayesian sample selection

Iteratively increases Confidence in Validation Predictions

by Bayesian Optimization

Can lead to Overfitting of Validation samples

How to avoid Overfitting?

Additional Test Samples?
Changes in Validation Samples?

Iteratively increases Confidence in Validation Predictions

by Bayesian Optimization

Can lead to Overfitting of Validation samples

Intricates scheduling in product development

How to avoid Overfitting?

Additional Test Samples?
Changes in Validation Samples?

Iteratively increases Confidence in Validation Predictions

by Bayesian Optimization

Can lead to Overfitting of Validation samples

How to avoid Overfitting?

Additional Test Samples?
Changes in Validation Samples?

Intricates scheduling in product development

When to stop Iterations?

Elbow method?
Fixed number of iterations?

Summary

Summary

#1

Neural networks are by themselves **not suitable** for multivariate calibration in NIR due to missing generalizability

Summary

#1

Neural networks are by themselves **not suitable** for multivariate calibration in NIR due to missing generalizability

#2

A preceded embedding step with deep **non-linear decorrelation** using energy based modeling fixes this issue

Summary

#1

Neural networks are by themselves **not suitable** for multivariate calibration in NIR due to missing generalizability

#2

A preceded embedding step with deep **non-linear decorrelation** using energy based modeling fixes this issue

#3

The incorporation of non-linearity increases the **accuracy** and the **efficiency** of predictions

Summary

#1

Neural networks are by themselves **not suitable** for multivariate calibration in NIR due to missing generalizability

#2

A preceded embedding step with deep **non-linear decorrelation** using energy based modeling fixes this issue

#3

The incorporation of non-linearity increases the **accuracy** and the **efficiency** of predictions

#4

The incorporation of non-linearity intricates **feature selection** and the **design of experiments**

Thank you for your attention!