LIÈGE université
Gembloux
Agro-Bio Tech

Unraveling the Fascination of Advancing Complexity in Spectrometry Models

Prof. Hélène Soyeurt

4th edition of the Spectroscopy and Chemometrics Workshop – 29th April 2024

LIÈGE université
Gembloux
Agro-Bio Tech

Study-case :

Milk mid-infrared spectrometry

1883-1887: First dairy companies in Belgium

1903: First international congress of dairy industry in Belgium ➔ Common approach

1903: Creation of International Dairy Federation (IDF)
Starting research to standardize and to discover new analytical, simple, reliable and reproducible methods

1953: Belgian law : label A and AA

1957: Milk recording on individual cows

1964: Routine milk analysis for quality and composition for the milk payment

Industry and farmers

1897: First national congress of dairy industry
Need for solid content quantification to fix milk price

60's: IRMA : InfraRed Milk Analyzer : 3 MIR Filters %Fat,%PROT,%Lactose + water absorption

1970: First MilkoScan : one cell measurement

Milk spectroscopic analysis

1927:
First spectroscopic analysis
Filtered UV analysis on milk

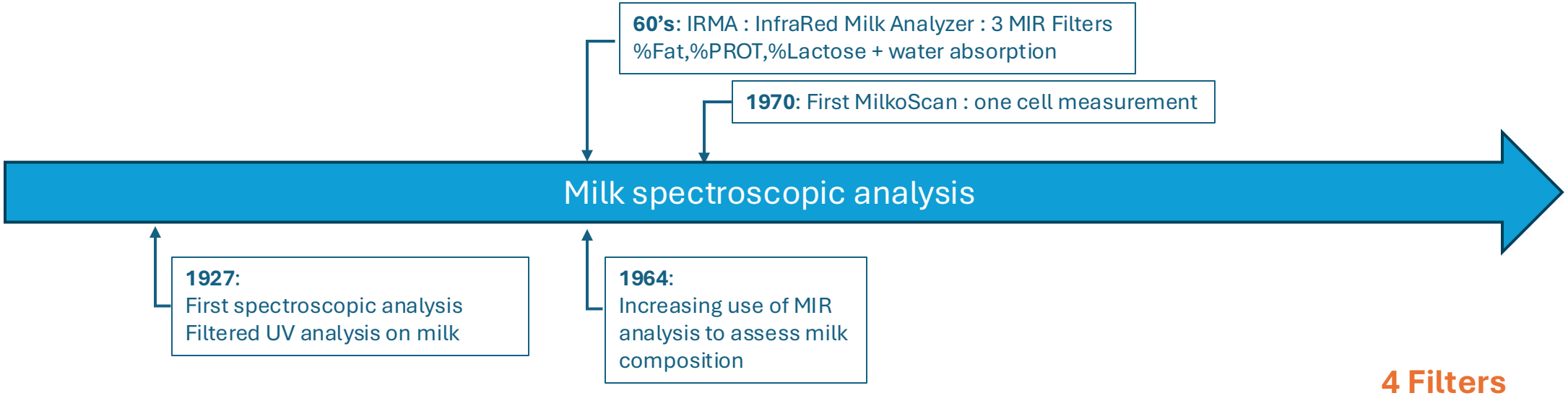
1964:
Increasing use of MIR analysis to assess milk composition

1990: FT-MIR : Anadys

Algorithms based on milk analysis

1960 : use of multivariate linear regression

1991: First PLS model for milk analysis.


60's: IRMA : InfraRed Milk Analyzer : 3 MIR Filters
%Fat,%PROT,%Lactose + water absorption
1970: First MilkoScan : one cell measurement
Milk spectroscopic analysis
1927:
First spectroscopic analysis
Filtered UV analysis on milk
1964:
Increasing use of MIR analysis to assess milk composition
4 Filters

Prior to making a comparison of the two calibration approaches, it will be useful to review the basic principles of infrared milk analysis. A much more detailed discussion of basic principles of infrared analysis is available elsewhere (3). In general, four filters are used in most infrared analyzers. The following predominant wavelengths are commonly used: 5.723 µm for fat (C=O at the ester linkage), 6.465 µm for protein (N-H bond with the peptide bond), 9.610 µm for lactose (C-OH bond), and 3.48 µm for fat (CH_2 groups). These filters are referred to as fat A, protein, lactose, and fat B, respectively. A separate reference filter (i.e., wavelength) is used in combination with each of these wavelengths in most instruments. Some new instrument models (Multispec Dairy lab) use no reference filter (it uses water reading at the same wavelength) or only uses one separate reference filter wavelength for all channels.

4 Filters

1,747cm^{-1} for fat (C=0 ester linkage) ➔ Fat A
1,547 cm^{-1} for protein
1,041 cm^{-1} for lactose
2,873 cm^{-1} for fat (CH2 groups) ➔ Fat B

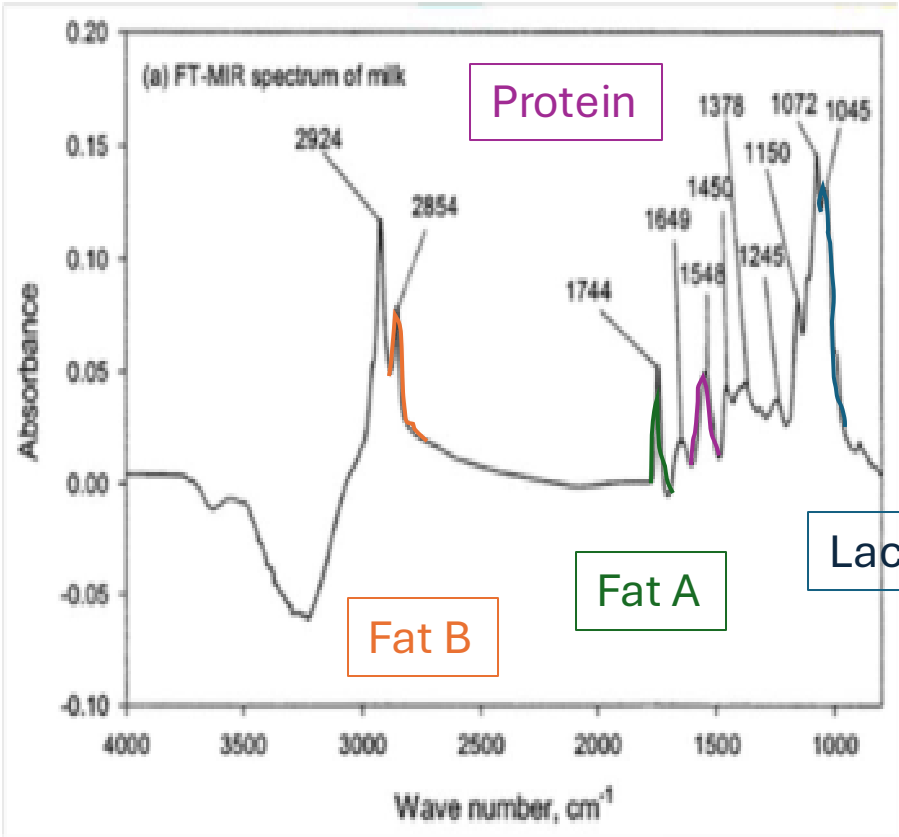
SYMPOSIUM: INSTRUMENTAL METHODS FOR MEASURING COMPONENTS OF MILK

Infrared Milk Analysis – Challenges for the Future[1]

Received August 31, 1988.
Accepted November 14, 1988.
[1]Mention of any equipment supplier is for scientific accuracy only and does not indicate any product endorsement by the authors or Cornell University.

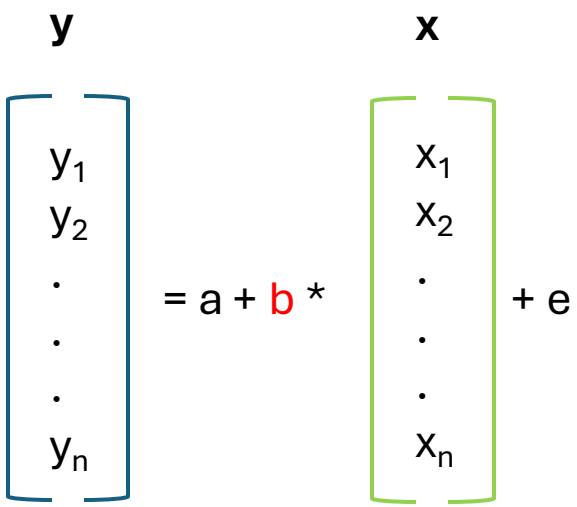
D. M. BARBANO and J. L. CLARK
Department of Food Science
Cornell University
Ithaca, NY 14853

1989 J Dairy Sci 72:1627–1636



Linearity between IR absorption and trait

Univariate linear regression

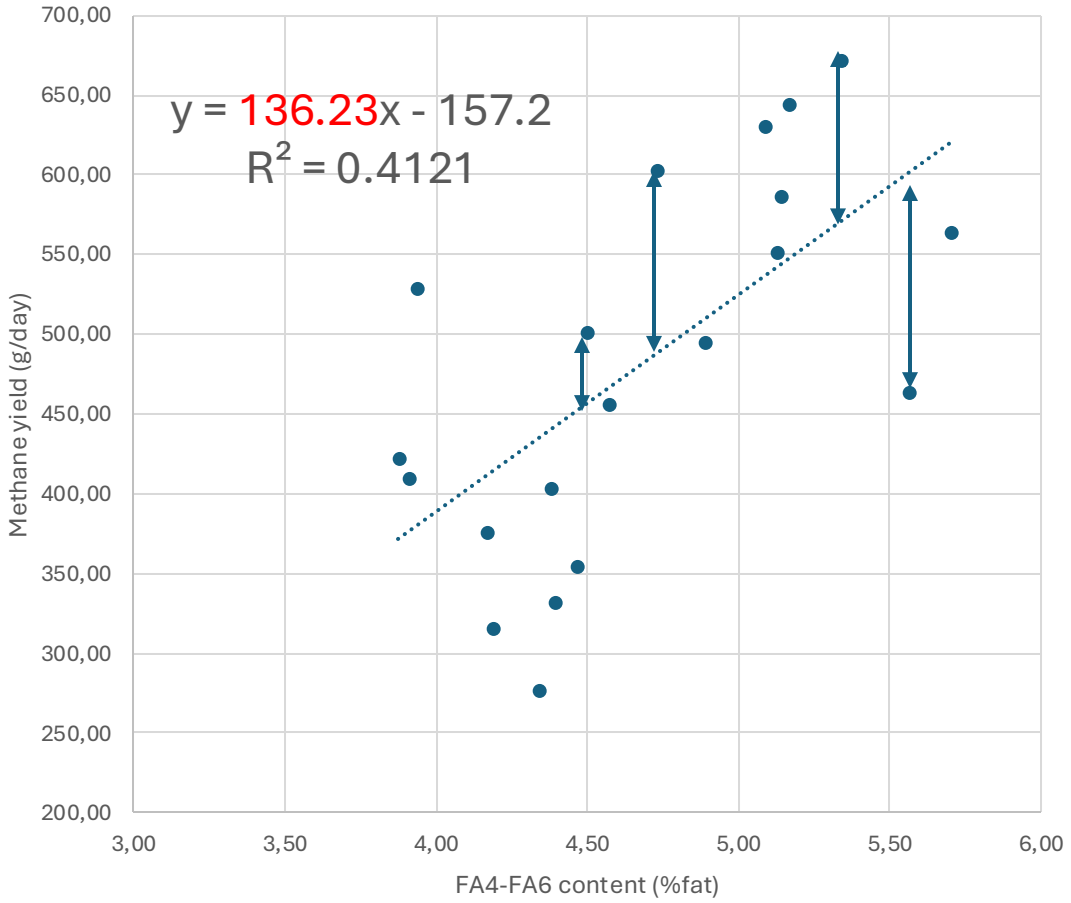


y $\rightarrow \bar{y}, \sigma_y$ x $\rightarrow \bar{x}, \sigma_x$

Filters :

- Protein
- Lactose

Univariate linear regression

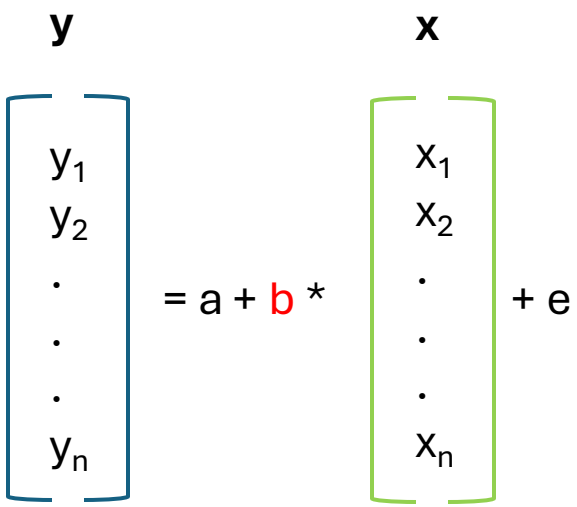


$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = a + b * \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix} + e$$

Minimize the distance between the records and the predictions

Linearity between IR absorption and trait

Univariate linear regression



y $\rightarrow \bar{y}, \sigma_y$ x $\rightarrow \bar{x}, \sigma_x$

Filters :

- Protein
- Lactose

→ %fat ?

Use the information coming from 2 filters

Multivariate linear regression

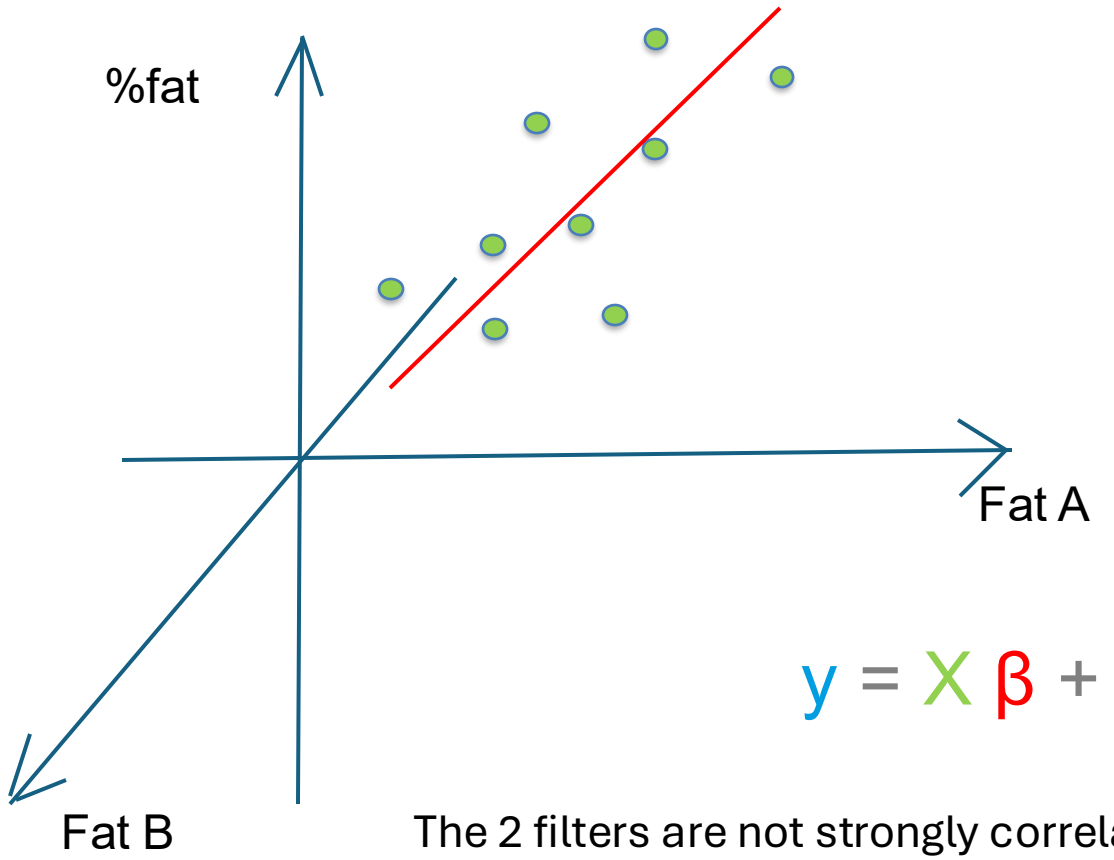
$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = b_1 \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix} + b_2 \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix} + e$$

%fat FAT A Fat B

Variable to be explained

Explaining variables

Regression coefficients



$$y = X \beta + e$$

The 2 filters are not strongly correlated.

Regression

Estimation of regression coefficients

$$y = X\beta + e$$

Ordinary least squares (OLS) estimation

$$\min (e^2 = (y - X\beta)^2)$$

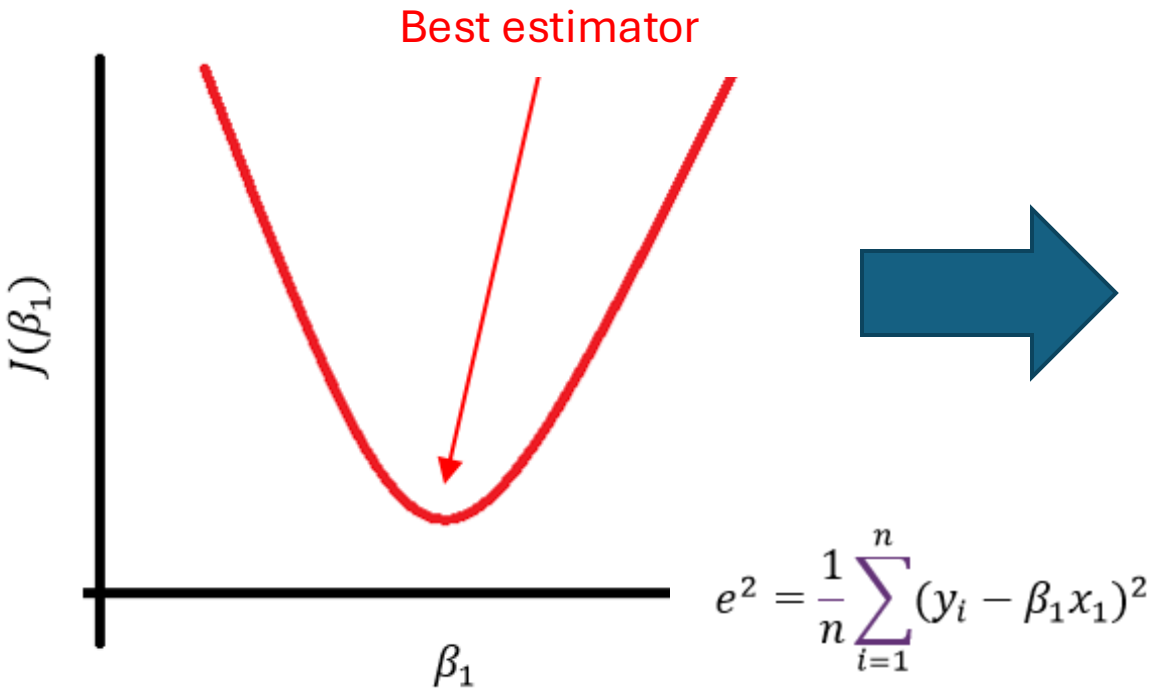
$$\rightarrow \frac{\partial (y - X\beta)^2}{\partial (\beta)} = 0 \rightarrow \beta = (X'X)^{-1}X'y$$

The inversion can be sometimes problematic when features are highly correlated.

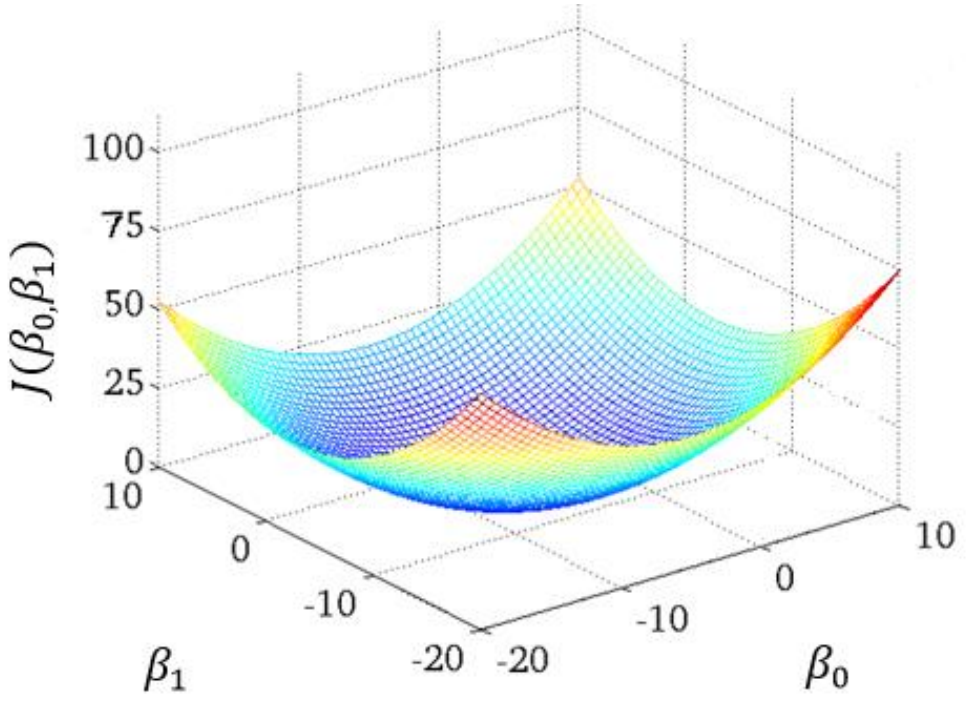
Regression

Estimation of regression coefficients by iteration

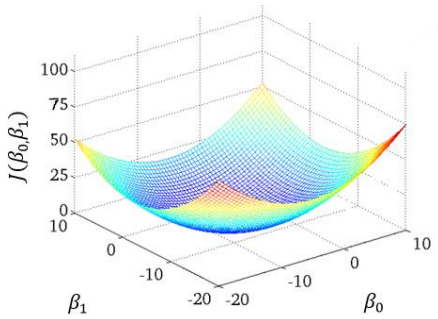
min $((y - X\beta)^2)$ ➔ $(y - X\beta)^2$ is the cost function $(J(\beta))$



$$e^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_1 x_1)^2$$



Regression



Estimation of regression coefficients by iteration

min $((y - X\beta)^2)$ ➔ $(y - X\beta)^2$ is the cost function $(J(\beta))$

β coefficients are estimated from partial derivatives

$$\frac{\partial}{\partial \beta_0} J(\beta_0, \beta_1)$$

$$\frac{\partial}{\partial \beta_1} J(\beta_0, \beta_1)$$

Iteration

$$temp0 = \beta_0 - \alpha \frac{\partial}{\partial \beta_0} J(\beta_0, \beta_1)$$

$$temp1 = \beta_1 - \alpha \frac{\partial}{\partial \beta_1} J(\beta_0, \beta_1)$$

$$\beta_0 = temp0$$

$$\beta_1 = temp1$$

Regression

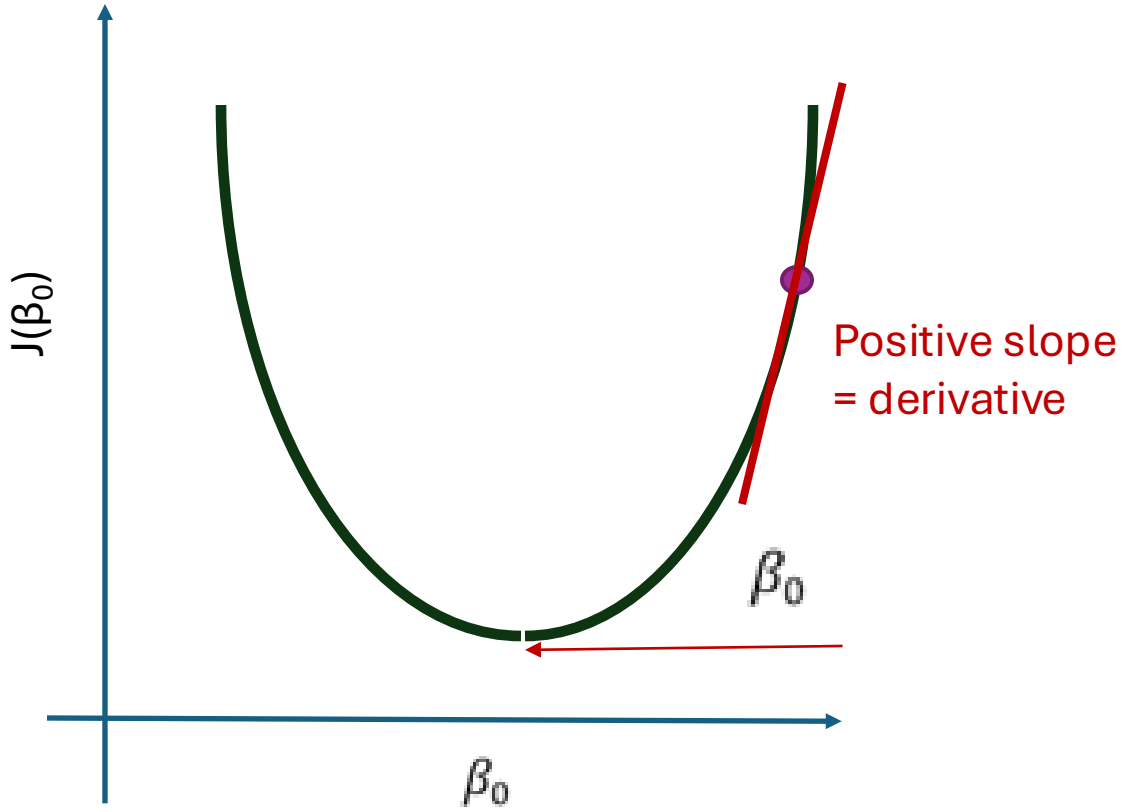
Estimation of regression coefficients by iteration

From univariate regression

$$temp0 = \beta_0 - \alpha \frac{\partial}{\partial \beta_0} J(\beta_0, \beta_1)$$

$$\beta_0 = temp0$$

Farest we are from the best estimator, bigger will be the derivative value (=slope) !!



Regression

Estimation of regression coefficients by iteration

From univariate regression

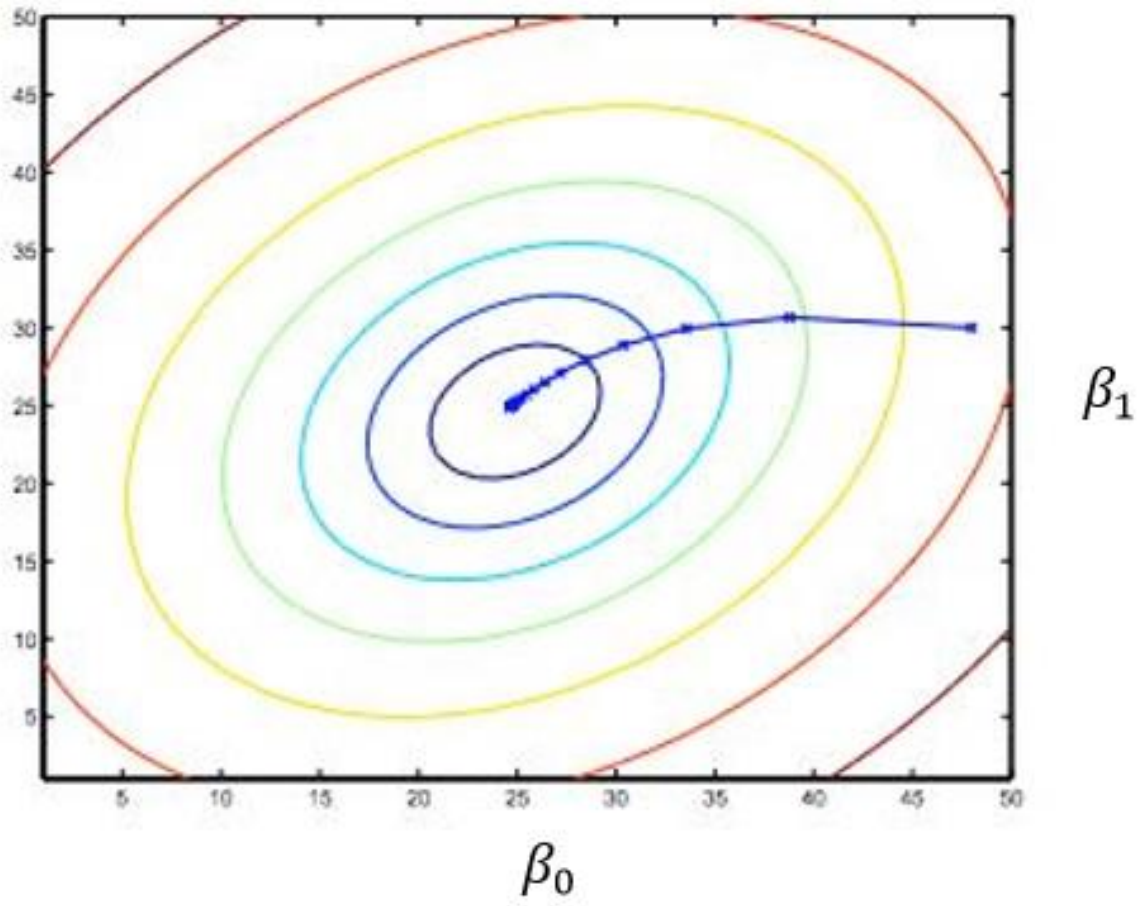
$$temp0 = \beta_0 - \alpha \frac{\partial}{\partial \beta_0} J(\beta_0, \beta_1)$$

$$temp1 = \beta_1 - \alpha \frac{\partial}{\partial \beta_1} J(\beta_0, \beta_1)$$

$$\beta_0 = temp0$$

$$\beta_1 = temp1$$

Gradient descent



Regression

Estimation of regression coefficients by iteration

From univariate regression

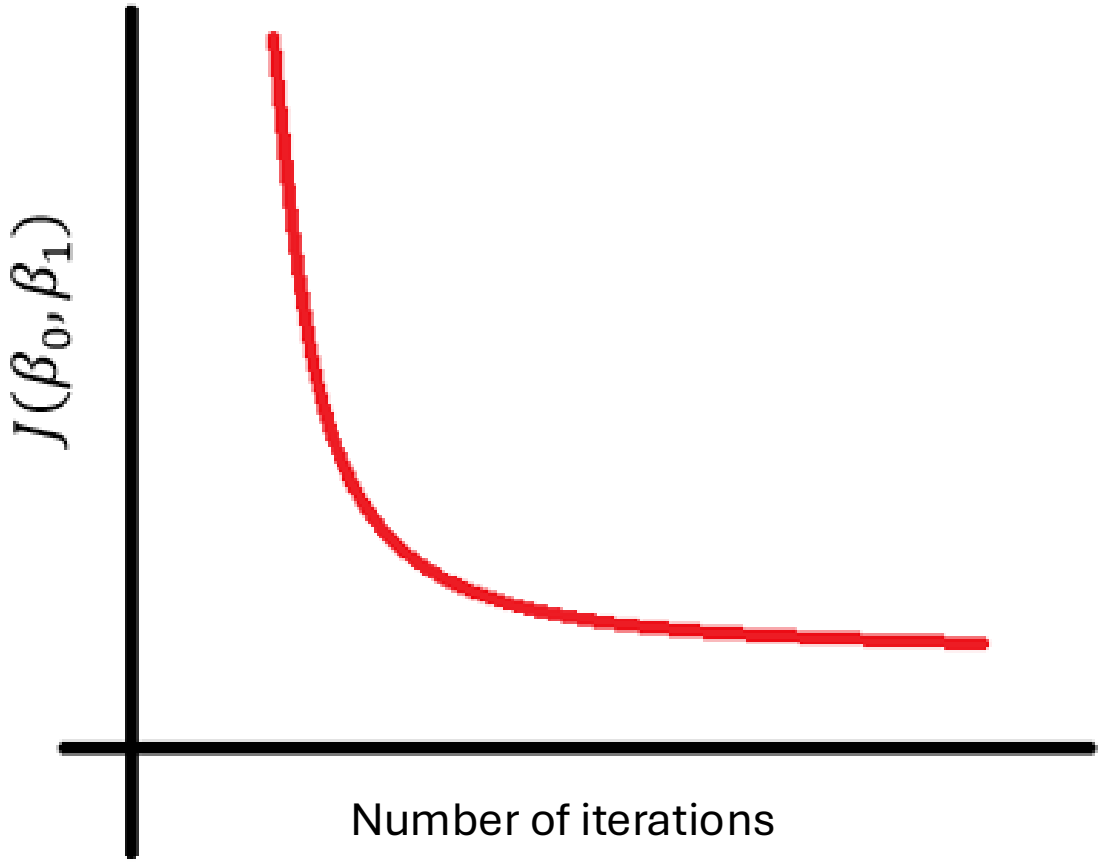
$$temp0 = \beta_0 - \alpha \frac{\partial}{\partial \beta_0} J(\beta_0, \beta_1)$$

$$temp1 = \beta_1 - \alpha \frac{\partial}{\partial \beta_1} J(\beta_0, \beta_1)$$

$$\beta_0 = temp0$$

$$\beta_1 = temp1$$

Gradient descent



When have we started to machine learning / artificial intelligence ?

Tom Mitchell introduces the concept of « Learning by experience »

« A computer program is said to learn from experience E with respect to some task (T) and some performance measure P, if its performance on T, as measured by P, improves with experience E. »

➔ **When we have started to estimate regression coefficients using interative procedure.**

Model complexity

- Structure
  - Number of features or combined features
- Number of parameters
  - Optimized parameters

The complexity of models changes

- Structure
- Number of parameters

Models already tested in milk spectroscopic analysis

Multivariate linear regression

60's: IRMA : InfraRed Milk Analyzer : 3 MIR Filters %Fat,%PROT,%Lactose + water absorption

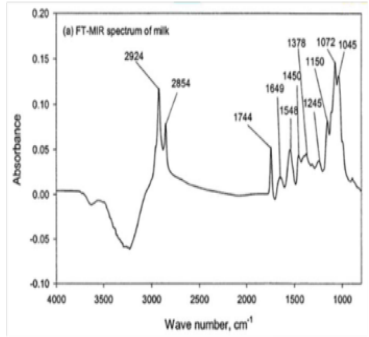
1970: First MilkoScan : one cell measurement

Milk spectroscopic analysis

1927:
First spectroscopic analysis
Filtered UV analysis on milk

1964:
Increasing use of MIR analysis to assess milk composition

1990: FT-MIR : Anadys

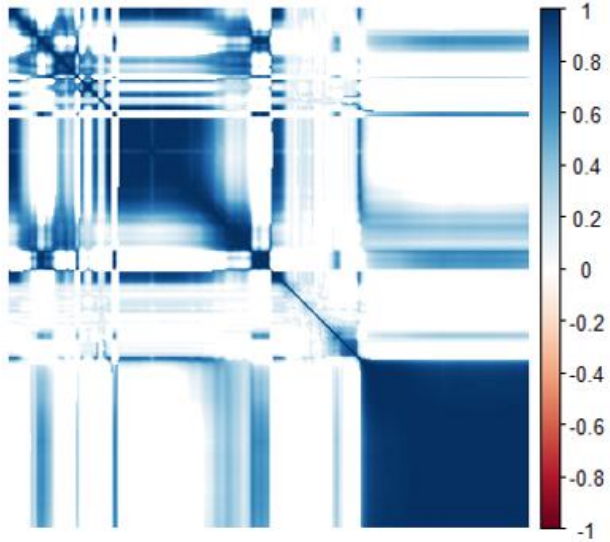


Higher dimensionality

Multivaried linear regression

$$Y = \mu + b_1 x_1 + b_2 x_2 + e$$

- Low computational resources
- Need a number of samples higher than the number of features in the regression
- Features cannot be highly correlated



Different possibilities

(1) Make a **selection** of features

(2) Force to **share** the informations between correlated features

(3) Create **new less correlated features** based on linear combination

Different possibilities

Penalized regression

(1) Make a **selection** of features

(2) Force to **share** the informations between correlated features

(3) Create **new less correlated features** based on linear combination

Penalized Regression

Ridge

- Solve the colinearity problem by reducing the value of b coefficients

Regularization parameter

$$\min \left[\sum_{i=1}^{n} (y_i - \beta_1 x_1 - \cdots - \beta_q x_q)^2 + \lambda \sum_{j=1}^{q} \beta_j^2 \right]$$

OLS

Penalty

If $= 0$ ➔ simple linear regression

If $\uparrow$, penality is growing

Penalized Regression

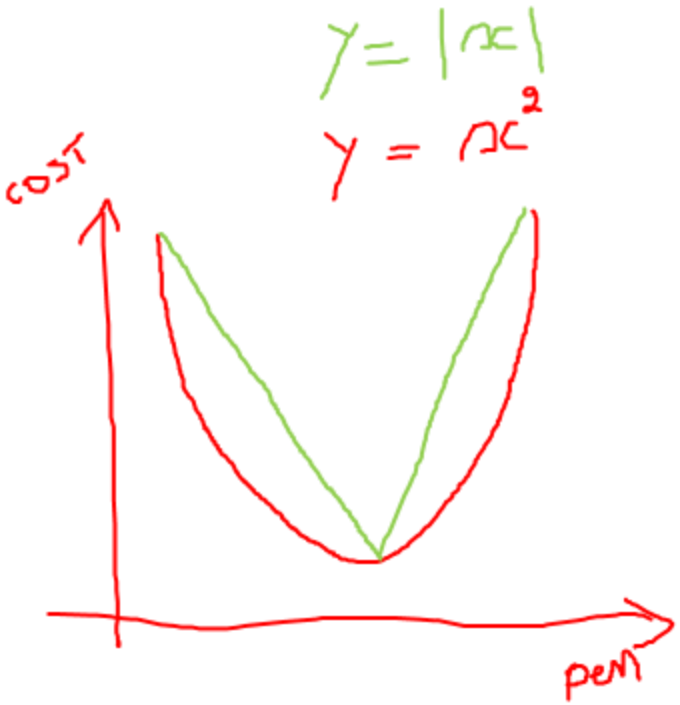
Lasso

- Lasso (*Least Absolute Shrinkage and Selection Operator*)

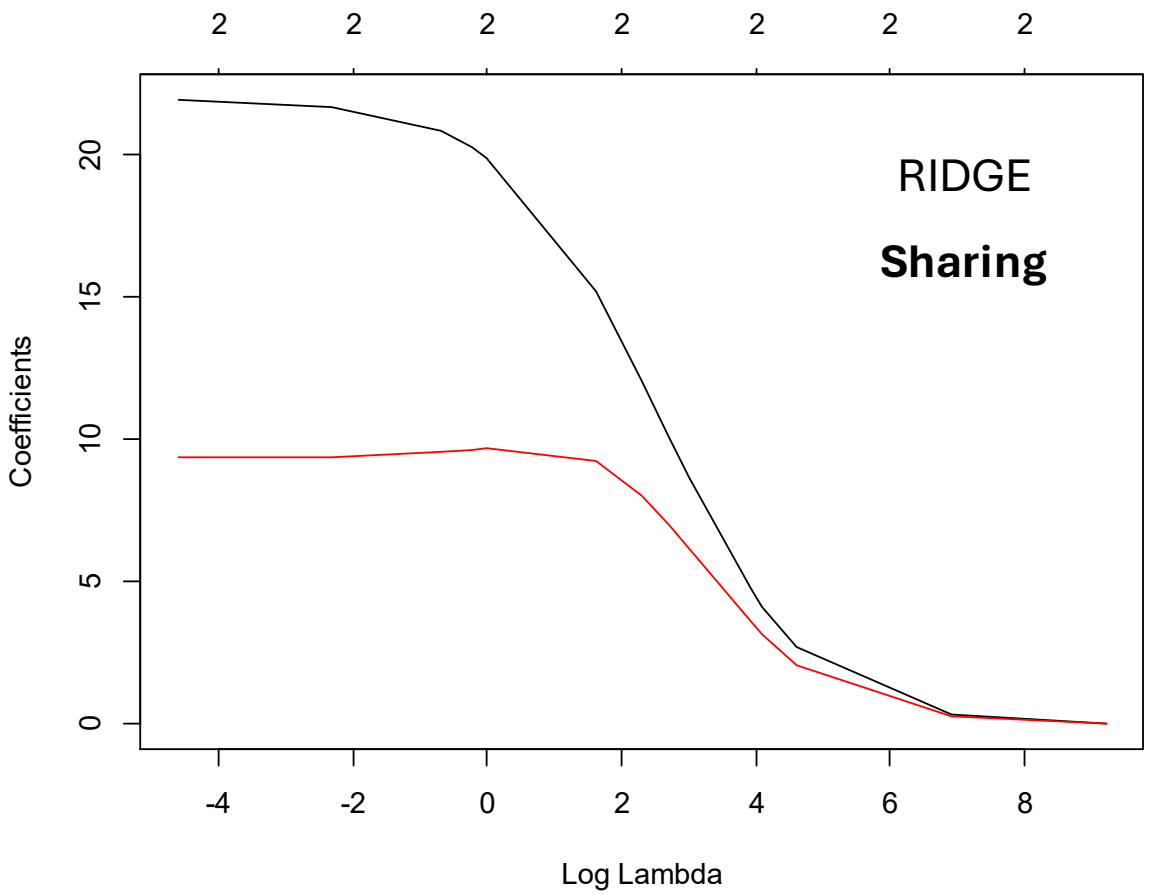
$$\min \left[\sum_{i=1}^{n} (y_i - \beta_1 x_1 - \cdots - \beta_q x_q)^2 + \lambda \sum_{j=1}^{q} |\beta_j| \right]$$

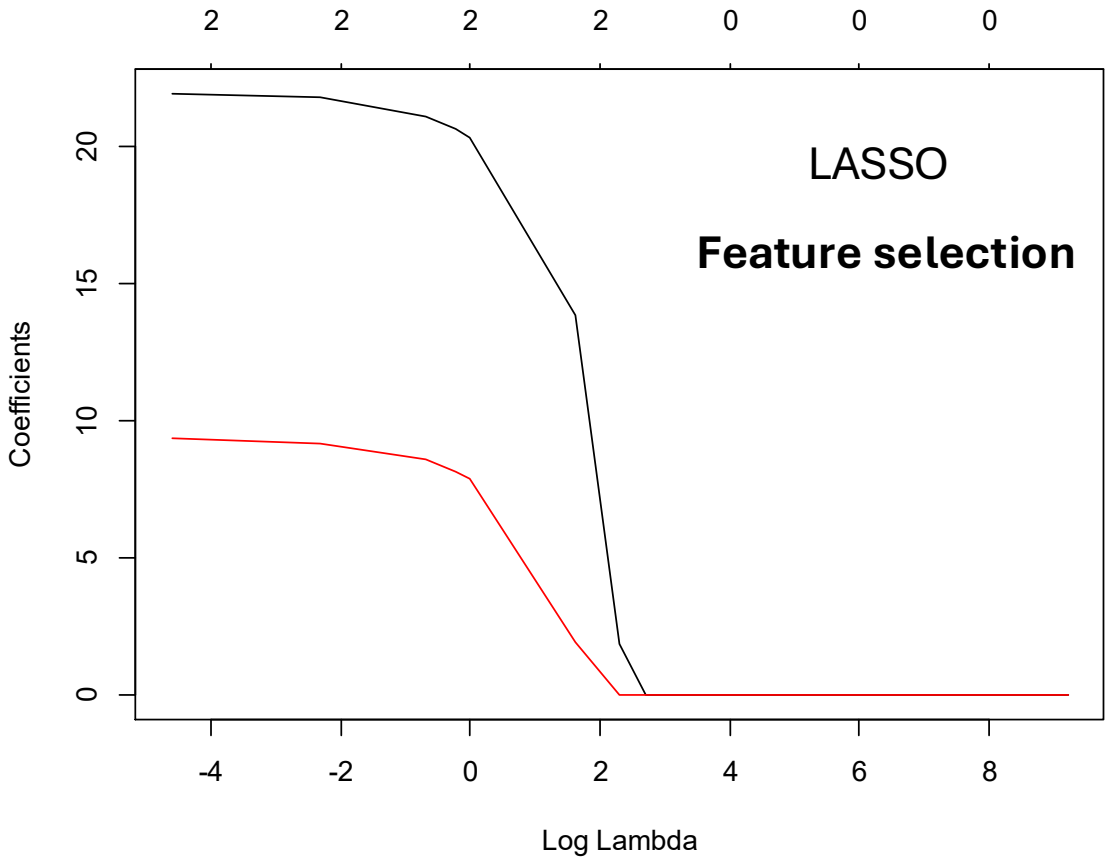
OLS

Penalty



Ridge vs. Lasso regressions





Penalized Regression

ElasticNet

- Combination between Ridge and Lasso

$$\sum_{i=1}^{n}(y_i - \beta_1 x_1 - \cdots - \beta_q x_q)^2 + \lambda_2 \sum_{j=1}^{q} \beta_j^2 + \lambda_1 \sum_{j=1}^{q} |\beta_j|$$

OLS

Penalty

But, the complexity of models changes

- Structure
- Number of parameters

Models already tested in milk spectroscopic analysis

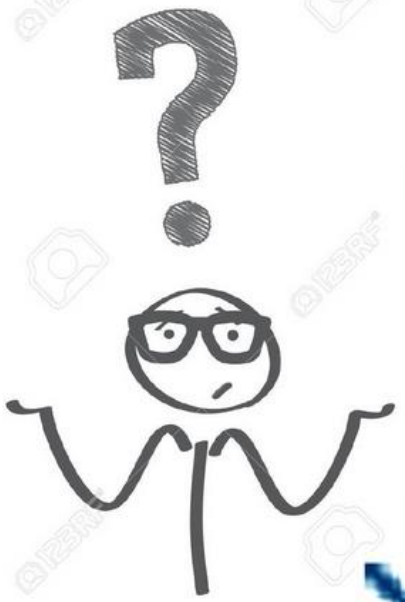
Multivaried linear regression	Penalized linear Regression (Ridge, Lasso, Elastic)

Different possibilities

(1) Make a **selection** of features

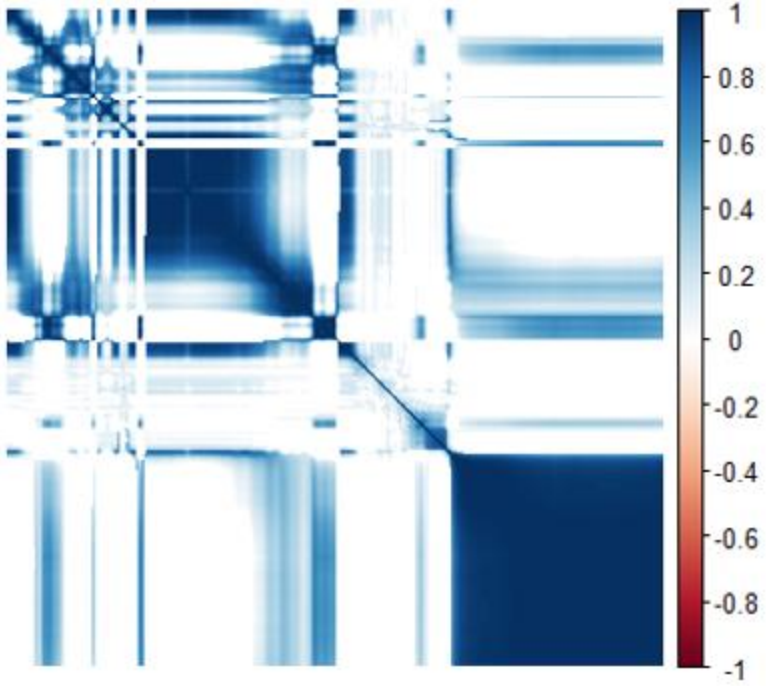
(2) Force to **share** the informations between correlated features

(3) Create **new less correlated features** based on linear combination

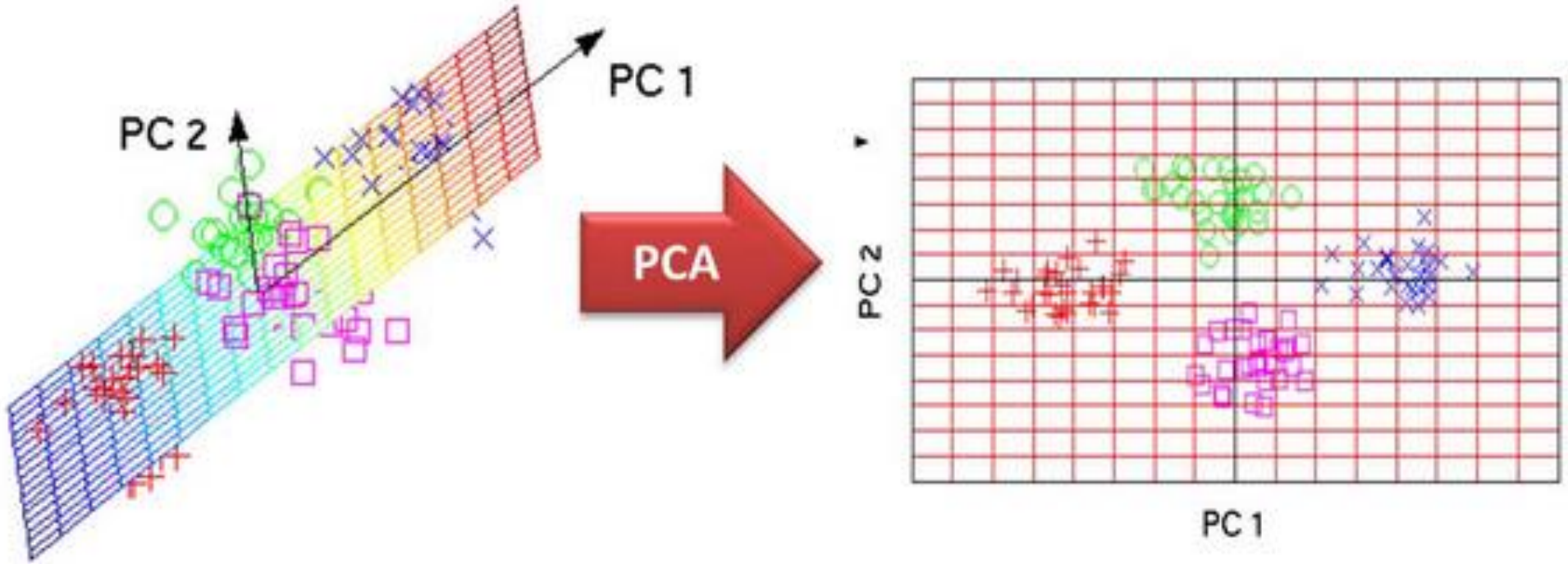


Nvariable in the regression is limited to the number of records – the number of estimated coefficients

Spectral variables (X variables) are correlated between them → colinearity



What can we do?



Principal Components Analysis

- Reducing the dimensionality of the X matrix
- Principal components are independent
  - Correlations between them = 0

Principal Components Analysis

Records (n x p)
- n = number of observations
- p = number of variables (e.g., 1,060)

Standardization of each j explaining variable

$$x'_{ij} = \frac{(x_{ij} - \bar{x}_j)}{\hat{\sigma}_j}$$

(1) $z_{ji} = u_{1j}x_{i1} + u_{2j}x_{i2} + : : : + u_{pj}x_{ip}$
(2) $u^2_{1j} + u^2_{2j} + : : : + u^2_{pj} = 1$
(3) $u_{1j}u_{1k} + u_{2j}u_{2k} + : : : + u_{pj}u_{pk} = 0; \forall j \neq k$
(4) Maximum variance for z_j

Principal Components Regression (PCR)

$$y = X\,\beta + e$$

$$y = X_{PCA}\,\beta + e$$

PCA allows to maximize the X variability and reduce the number of considered x variables if they are correlated between them.

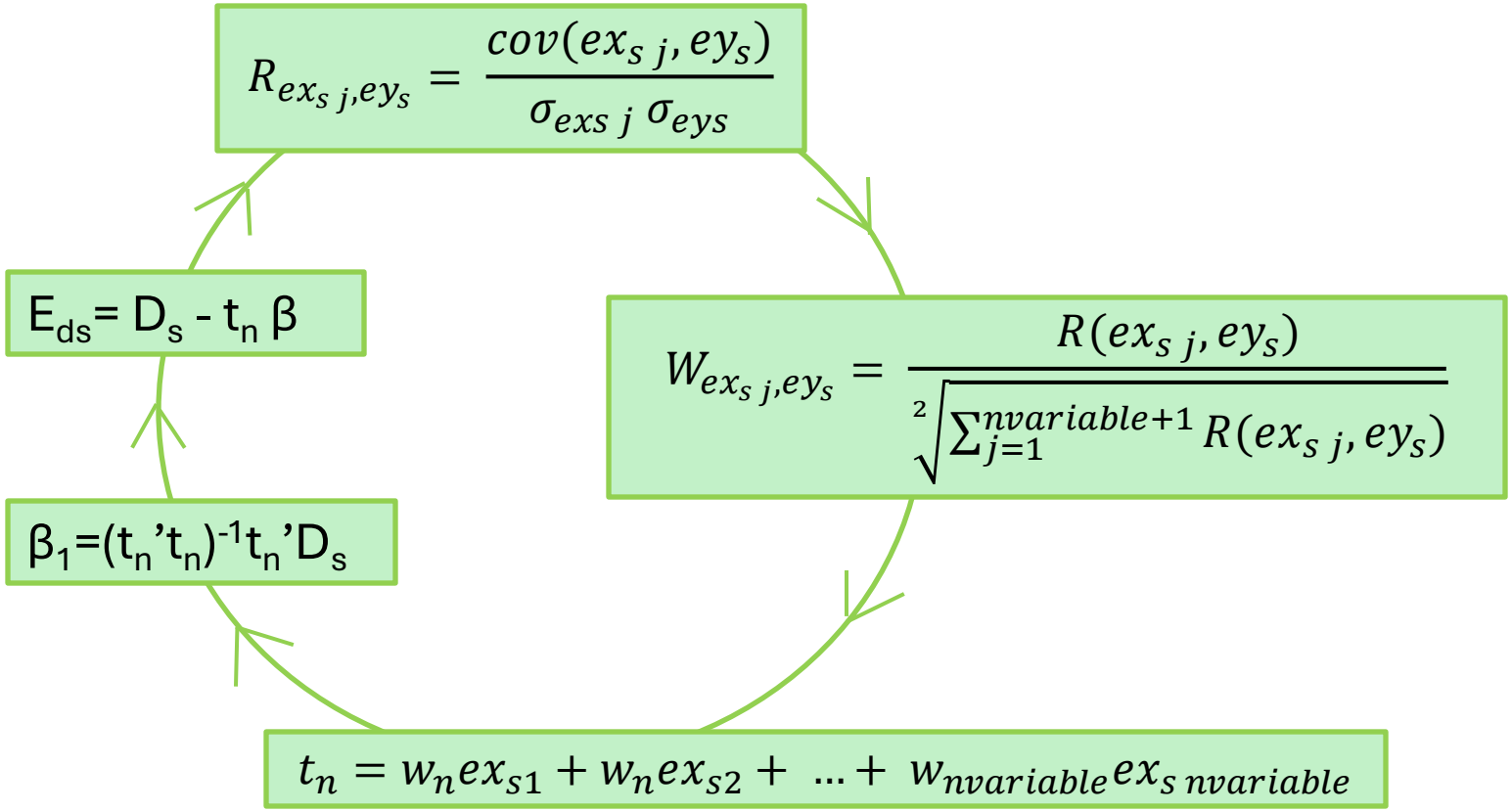
Similar to multivariate linear regression but now the features are not correlated.

Partial Least Squares Regression (PLS)

- Advantages of multivariate regression without inconvenients
  - No colinearity problem
  - No problem to model with a large dataset
  - No problem to interpret

- Take into account simultaneously **X and y variabilities**

PLS method – NIPALS algorithm



But, the complexity of models changes

- Structure
- Number of parameters

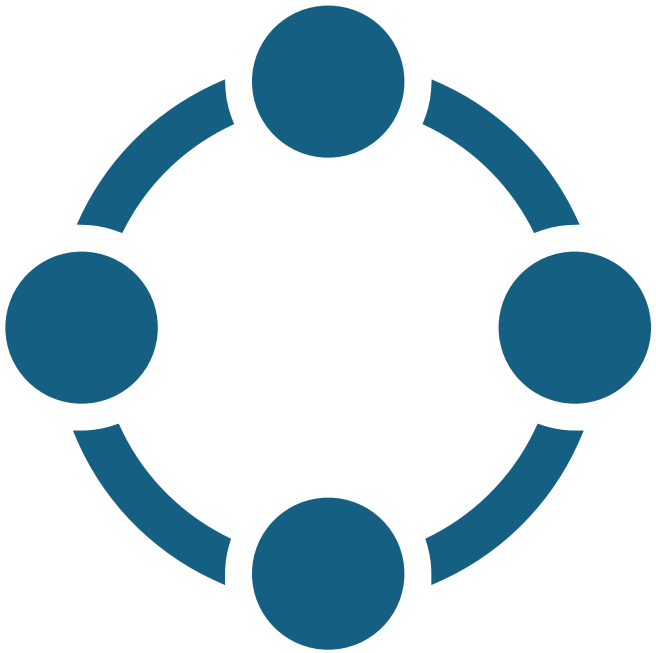
Models already tested in milk spectroscopic analysis

Multivaried linear regression

Penalized linear Regression (Ridge, Lasso, Elastic)

Principal Component regression

Partial Least Squares Regression



Why other kinds of model?

Difficulty to consider no linear relationships

- Use polynomials

A discriminant modelling method

SVM

Support Vector Machines

V.N. Vapnik (The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995)

- Discrimination
  - two classes
  - multiclass
- Regression

A discriminant modelling method

SVM

Notations

⇒ Consider a learning data set $\{(x^1,y^1),\ldots,(x^m,y^m)\}$ where y_i is the class label of vector x_i

⇒ Spectra lie in X whereas labels lie in $y = \{-1,1\}$

Aim

⇒ build a decision function $f(x)$ able to assign a label from the observation of a certain number of input-output pairs $\{(x^i,y^i)\ \ i=1,\ldots m\}$

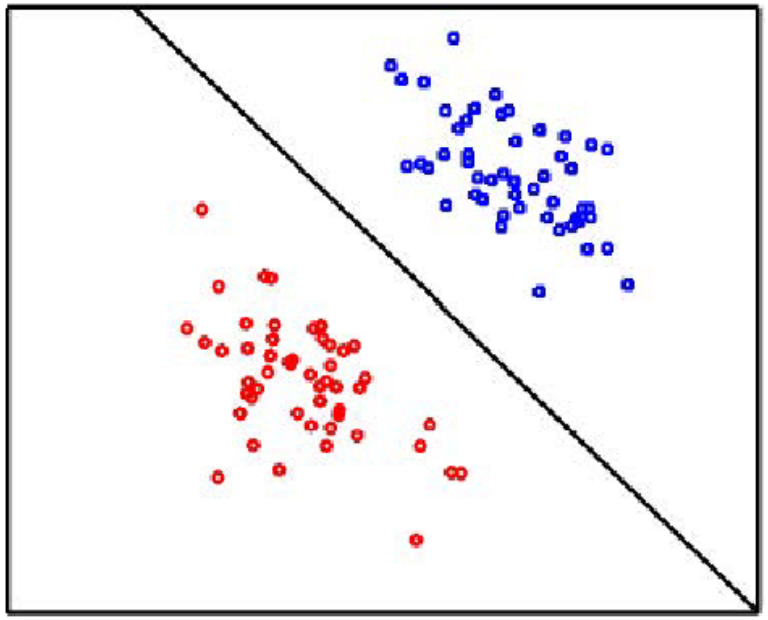
$$f(x) : X \longrightarrow \{+1,-1\}$$

A discriminant modelling method

SVM

Hyperplane classifier

⇒ We assume here that a linear classifier can be implemented



Quantification ➔ distance from the hyperplane

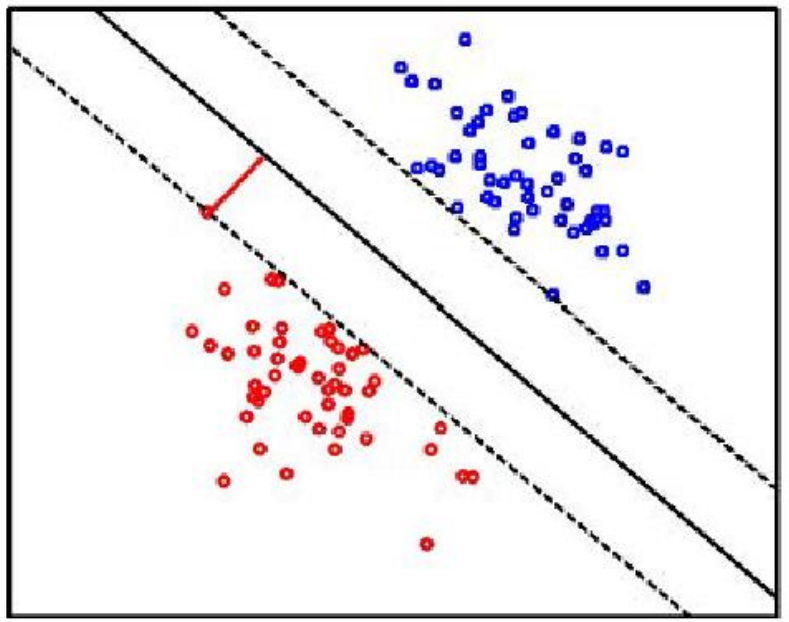
(Source : Manuel Davy, CNRS (Equipe ADTS)
October 11, 2003, Nantes, France)

A discriminant modelling method

SVM

Suggestion

⇒ The one that maximises the margin within the learning data set



C optimizer.

Classifier that maximises the margin → Optimal sparating hyperplane
→ better generalization is expected

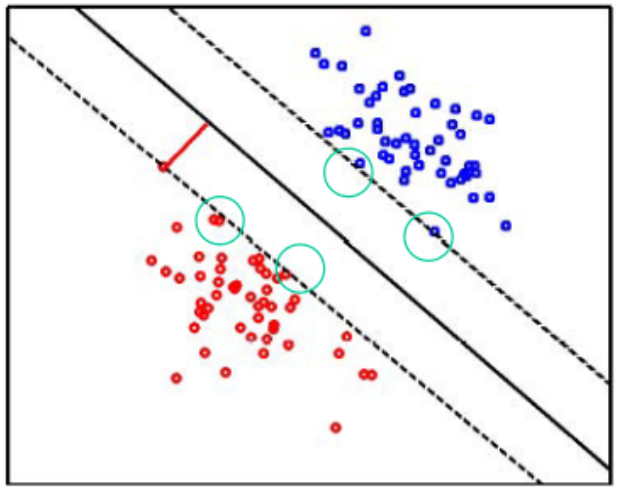
A discriminant modelling method

SVM

$\Rightarrow$ The decision function is

$$f(x) = sign\left(\sum_{i=1}^{m} y_i \alpha_i < x, x_i > + b\right)$$

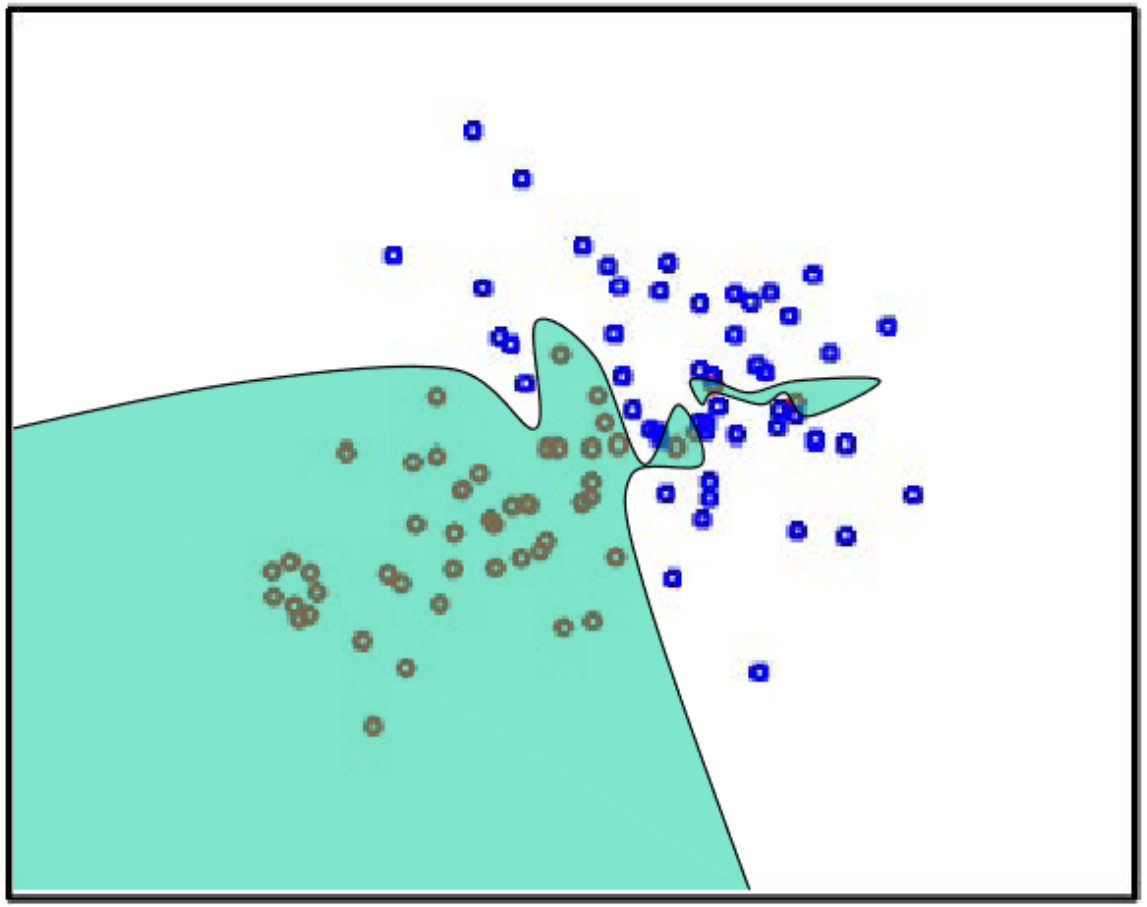
$\Rightarrow$ α_i 's are nonzero only for a small number of learning vectors, samples at the margin called Support Vectors



$\Rightarrow$ Then, one only considers support vectors (sparse solution)

$$f(x) = sign\left(\sum_{i=1}^{SV} y_i \alpha_i < x, x_i^{sv} > + b\right)$$

Or a nonlinear classifier?



But, the complexity of models changes

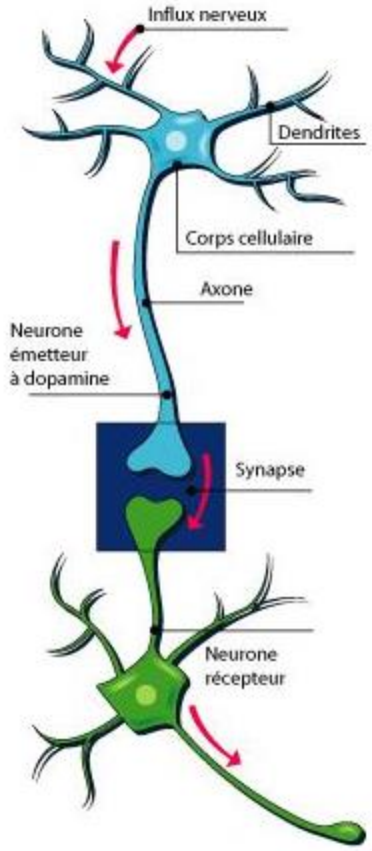
- Structure
- Number of parameters

Models already tested in milk spectroscopic analysis

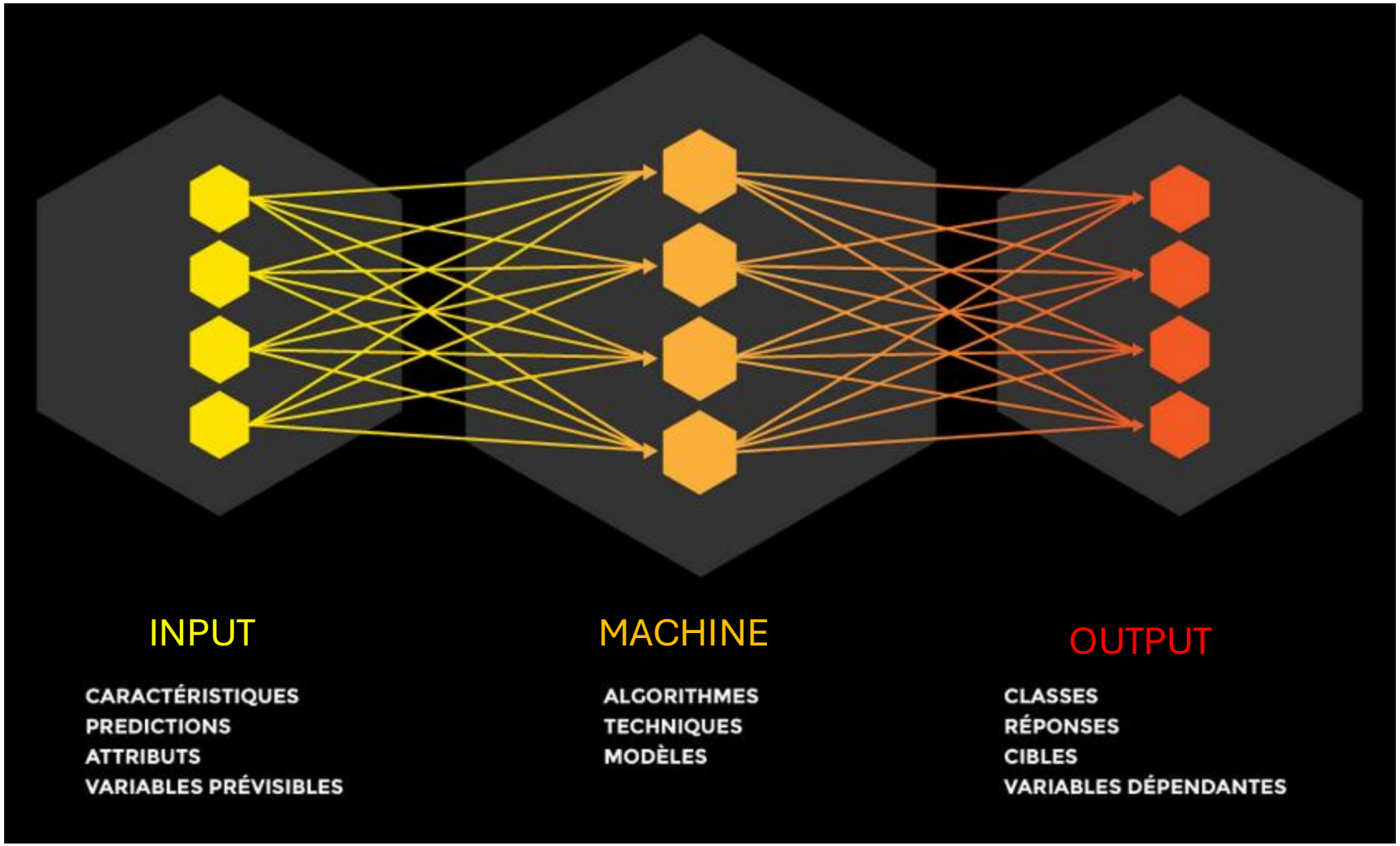
- Multivaried linear regression
- Penalized linear Regression (Ridge, Lasso, Elastic)
- Principal Component regression
- Partial Least Squares Regression
- Support vector machine

Neural network

- Artificial neural network (ANN)
- Basic of deep learning
- McCulloch-Pitts (1943) represents the neuron as a binary tool
- Great interest in the 80’s and 90’s
- Mimic the structure of brain composed of neurons and synapses

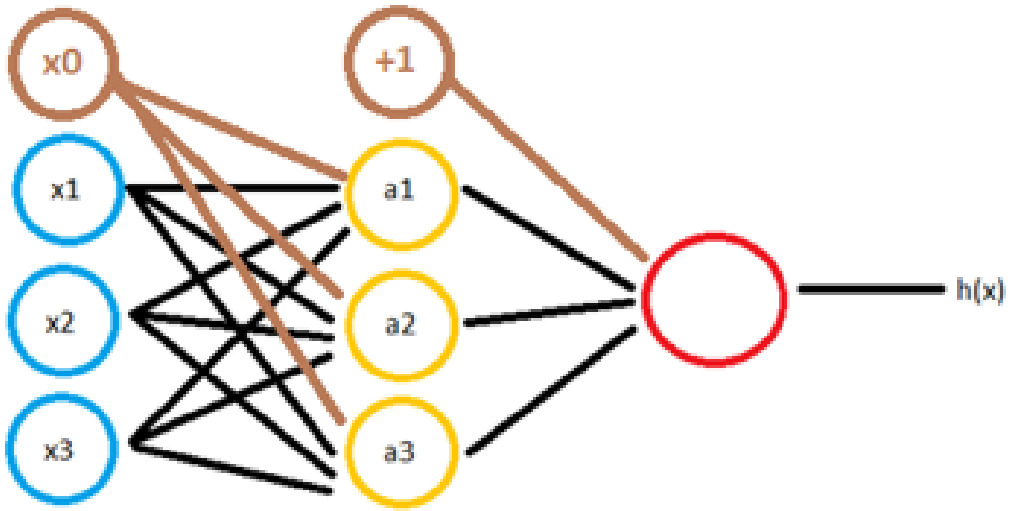


First idea



@Richard, 2017

Artificial neural network



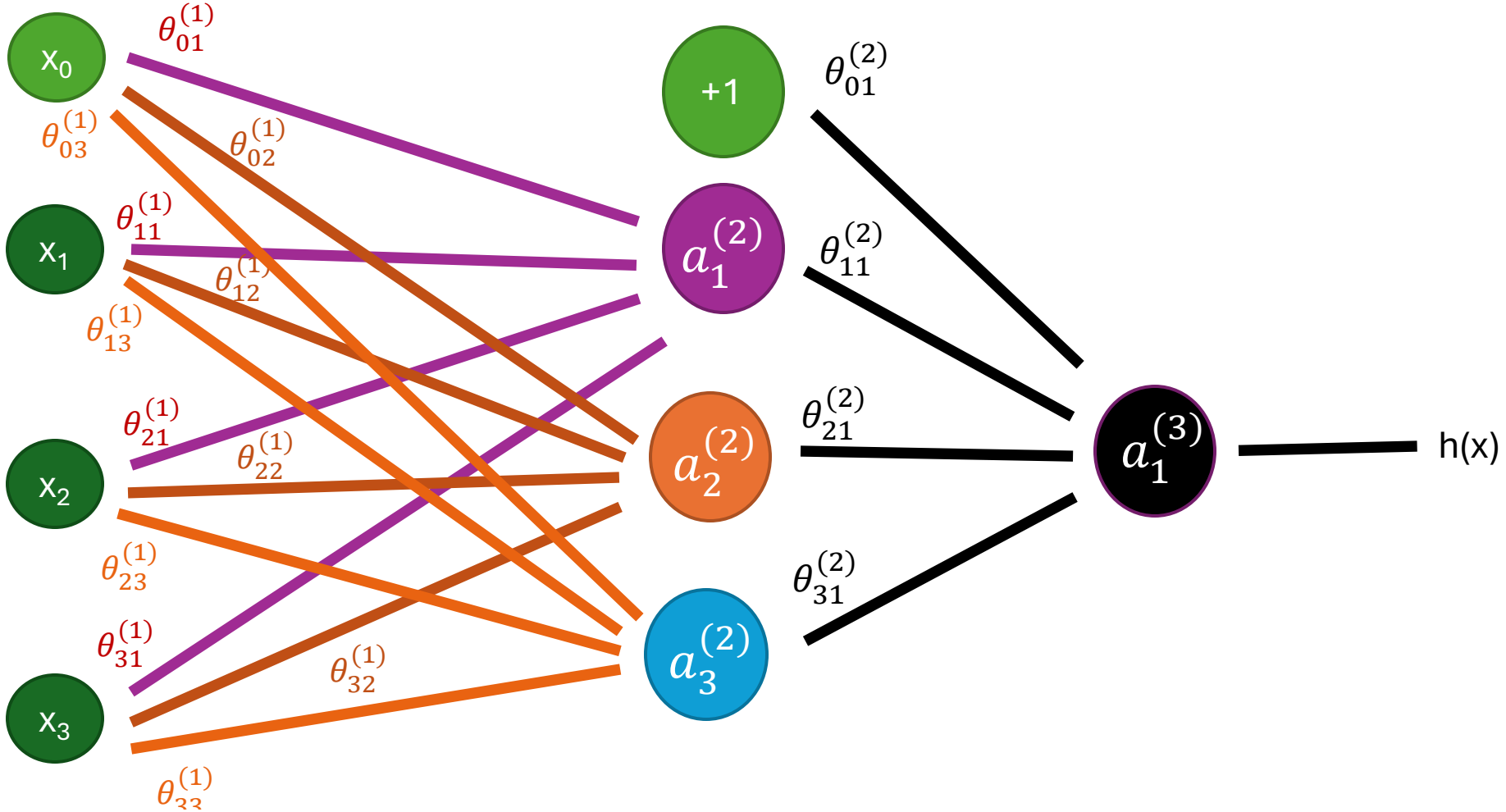
Combination function calculate the value of a node from the other connected nodes and their corresponding weights

Transfer function define the output value

- E.g., sigmoid or logistic function
  - Linear behavior near to 0 and not linear at the extremities
  - Model linear and not linear phenomena

$$g(x) = \frac{1}{1 + \exp(-x)}$$

$$h(x) = a_1^{(3)} = g(\theta_{01}^{(2)}.a_0^{(2)} + \theta_{11}^{(2)}.a_1^{(2)} + \theta_{21}^{(2)}.a_2^{(2)} + \theta_{31}^{(2)}.a_3^{(2)})$$

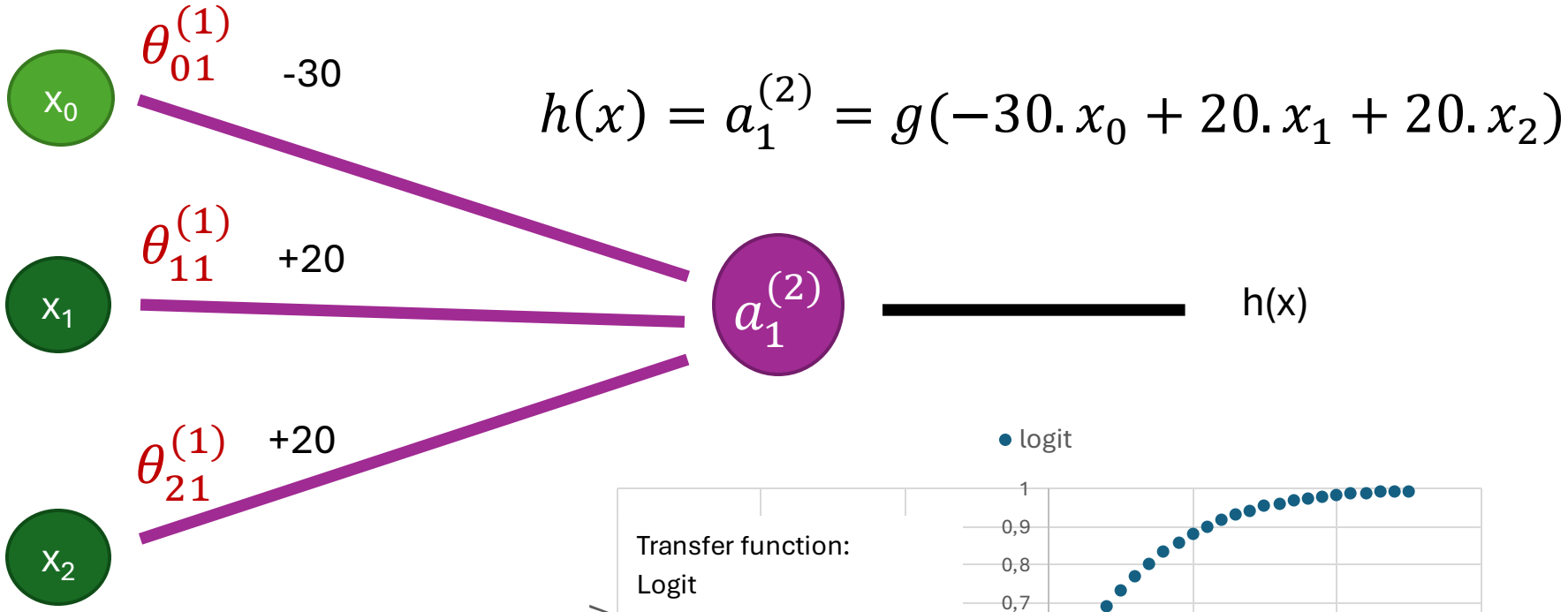


$$a_1^{(2)} = g(\theta_{01}^{(1)}.x_0 + \theta_{11}^{(1)}.x_1 + \theta_{21}^{(1)}.x_2 + \theta_{31}^{(2)}.x_3)$$

$$a_2^{(2)} = g(\theta_{02}^{(1)}.x_0 + \theta_{12}^{(1)}.x_1 + \theta_{22}^{(1)}.x_2 + \theta_{32}^{(2)}.x_3)$$

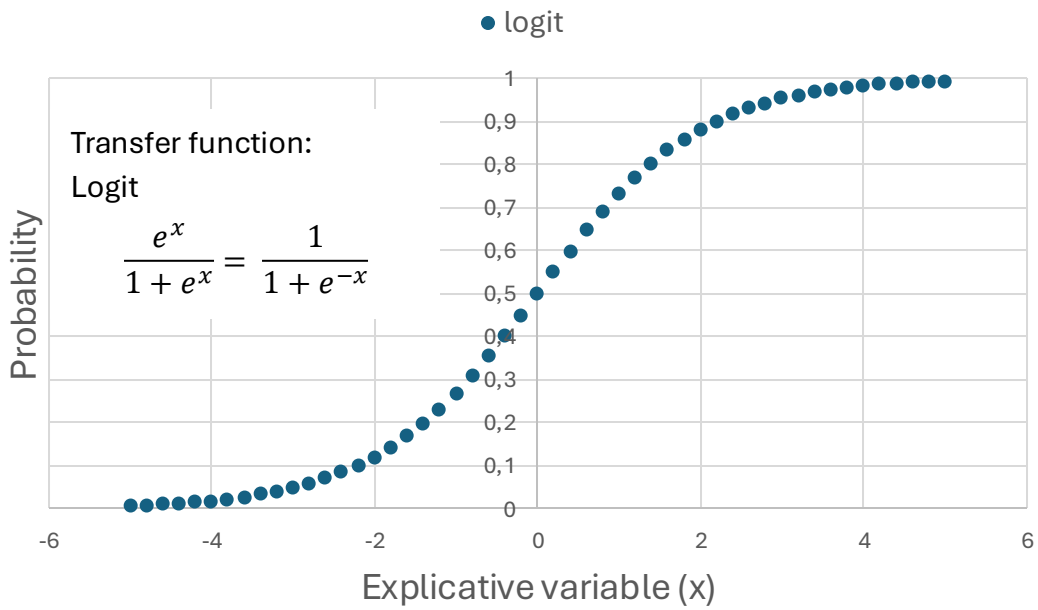
$$a_3^{(2)} = g(\theta_{03}^{(1)}.x_0 + \theta_{13}^{(1)}.x_1 + \theta_{23}^{(1)}.x_2 + \theta_{33}^{(2)}.x_3)$$

Concrete example



$$h(x) = a_1^{(2)} = g(-30.x_0 + 20.x_1 + 20.x_2)$$

x0	x1	x2	h(x)
+1	0	0	g(-30)
+1	1	1	g(10)



Weight estimation

Back-propagation of error gradient

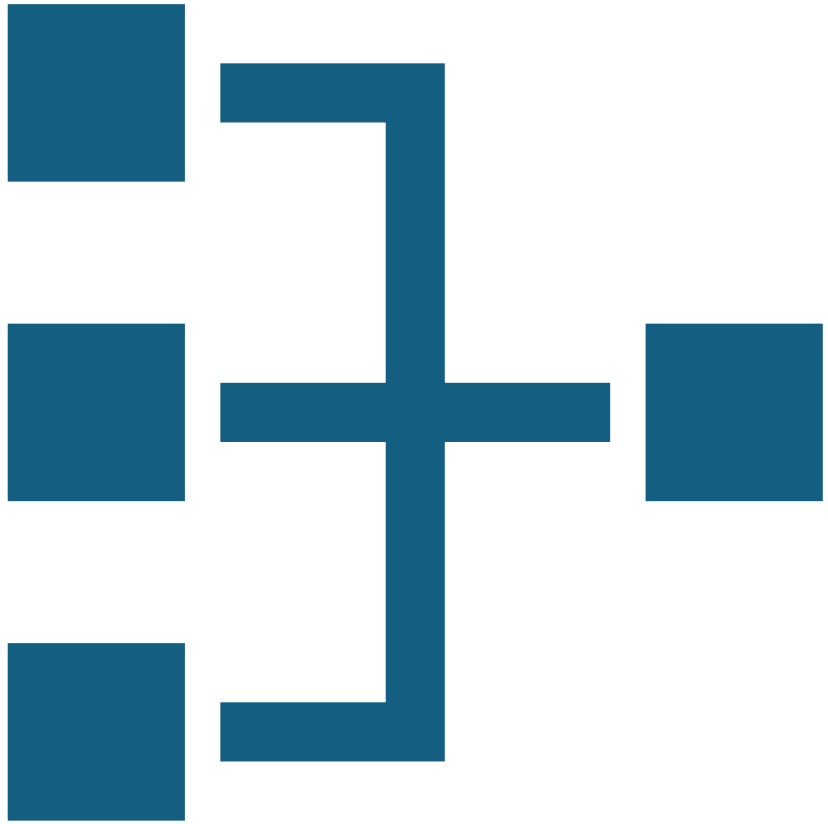
- The oldest algorithm
- Minimize the prediction error similarly to the ordinary least squares
- Iterative estimation
  - Start : attribution of random value to all weights ➔ first prediction ➔ error
  - The values will be modified at each iteration
  - Higher will be the error, stronger will be the change of the weight value.

N = number of samples
K = number of nodes

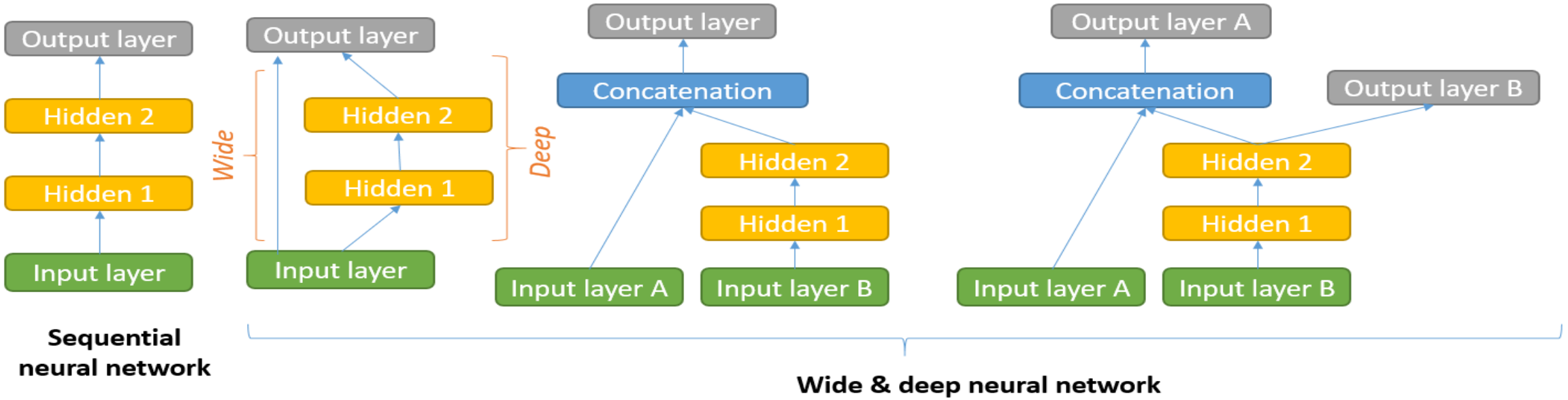
$$e^2 = \sum_{i=1}^{n} \sum_{j=1}^{k} (\hat{y}_{ij} - y_{ij})^2$$

Optimization

- Complex process to optimize the architecture of the neural network
  - Number of nodes/ units on the hidden layer (size)
  - Number of layers (often one hidden layer)
  - Penalty (decay)
  - Number of iterations
- Need the cross-validation to assess the model quality
- Each run can give different cross-validation results due to the random initialisation ➔ fix the seed
- For a first approach:
  - High number of iterations
  - Grid to test the best number of nodes and decay value



Wide & deep neural network



Decrease the weights

- Feature selection
  - Lasso
  - Random Forest
  - …
- PCR components
  - PCAnet
- PLS components

But, the complexity of models changes

- Structure
- Number of parameters

Models already tested in milk spectroscopic analysis

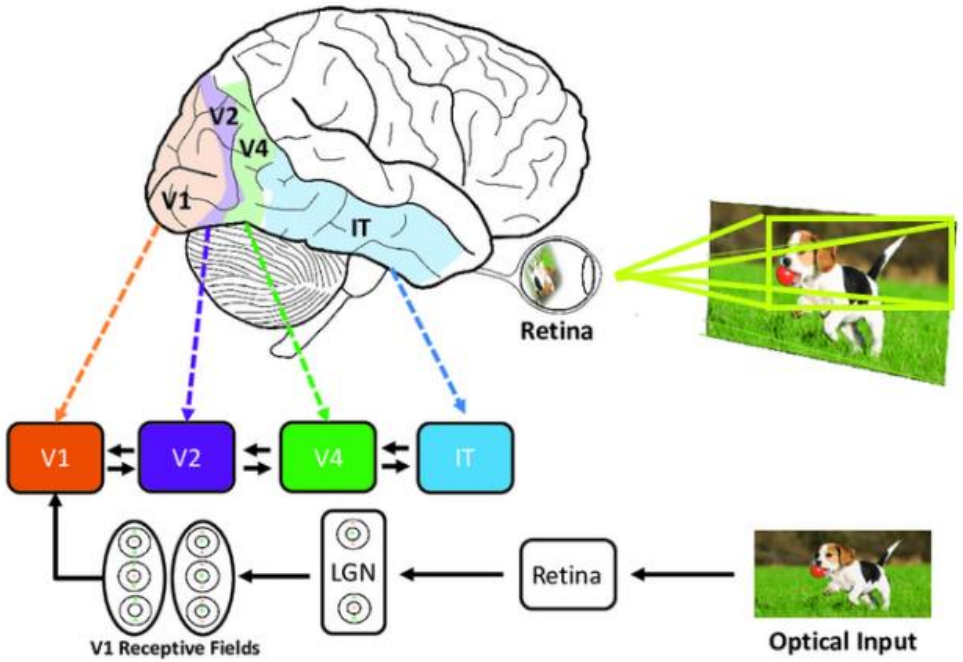
- Multivaried linear regression
- Penalized linear Regression (Ridge, Lasso, Elastic)
- Principal Component regression
- Partial Least Squares Regression
- Support vector machine
- Neural Network – Multilayer perceptron

Multi-layer perceptron

Rapid increase of the number of features

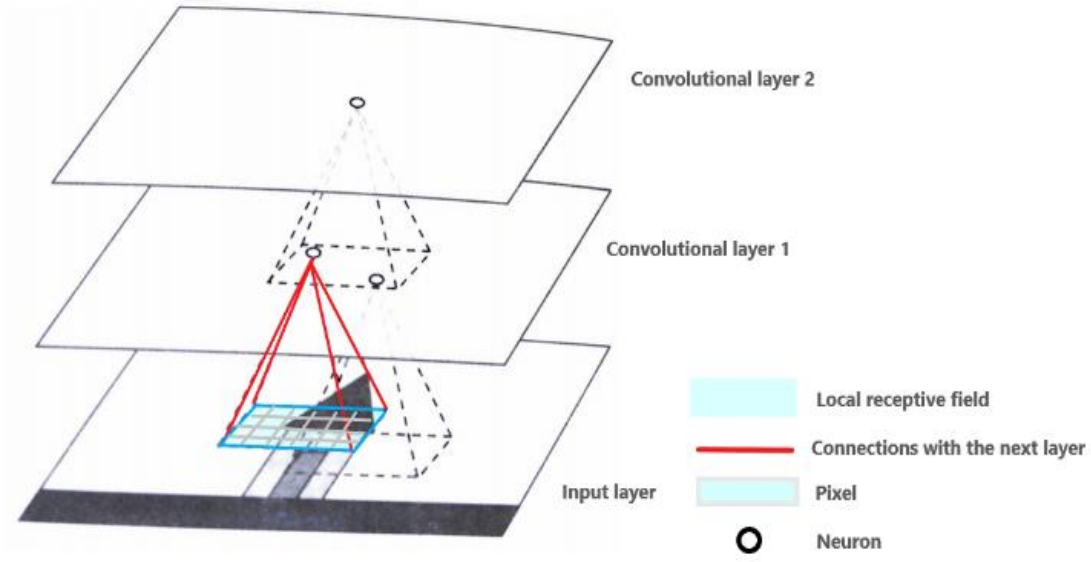
Convolutional neural network (cANN)

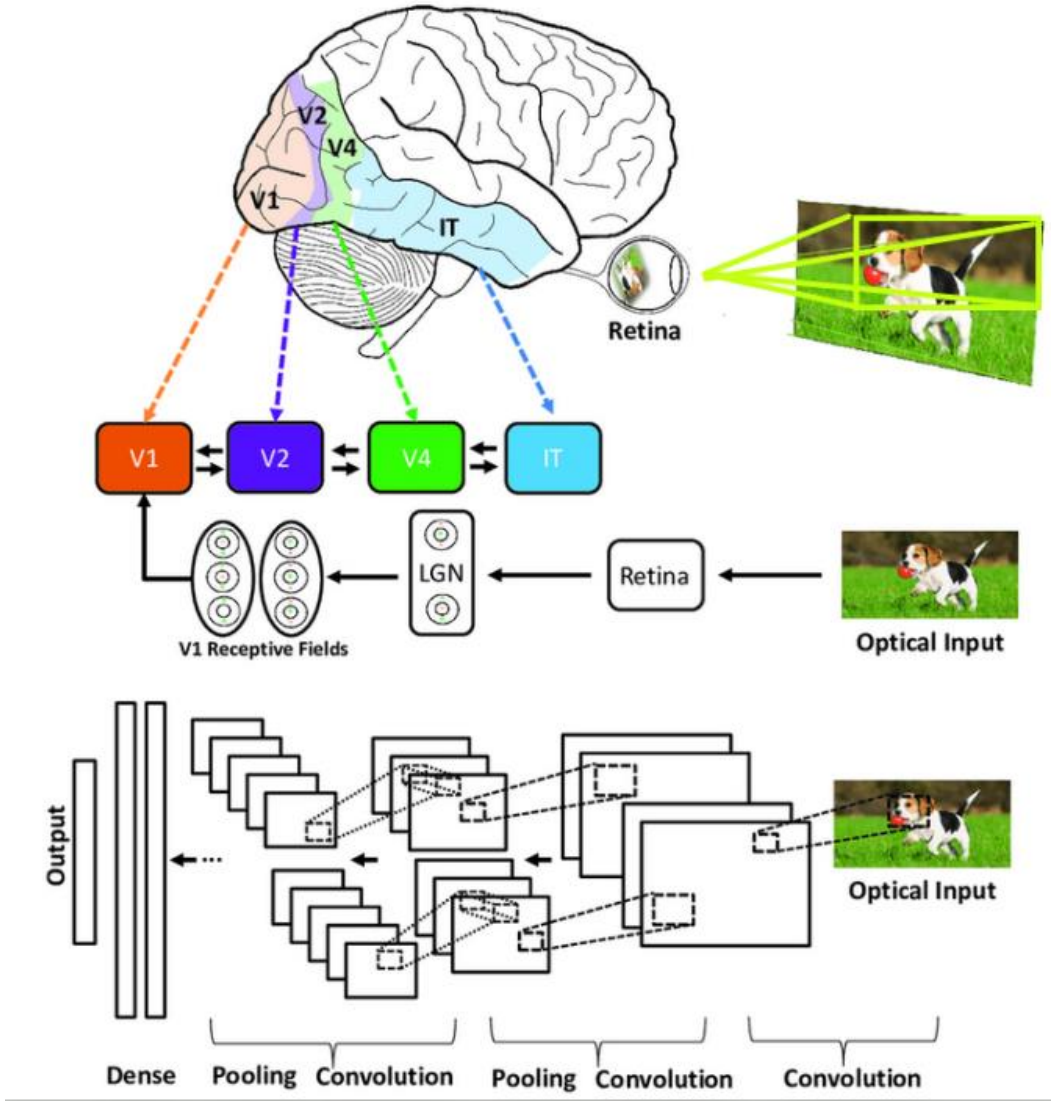
David H. Hubel and Torsten Wiesel in the 50's observed that several neurons in the visual cortex of cats in the brain focus on a restrict region of the visual field and interact only if visual stimuli occurred in this region.



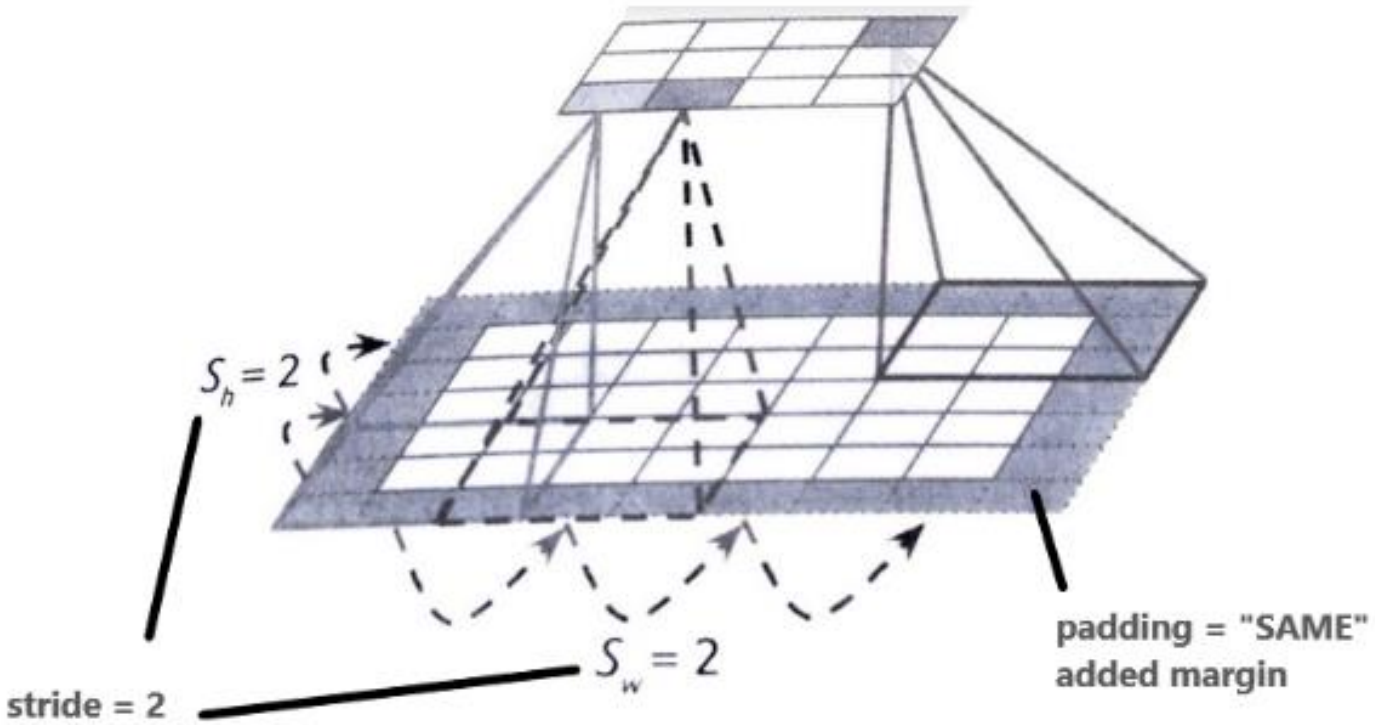
➔ No interest to link all neurons
➔ Local receptive field (LRF)

Local receptive field (LRF)



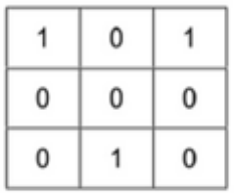


Padding and stride



filter

Input image
Padding = valid (no margin)
Stride = 1

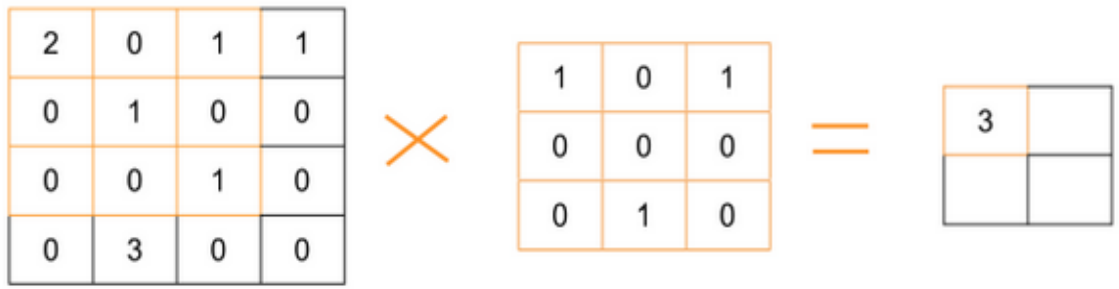


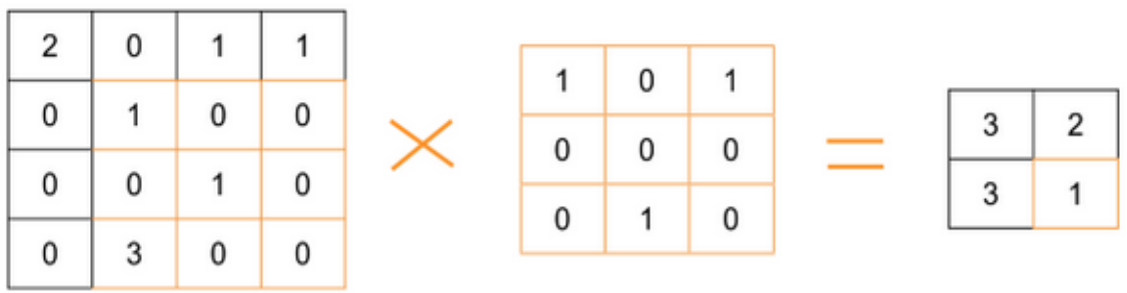
Filter (3x3)

Output image with
stride = 1

Multiplication ==========

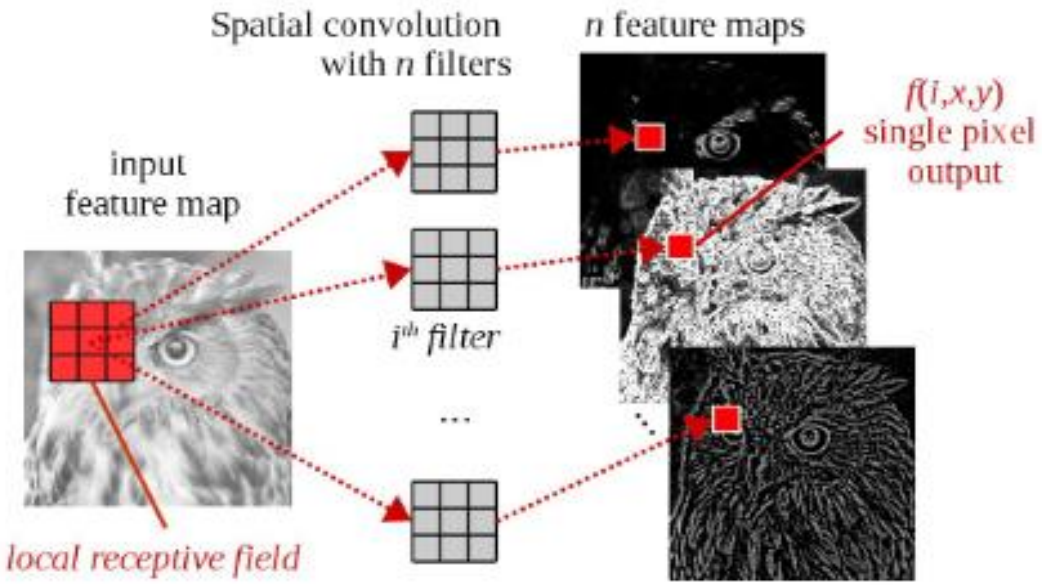
(2x1)+(0*0)+(1*1)
+(0*0)+(1*0)+(0*0)
+(0*0)+(0*1)+(1*0)
= 3

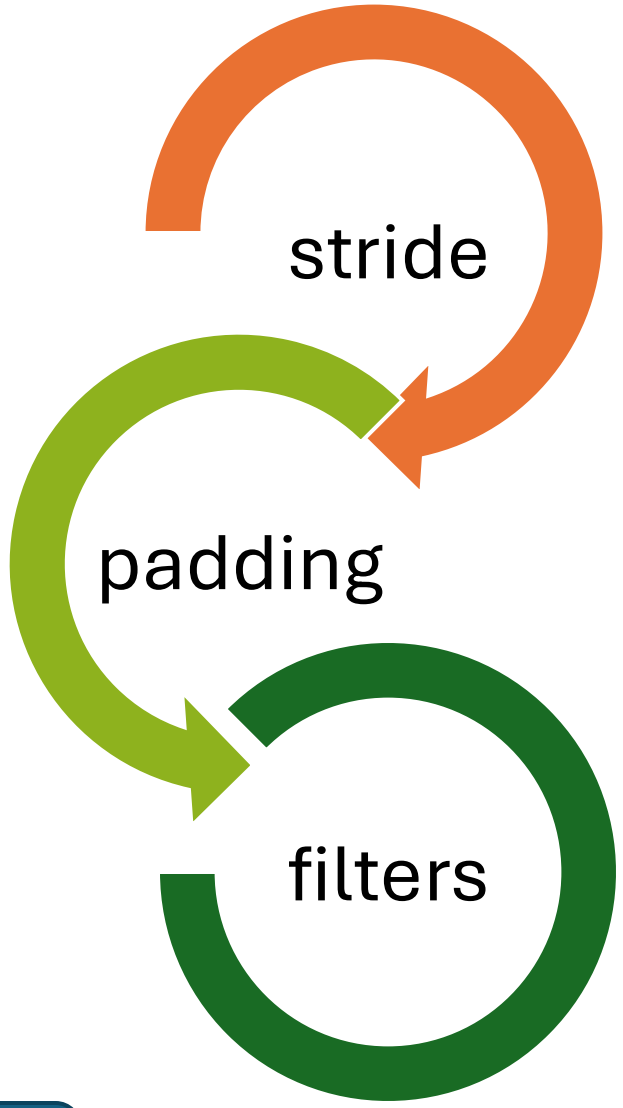




Final output image

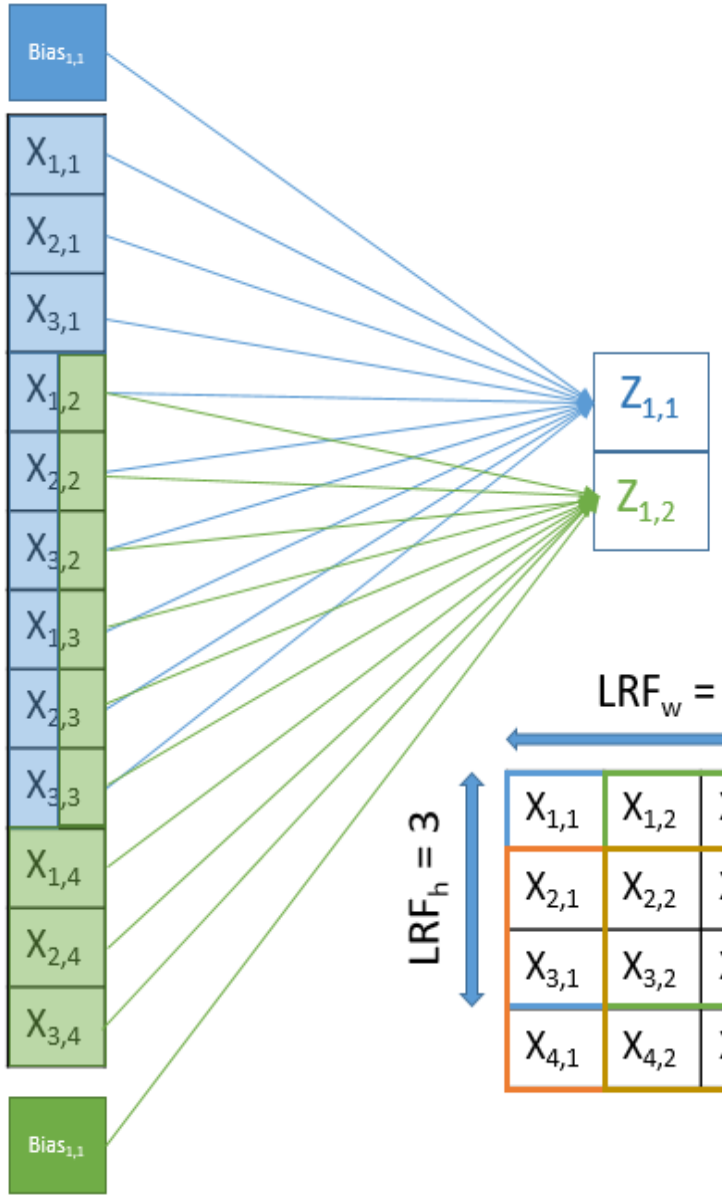
Feature maps





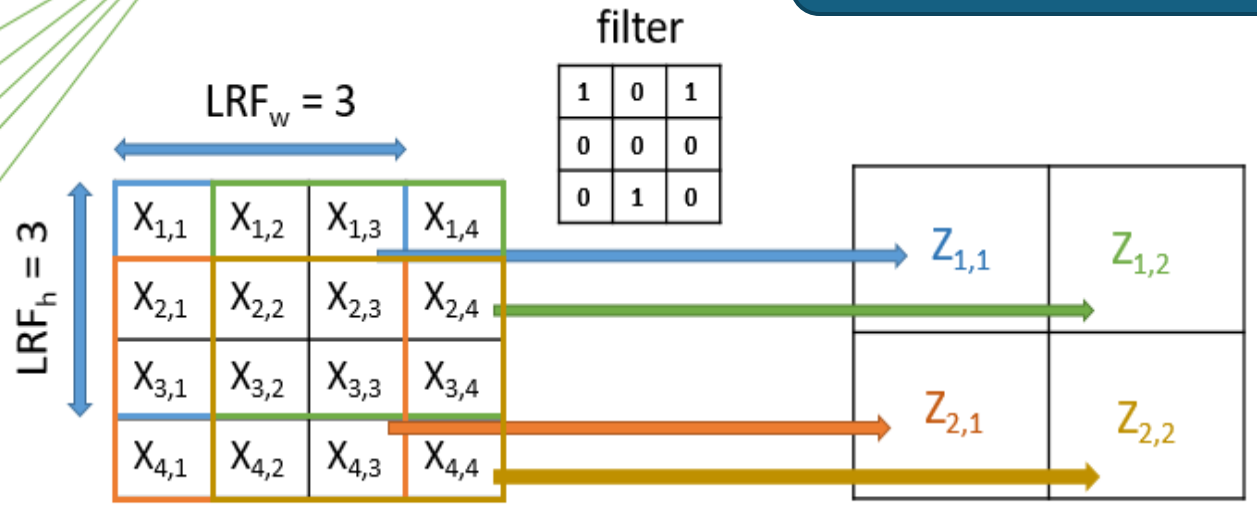
Where are the weights ?

Feature map



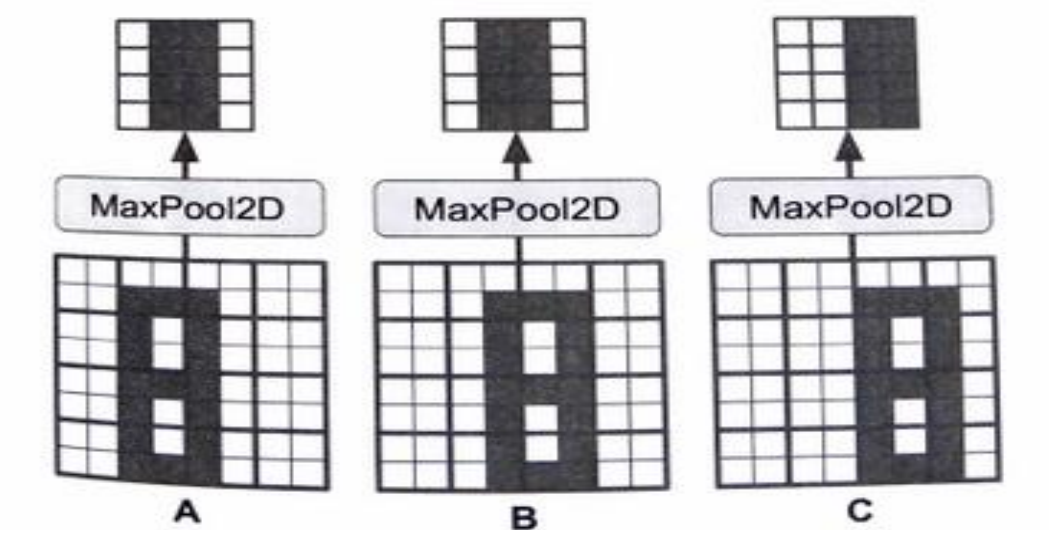
$$z_{1,1} = (f_{1,1} * w_{1,1} * x_{1,1}) + (f_{1,2} * w_{1,2} * x_{1,2}) + (f_{1,3} * w_{1,3} * x_{1,3})$$
$$+(f_{2,1} * w_{2,1} * x_{2,1}) + (f_{2,2} * w_{2,2} * x_{2,2}) + (f_{2,3} * w_{2,3} * x_{2,3})$$
$$+(f_{3,1} * w_{3,1} * x_{3,1}) + (f_{3,2} * w_{3,2} * x_{3,2}) + (f_{3,3} * w_{3,3} * x_{3,3})$$
$$+ bias_{1,1}$$

Not all neurons are connected

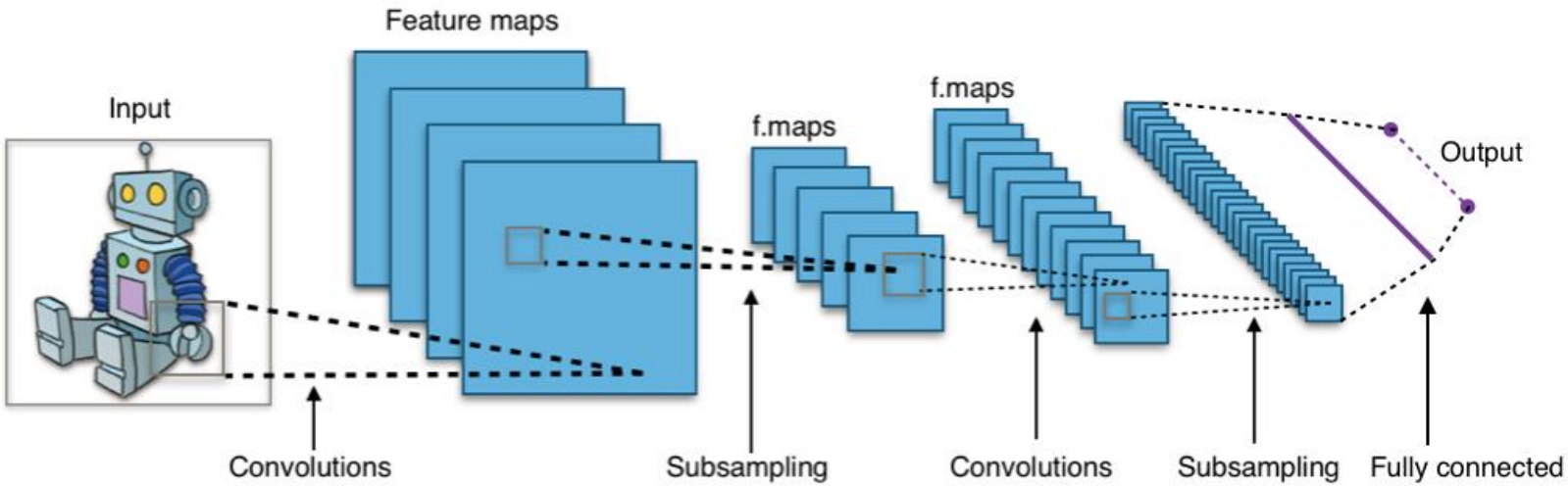


Pooling layer

- decrease the dimensionality of the layer by sub-setting the image
- 2*2 is preferred
- Max or mean is used



cANN



Need to treat the spectrum as an image.

But, the complexity of models changes

- Structure
- Number of parameters

Models already tested in milk spectroscopic analysis

Multivaried linear regression

Penalized linear Regression (Ridge, Lasso, Elastic)

Principal Component regression

Partial Least Squares Regression

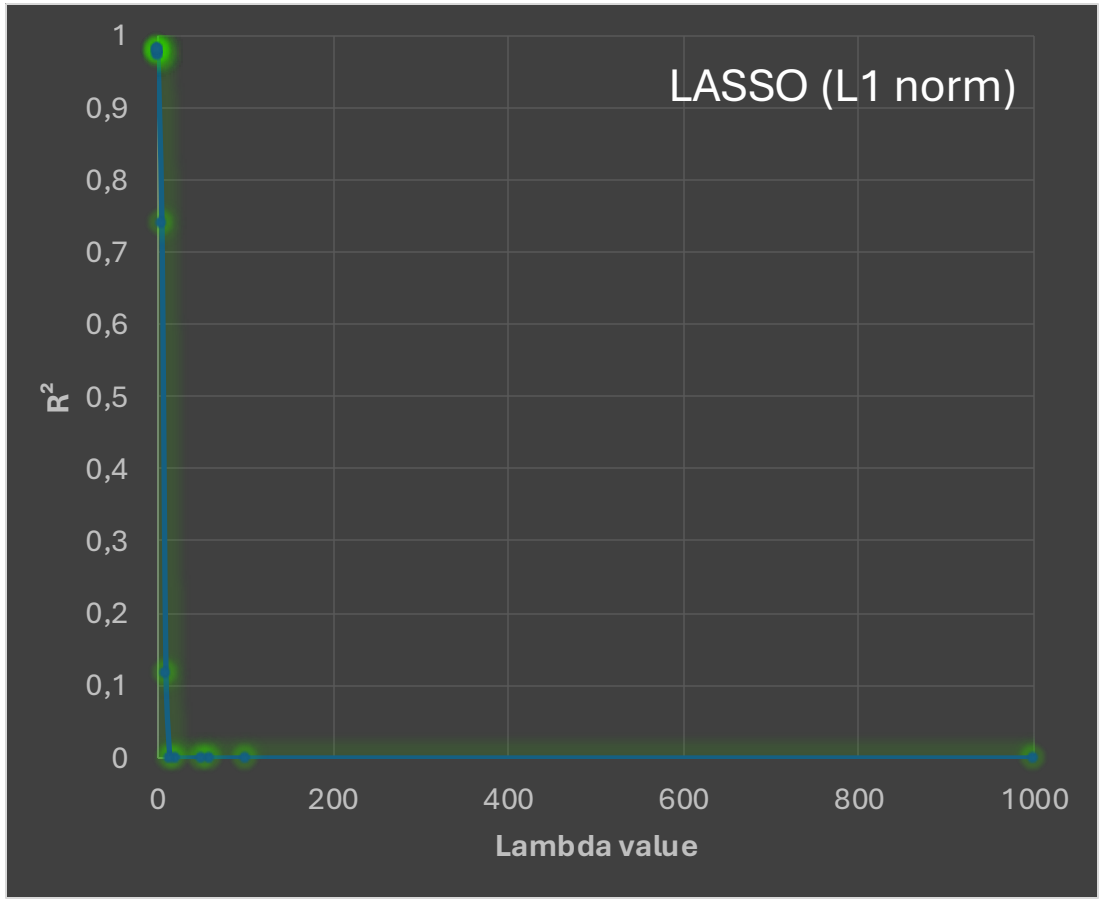
Support vector machine

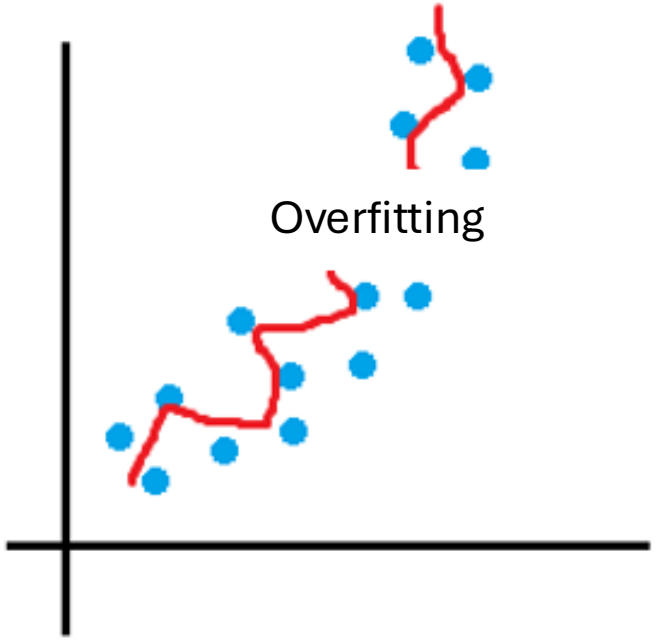
Neural Network – Multilayer perceptron

Convolutional neural network

Linear until non-linear relationships

But the complexity of models changes





Higher features, higher train R^2

Impact of validation

- External validation:
  - Independent datasets
  - Take care to the hidden dependencies
- Internal validation:
  - Simple: Random splitting of the dataset
    - 2/3 calibration and 1/3 validation
  - Cross-validation
  - Bootstrapping …

Following the validation, the results will be different

But, the complexity of models changes

- Structure
- Number of parameters

Models already tested in milk spectroscopic analysis

Multivaried linear regression

Penalized linear Regression (Ridge, Lasso, Elastic)

Principal Component regression

Partial Least Squares Regression

Support vector machine

Neural Network – Multilayer perceptron

Convolutional neural network

Linear until non-linear relationships

Higher computational resources

Example with saturated fatty acids

Method	N	Parameters	RMSEcv		R^2cv		Validation		
			Mean	SD	Mean	SD	N	RMSE_v	R^2v
ridge	1500	1.00E-04	0.10	0.01	0.98	0.00	500	0.10	0.99
lasso	1500	0.1	0.39	0.05	0.79	0.06	500	0.23	0.93
enet	1500	0.525_1e-04	0.09	0.01	0.99	0.00	500	0.10	0.99
pls	1500	28	0.09	0.01	0.99	0.00	500	0.10	0.99
pcr	1500	29	0.12	0.02	0.97	0.01	500	0.13	0.98
rf	1500	289	0.17	0.01	0.95	0.00	500	0.20	0.94
RRF	1500	145_0.01_1	0.17	0.02	0.95	0.00	500	0.19	0.95
pcaNNet	1500	5_0.001	1.84	0.03	0.19	0.23	500	1.88	0.47

There is not only one interesting model …

Journal of Dairy Science
Volume 103, Issue 12, December 2020, Pages 11585-11596

Research

A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra

H. Soyeurt [1], C. Grelet [2], S. McParland [3], M. Calmels [4], M. Coffey [5], A. Tedde [1], P. Delhez [1 6], F. Dehareng [2], N. Gengler [1]

Table 1. The 10-fold cross-validation and external-validation performances for predicting lactoferrin content in milk using 4 different machine learning algorithms[1]

		PLSR	PLS + Linear SVR	PLS + Polynomial SVR	PLS + ANN
Selection function		**oneSE**	**oneSE**	**best**	**best = oneSE[2]**
Calibration (n = 5,541)	Parameters	nLV[3] = 23	C[4] = 5	degree = 3; scale = 0.01; C = 1	size = 4; decay = 0.5
	R^2c	0.53	0.53	0.64	0.60
	RMSEc	140.94	144.32	125.89	130.59
Cross-validation	R^2cv	0.51	0.53	0.56	0.55
	R^2cv SD	0.03	0.03	0.03	0.03
	RMSEcv	144.31	144.60	138.40	139.01
	RMSEcv SD	5.77	5.61	8.08	5.05
	RPD	1.43	1.42	1.49	1.48
External validation (n = 836)	R^2v	0.61	0.63	0.62	0.60
	RMSEv	163.76	174.92	166.75	162.17

Soyeurt et al., JDS 2020

Example : **Lactoferrin**

Mean : 260 mg/L

|Difference| $_{\text{PLS vs ANN}}$: 1.59 mg/L ➔ 0.61%

|Difference| $_{\text{PLS vs LSVM}}$: 11.16 mg/L ➔ 4.29%

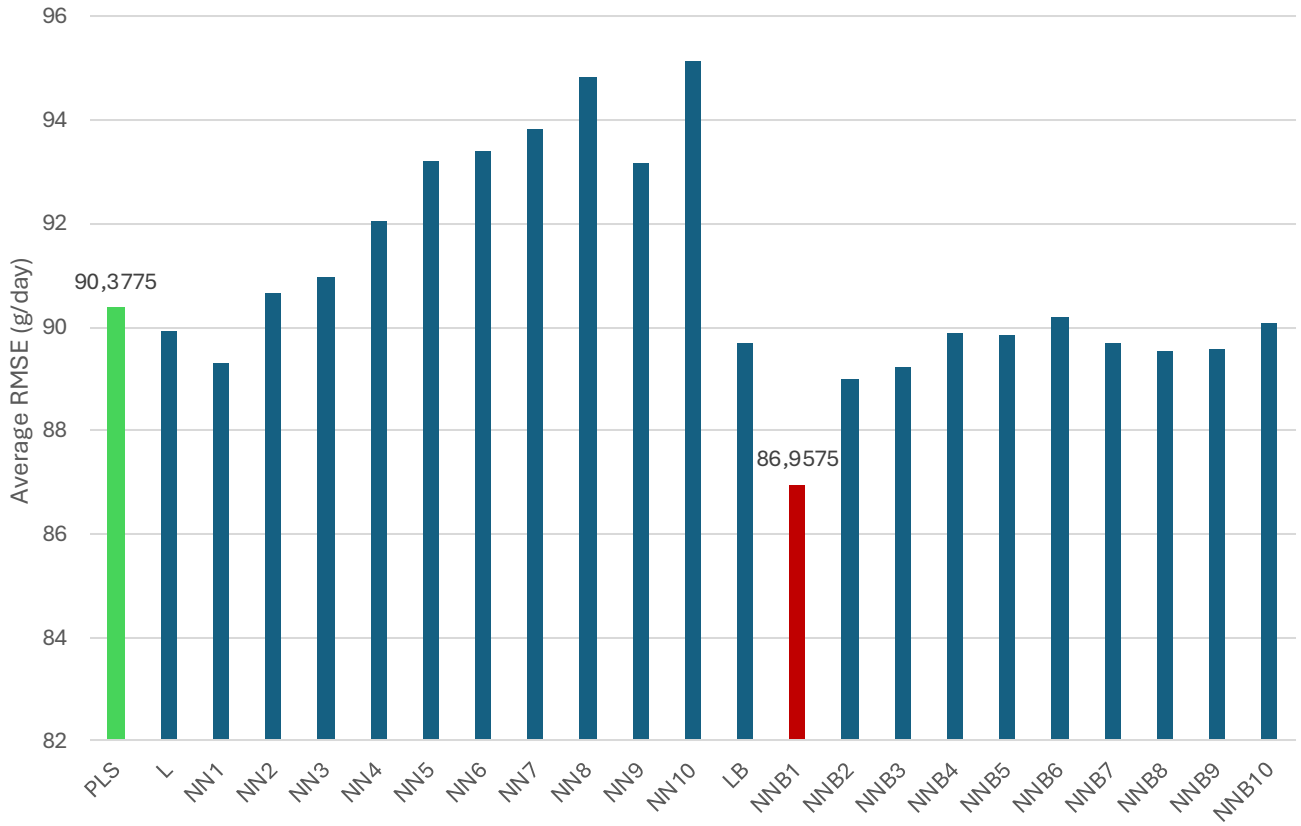
|Difference| $_{\text{PLS vs PSVM}}$: 2.99 mg/L ➔ 1.15%

|Difference$_{\text{| ANN vs LSVM}}$: 12.75 mg/L ➔ 4.90%

|Difference$_{\text{| ANN vs PSVM}}$: 4.58 mg/L ➔ 1.76%

|Difference$_{\text{| PSVM vs LSVM}}$: 8.17 mg/L➔ 3.14%

Small differences !



Journal of Dairy Science
Volume 105, Issue 10, October 2022, Pages 8272-8285

Research

Predicting methane emission in Canadian Holstein dairy cattle using milk mid-infrared reflectance spectroscopy and other commonly available predictors via artificial neural networks

Saeed Shadpour [1], Tatiane C.S. Chud [1], Dagnachew Hailemariam [2], Graham Plastow [2], Hinayah R. Oliveira [1], Paul Stothard [2], Jan Lassen [3], Filippo Miglior [1 4], Christine F. Baes [1], Dan Tulpan [1], Flavio S. Schenkel [1]

Example : **Methane**

Mean : +/- 400g/jour

|Difference| $_{\text{PLS vs NN}}$: 3g/day ➔ 0.75%

Method	Predictor	Statistic					Method						
		Bias	RMSE	r	b	RPIQ			Bias	RMSE	r	b	RPIQ
Partial least squares	Previous						Neural networks	Previous					
	a.m.	-0.23 (8.97)	43.04 (2.36)	0.64 (0.03)	0.94 (0.11)	1.76 (0.16)		a.m.	0.12 (8.66)	42.50 (1.64)	0.64 (0.02)	0.90 (0.11)	1.76 (0.16)
	p.m.	0.03 (9.49)	43.73 (2.48)	0.63 (0.03)	0.93 (0.11)	1.74 (0.16)		p.m.	-0.11 (10.00)	43.15 (2.04)	0.64 (0.03)	0.87 (0.10)	1.73 (0.18)
	a.m. and p.m.	-0.01 (8.89)	42.15 (1.52)	0.65 (0.03)	0.94 (0.10)	1.77 (0.17)		a.m. and p.m.	-0.42 (8.26)	41.14 (1.86)	0.67 (0.02)	0.90 (0.09)	1.81 (0.16)
	Average a.m. and p.m.	-0.09 (8.81)	41.76 (1.98)	0.66 (0.03)	0.95 (0.11)	1.79 (0.18)		Average a.m. and p.m.	-0.42 (7.96)	41.45 (1.45)	0.66 (0.03)	0.91 (0.12)	1.80 (0.15)
	Following							Following					
	a.m.	-0.49 (9.34)	44.62 (1.46)	0.61 (0.02)	0.95 (0.09)	1.71 (0.11)		a.m.	-0.73 (8.93)	43.59 (1.42)	0.64 (0.02)	0.92 (0.11)	1.75 (0.11)
	p.m.	-0.06 (9.93)	45.23 (0.90)	0.60 (0.04)	0.93 (0.12)	1.69 (0.12)		p.m.	-0.36 (9.08)	44.36 (0.37)	0.62 (0.03)	0.89 (0.07)	1.72 (0.11)
	a.m. and p.m.	-0.20 (9.14)	44.14 (0.76)	0.62 (0.03)	0.94 (0.10)	1.73 (0.11)		a.m. and p.m.	-0.96 (8.84)	42.48 (1.02)	0.66 (0.03)	0.89 (0.08)	1.80 (0.13)
	Average a.m. and p.m.	-0.56 (9.43)	44.00 (1.40)	0.63 (0.03)	0.95 (0.10)	1.74 (0.11)		Average a.m. and p.m.	-0.46 (8.23)	42.33 (1.38)	0.66 (0.03)	0.92 (0.09)	1.81 (0.13)
	Flanking							Flanking					
	a.m.	-0.42 (9.08)	38.09 (1.87)	0.68 (0.03)	0.95 (0.11)	1.87 (0.17)		a.m.	0.16 (8.97)	38.23 (1.50)	0.68 (0.03)	0.92 (0.12)	1.86 (0.17)
	p.m.	-0.11 (9.71)	39.25 (1.57)	0.66 (0.03)	0.93 (0.12)	1.81 (0.15)		p.m.	-0.46 (9.27)	39.16 (0.94)	0.66 (0.02)	0.87 (0.07)	1.82 (0.14)
	a.m. and p.m.	-0.10 (8.83)	37.98 (1.36)	0.68 (0.03)	0.95 (0.11)	1.87 (0.16)		a.m. and p.m.	-0.77 (8.43)	37.17 (1.53)	0.70 (0.02)	0.91 (0.09)	1.91 (0.15)
	Average a.m. and p.m.	-0.25 (8.92)	37.56 (2.00)	0.69 (0.03)	0.96 (0.11)	1.90 (0.16)		Average a.m. and p.m.	-0.33 (7.95)	37.46 (4.01)	0.71 (0.03)	0.92 (0.11)	1.95 (0.16)

ELSEVIER

Journal of Dairy Science
Volume 107, Issue 2, February 2024, Pages 978-991

Research

Predicting methane emissions of individual grazing dairy cows from spectral analyses of their milk samples

S. McParland, M. Frizzarin, B. Lahart, M. Kennedy, L. Shalloo, M. Egan, K. Starsmore, D.P. Berry

Example : **Methane**

Also low differences

Open Access Article

Identifying Health Status in Grazing Dairy Cows from Milk Mid-Infrared Spectroscopy by Using Machine Learning Methods

by Brenda Contla Hernández [1], Nicolas Lopez-Villalobos [2] and Matthieu Vignes [1,*]

[1] School of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand
[2] School of Agriculture and Environment, Massey University, Palmerston North 4442, New Zealand
* Author to whom correspondence should be addressed.

Animals **2021**, *11*(8), 2154; https://doi.org/10.3390/ani11082154

More nuanced in classification

Table 2. Performance of classification models obtained in 10 Monte Carlo cross-validation for classifying any health problem and healthy cows during lactation (early, mid and lactation) at two dairy farms during the 2016 production season [1].

Models [2]	Sensitivity	Specificity	Accuracy	PPV	NPV	AUC	MCC
PLS-DA	65.60 ± 5.97	79.59 ± 2.36	78.85 ± 2.23	15.25 ± 3.07	97.66 ± 0.5	72.59 ± 3.27	0.24 ± 0.04
RF	46.22 ± 8.62	79.26 ± 2.15	77.51 ± 1.75	10.94 ± 1.88	96.38 ± 0.73	62.74 ± 3.78	0.14 ± 0.04
SVM	66.39 ± 6.80	76.39 ± 2.92	75.84 ± 2.42	13.48 ± 1.62	97.61 ± 0.61	71.39 ± 2.37	0.22 ± 0.02
NN	61.74 ± 15.99	97.00 ± 2.85	95.16 ± 3.26	59.99 ± 26.20	97.87 ± 0.87	79.37 ± 9.16	0.58 ± 0.22
CNN	57.02 ± 12.70	92.5 ± 5.27	90.63 ± 4.98	33.82 ± 13.41	97.5 ± 0.75	74.76 ± 6.88	0.39 ± 0.13
ESA	57.15 ± 12.38	87.61 ± 6.19	86.02 ± 6.21	24.06 ± 13.07	97.36 ± 0.77	72.38 ± 8.48	0.31 ± 0.16
ESMJ	60.75 ± 5.98	83.57 ± 2.56	82.36 ± 2.27	17.18 ± 3.21	97.46 ± 0.55	72.16 ± 2.9	0.25 ± 0.04
ESWA	56.43 ± 14.56	85.13 ± 7.41	83.61 ± 7.36	21.33 ± 14.18	97.22 ± 0.97	70.78 ± 9.71	0.27 ± 0.17

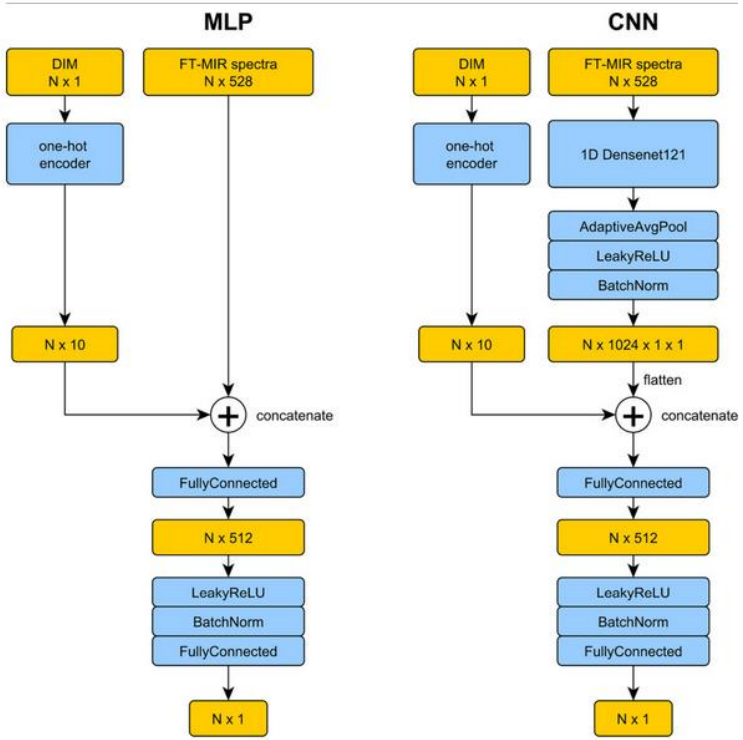
[1] These values correspond to the mean ± SD obtained by 10-fold Monte Carlo cross-validation for classifying any health problem (lameness, mastitis, reproductive disorder, etc.). From the cows' records, the positive cases were cows that had any illness (lameness, mastitis, reproductive disorder, etc.) and negative cases were cows who were healthy (no diagnosed disease); SD = Standard deviation; PPV = positive predicted value; NPV = negative predicted value; AUC = area under the receiver operating characteristic curve; MCC = Matthews correlation coefficient. [2] Models used to perform the classification: PLS-DA = partial least squares discriminant analysis, RF = random forest, SVM = support vector machine, NN = neural network, CNN = convolutional neural network, ESA = ensemble stacking average, ESMJ = ensemble stacking major voting and ESWA = ensemble stacking weighted average.

Journal of Dairy Science
Volume 105, Issue 4, April 2022, Pages 3615-3632

Research

Pregnancy status predicted using milk mid-infrared spectra from dairy cattle

K.M. Tiplady [1 2], M.-H. Trinh [1], S.R. Davis [1], R.G. Sherlock [1], R.J. Spelman [1], D.J. Garrick [2], B.L. Harris [1]



More nuanced in classification

Table 3. Model performance for multilayer perceptron (MLP) and convolutional neural network (CNN) approaches based on strategy 3 data[1]: accuracy (Acc), sensitivity (Sens), specificity (Spec), and area under the receiver operating characteristic curve (AUC) values within the training, herd-independent validation (VAL-Test) and pregnancy-associated glycoproteins validation (VAL-PAG) data sets

Deep learning approach[2] and model[3]	Training				Test validation (VAL-Test)				Glycoprotein-based validation (VAL-PAG)			
	Acc	Sens	Spec	AUC	Acc	Sens	Spec	AUC	Acc	Sens	Spec	AUC
MLP approach												
FT-MIR spectra	0.592	0.574	0.611	0.628	0.586	0.580	0.607	0.632	0.664	0.672	0.569	0.669
FT-MIR spectra + DIM	0.594	0.621	0.566	0.631	0.614	0.629	0.564	0.635	0.692	0.709	0.499	0.647
FT-MIR spectra (pre-adjusted for DIM)	0.559	0.554	0.564	0.583	0.562	0.567	0.547	0.581	0.554	0.547	0.636	0.636
CNN approach												
FT-MIR spectra	0.625	0.625	0.625	0.675	0.611	0.620	0.582	0.641	0.684	0.696	0.554	0.676
FT-MIR spectra + DIM	0.645	0.670	0.620	0.700	0.636	0.659	0.563	0.654	0.723	0.741	0.519	0.685
FT-MIR spectra (pre-adjusted for DIM)	0.982	0.975	0.988	0.998	0.668	0.790	0.273	0.551	0.759	0.805	0.266	0.564

The variability is bigger within the same model using different sets of features

Journal of Dairy Science
Volume 105, Issue 10, October 2022, Pages 8272-8285

Research

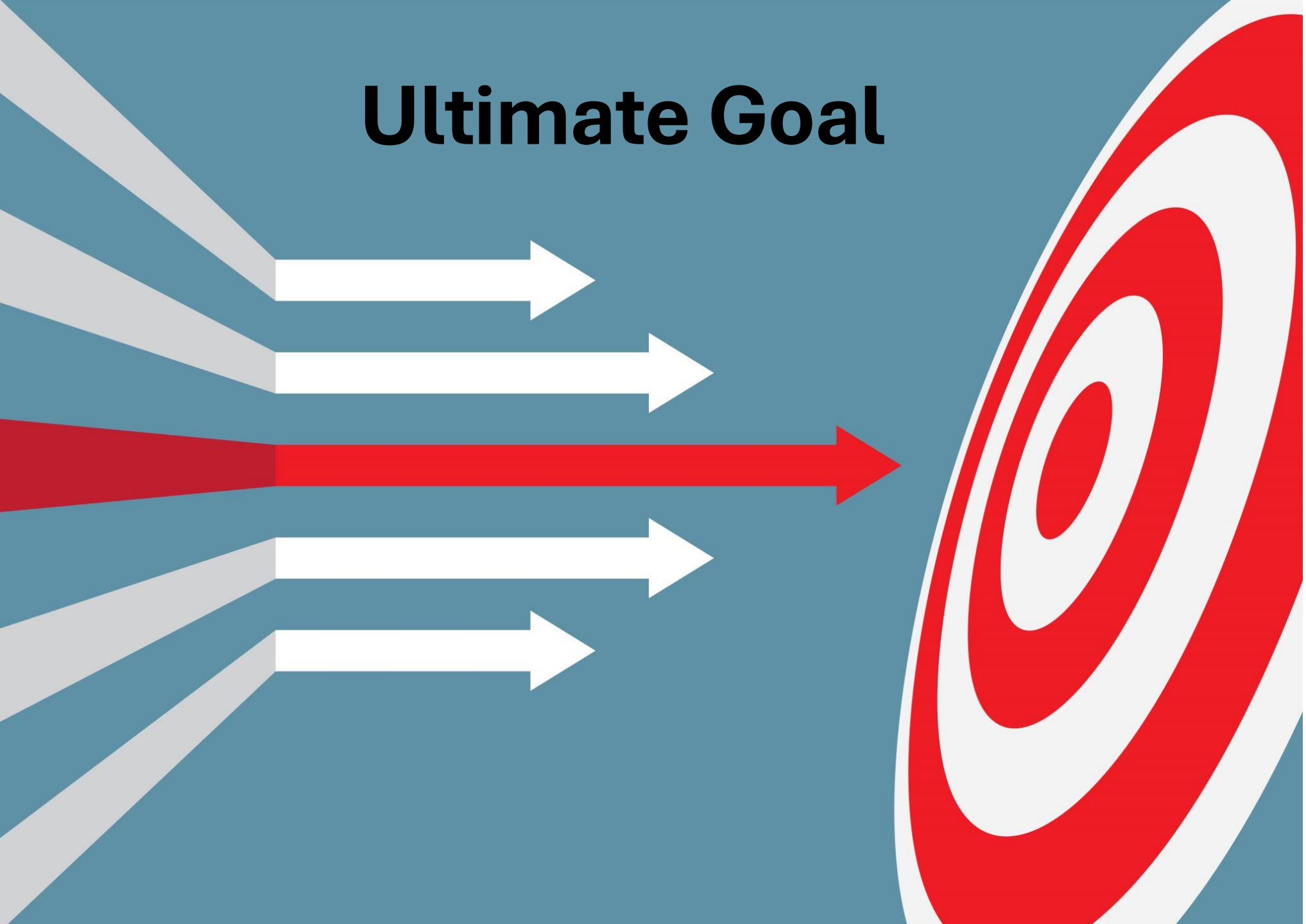
Predicting methane emission in Canadian Holstein dairy cattle using milk mid-infrared reflectance spectroscopy and other commonly available predictors via artificial neural networks

Saeed Shadpour [1], Tatiane C.S. Chud [1], Dagnachew Hailemariam [2], Graham Plastow [2], Hinayah R. Oliveira [1], Paul Stothard [2], Jan Lassen [3], Filippo Miglior [1 4], Christine F. Baes [1], Dan Tulpan [1], Flavio S. Schenkel [1]

Features	PLS	L	NN1	NN2	NN3	NN4	NN5	NN6	NN7	NN8	NN9	NN10	LB	NNB1	NNB2	NNB3	NNB4	NNB5	NNB6	NNB7	NNB8	NNB9	NNB10	SD
1	96.47	96.55	93.08	93.16	93.33	94.74	93.98	94.36	97.65	97.65	96.54	99.39	96.26	91.67	93.01	92.59	92.07	92.41	92.2	92.2	92.22	92.08	92.45	**2.26**
2	95.68	95.29	91.4	93.47	94.64	94.05	95.54	96.11	95.44	98.04	97.66	96.16	95.61	91.93	92.55	93.26	92.89	93.12	93.07	93.06	92.97	92.92	93.1	**1.78**
3	96.32	96.36	93.58	94.44	93.96	94.55	94.71	97.27	97.69	97.39	95.15	98.47	95.99	92.49	93.19	91.46	91.91	92.35	92.45	92.3	92.31	92.46	92.59	**2.13**
4	95.84	95.73	93.4	94.18	96.15	93.58	96.82	94.14	96.05	97.44	96.65	96.57	95.59	91.8	93.41	93.14	93.35	92.84	92.66	92.81	93.17	93.22	92.79	**1.67**
5	96.5	96.72	95.23	94.89	93.41	97.01	98.35	98.94	97.31	99.91	100.94	100.9	96.11	92.6	93.87	93.05	92.82	93.08	92.66	92.98	92.83	92.8	92.83	**2.84**
6	96.75	96.07	94.41	95.86	96.8	97.44	97.89	101.16	100.12	100.65	98.93	101.93	95.66	92.55	93.46	93.87	93.33	93.07	93.4	93.45	93.34	93.35	93.62	**2.94**
7	73.7	72.44	77.56	81.19	81.51	85.42	84.69	82.72	84.49	82.27	80.79	84.57	71.61	71.75	76.2	78.96	79.48	81.59	82.58	78.65	80.36	79.33	81.66	**4.15**
8	71.76	70.31	75.75	77.95	78.08	79.72	83.64	82.61	81.85	85.36	78.72	83.08	70.83	70.87	76.12	77.5	83.09	80.28	82.42	82.1	79.16	80.56	81.71	**4.44**
SD	**10.91**	**11.48**	**7.90**	**6.94**	**7.08**	**6.20**	**5.78**	**7.02**	**6.76**	**6.96**	**8.48**	**7.26**	**11.42**	**9.67**	**7.92**	**6.83**	**5.41**	**5.52**	**4.75**	**5.84**	**6.06**	**5.98**	**5.20**	

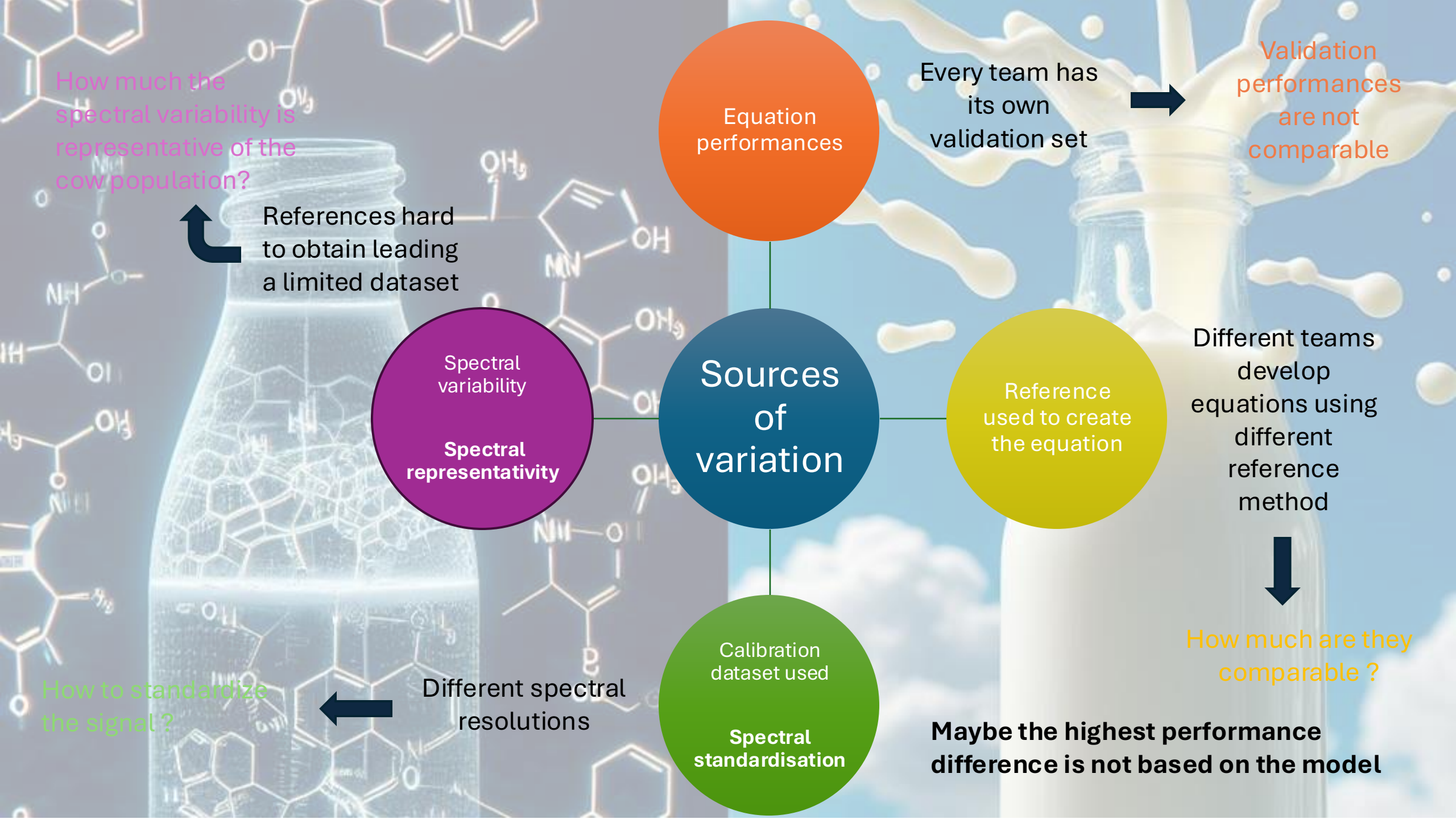
➔ **Not only focus on the model type**

Ultimate Goal



Having the same prediction for the same spectrum if you use a prediction model for the same trait

Not so simple …


How much the spectral variability is representative of the cow population?
References hard to obtain leading a limited dataset
Equation performances
Every team has its own validation set
Validation performances are not comparable
Spectral variability
Spectral representativity
Sources of variation
Reference used to create the equation
Different teams develop equations using different reference method
How much are they comparable ?
How to standardize the signal ?
Different spectral resolutions
Calibration dataset used
Spectral standardisation
Maybe the highest performance difference is not based on the model

Methods
Volume 186, February 2021, Pages 97-111

Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions

C. Grelet [a], P. Dardenne [a], H. Soyeurt [b], J.A. Fernandez [a], A. Vanlierde [a], F. Stevens [a], N. Gengler [b], F. Dehareng [a]

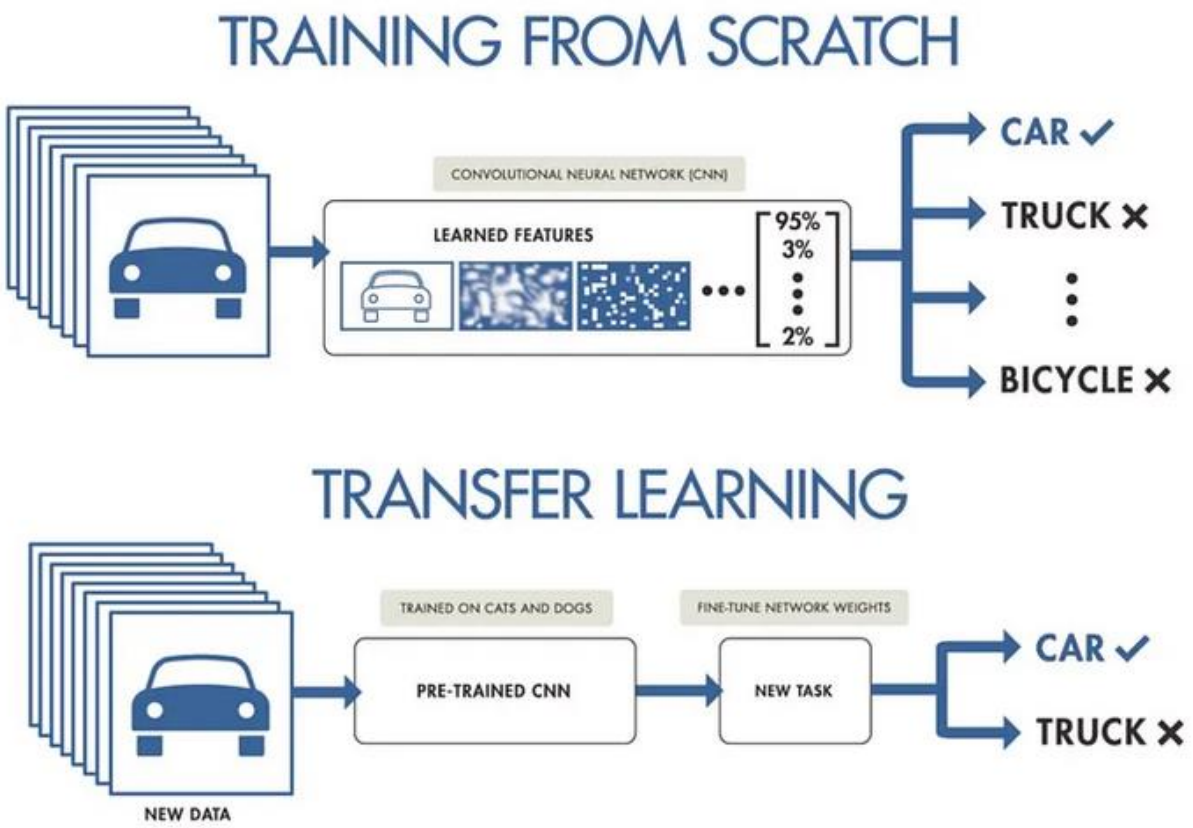
For regression mainly …

Highlights

- Robustness of models is essential to generate accurate phenotypes at a large scale.
- Variability in the reference and spectral data is a key element of robustness.
- In most cases, complexity of models is in contradiction with robustness.
- Models development should consider absorbance reliability of spectral regions.
- Quality assurance methods are necessary to evaluate models and predictions.

More complex models …

Transfer learning is interesting



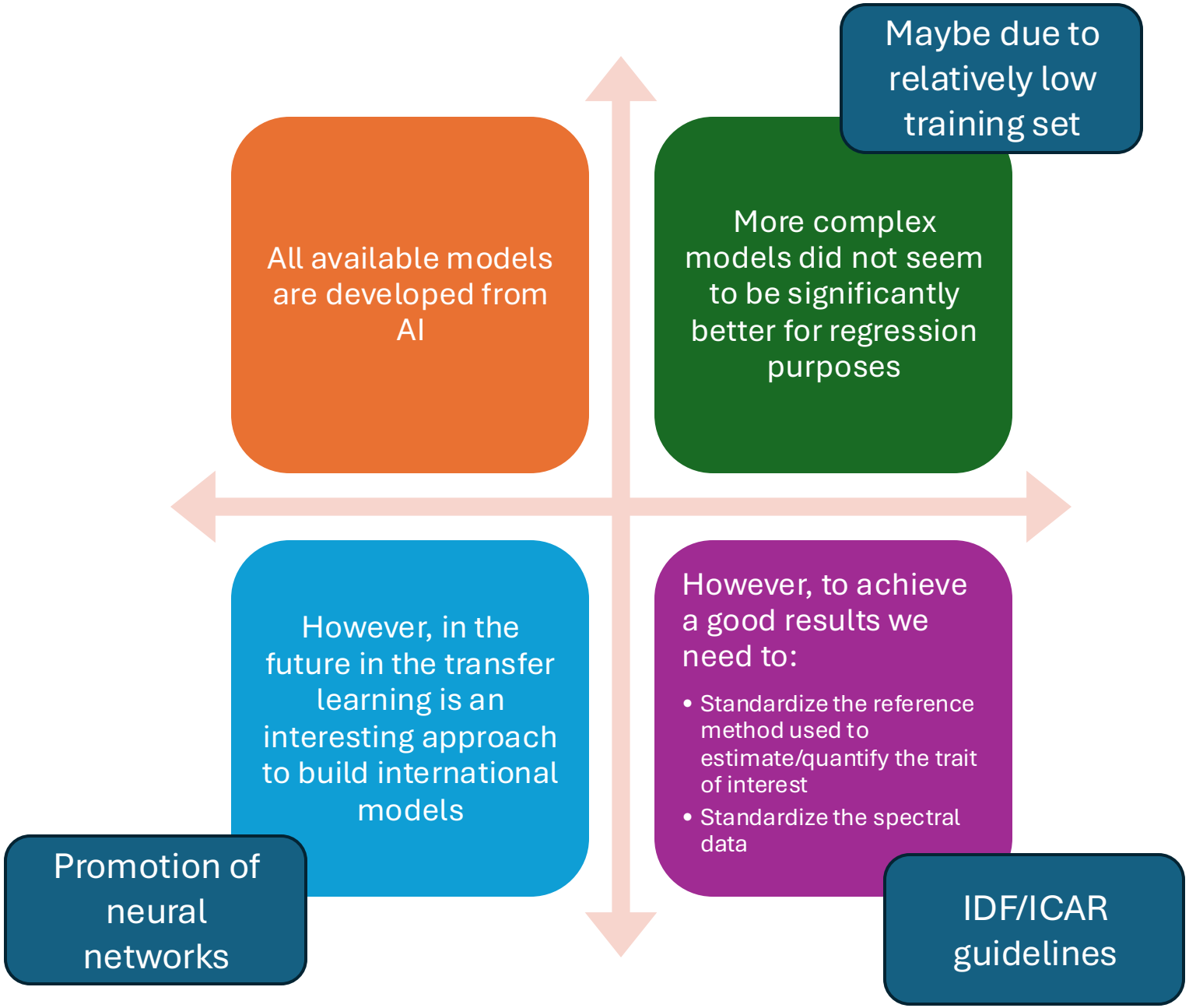
https://iamamadsiddiqui.medium.com/unleashing-the-power-of-transfer-learning-in-artificial-intelligence-e8a9deee62f3

Deeply used with image and video data source

More complex models …

- Transfer learning is interesting:
  - Large number of data is available
  - Difficulty to share data
  - Save computation time

Take home message



LIÈGE université
Gembloux
Agro-Bio Tech

Unraveling the Fascination of Advancing Complexity in Spectrometry Models

Prof. Hélène Soyeurt

4th edition of the Spectroscopy and Chemometrics Workshop – 29th April 2024