EPFL

ChE-609

Seyed Mohamad Moosavi, Kevin Maik Jablonka & Berend Smit

Basics of machine learning for chemistry and materials science

seyedmohamad.moosavi@epfl.ch @SMohMoosavi

Training the model (finding parameters w)

1

System

$x^* \in \mathbb{R}^{n+1}$

Features describing the system

ML model

$f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$

3

Parameters

$w \in \mathbb{R}^p$

2

Properties

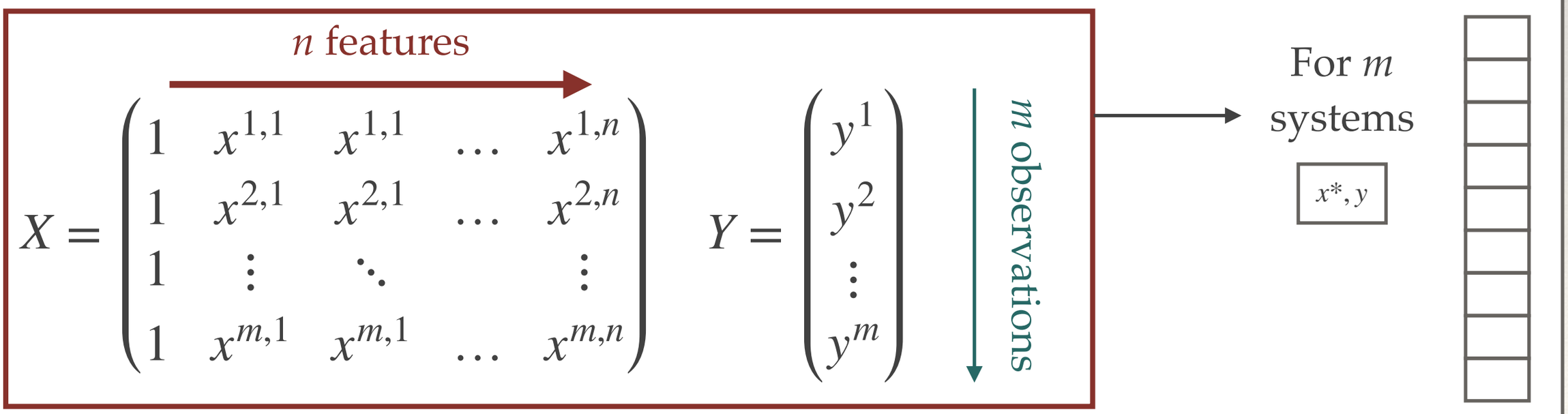
$y \in \mathbb{R}$

Energy, Gas uptake, …

For m systems

x^*, y

X, Y



n features

$$X = \begin{pmatrix} 1 & x^{1,1} & x^{1,1} & \dots & x^{1,n} \\ 1 & x^{2,1} & x^{2,1} & \dots & x^{2,n} \\ 1 & \vdots & \ddots & & \vdots \\ 1 & x^{m,1} & x^{m,1} & \dots & x^{m,n} \end{pmatrix} \quad Y = \begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^m \end{pmatrix}$$

m observations

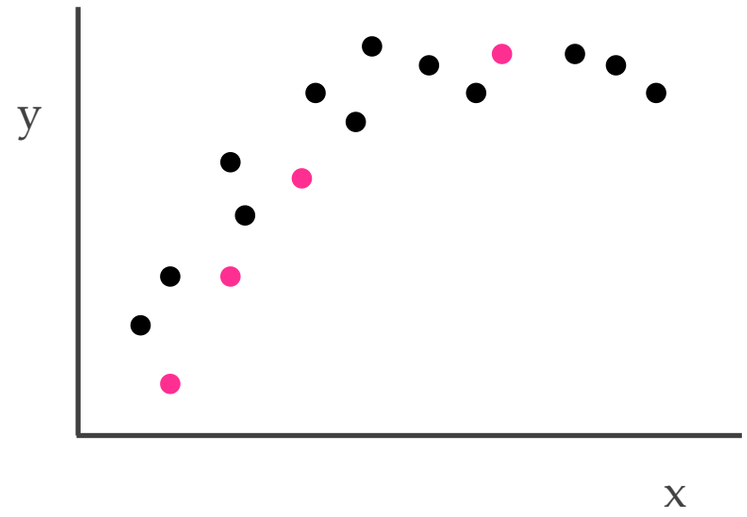
Training the model (finding parameters *w*)

Partition data

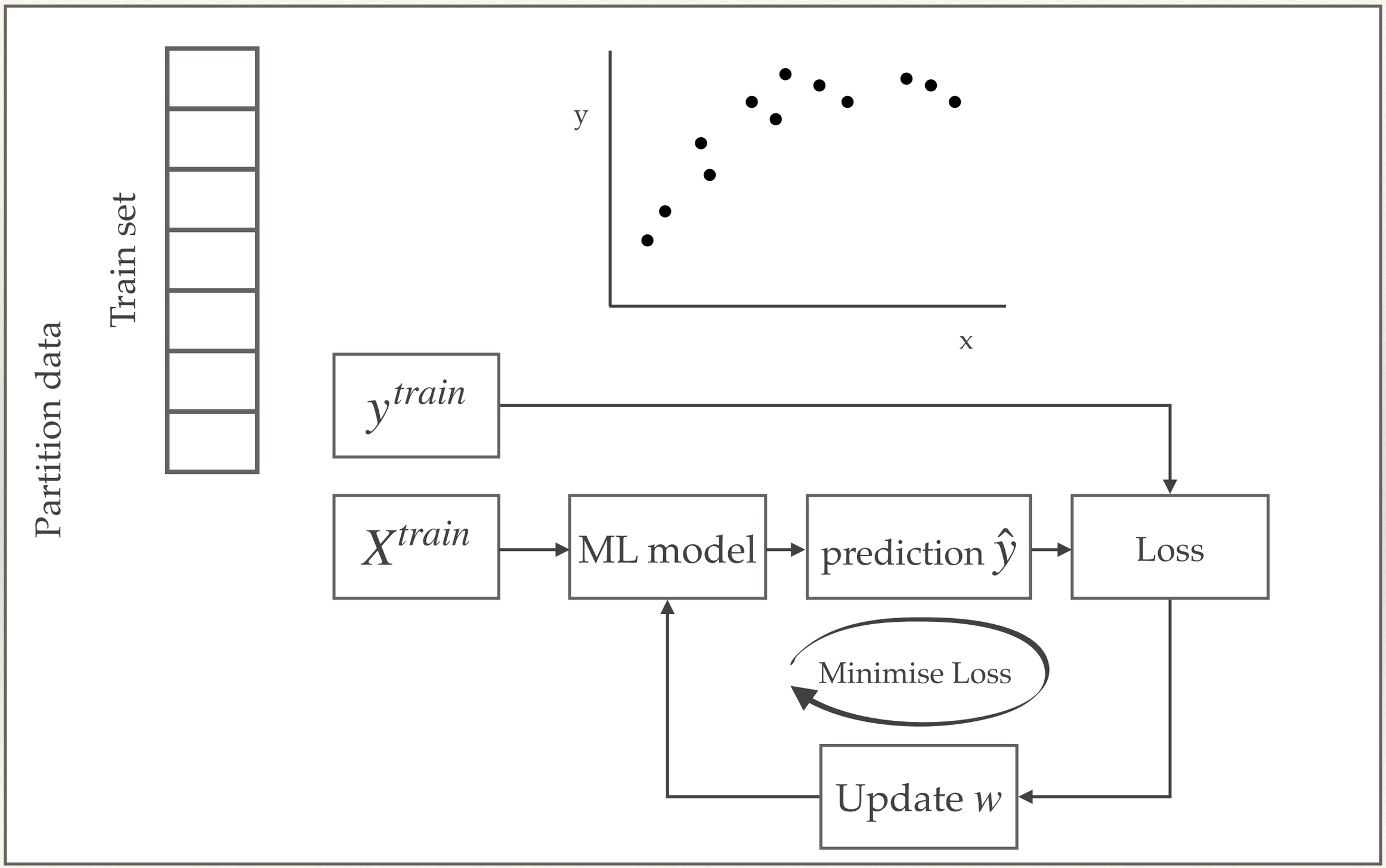
Train set

Test set

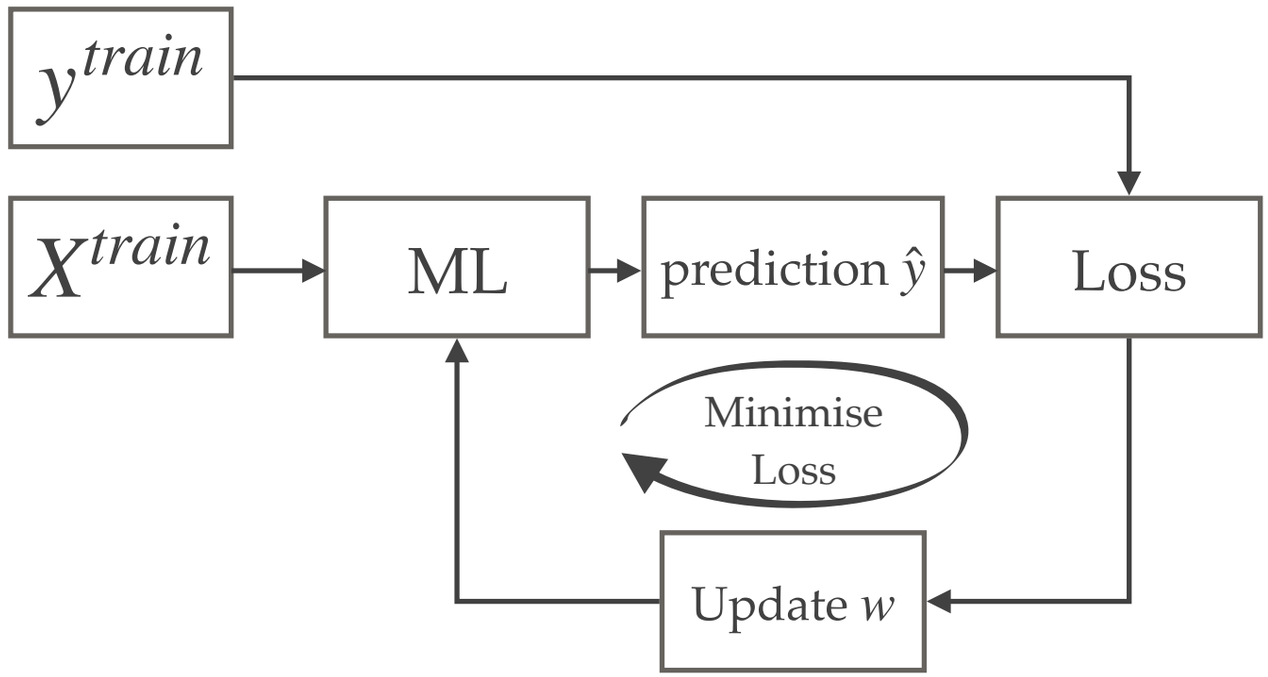
Only for model evaluation!

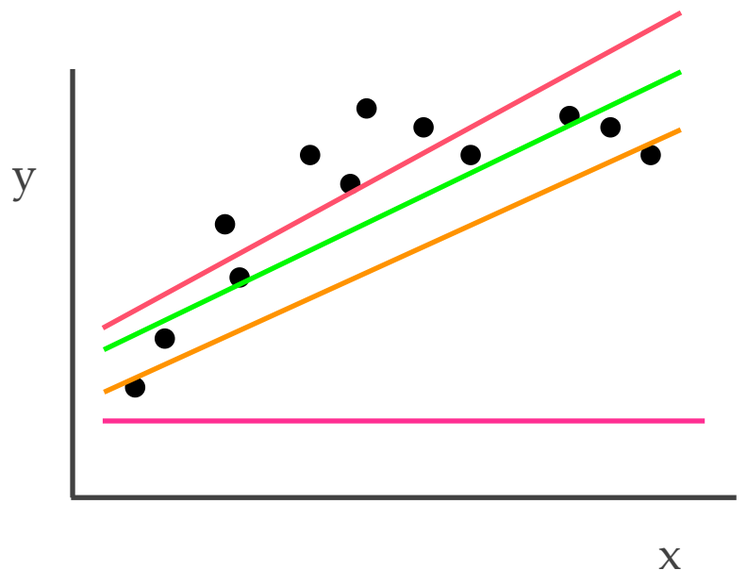


Training the model (finding parameters *w*)



Linear regression





A very simple case of 1 dimensional feature:

$$\hat{y}_{ML}(x^*) = x^* w = [1 \quad x^*] \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

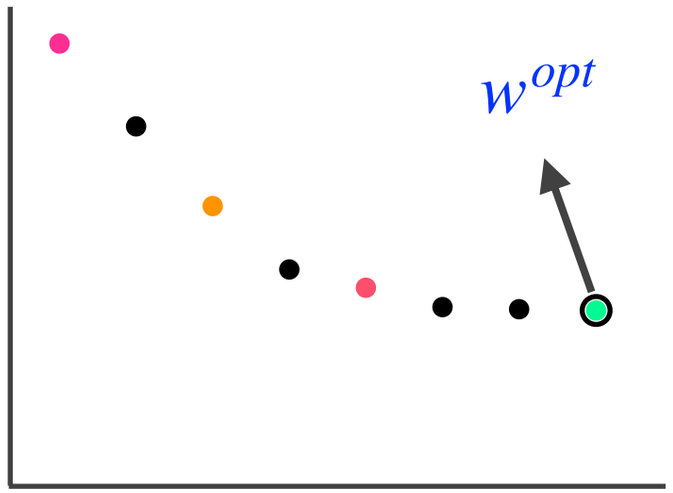
Matrix form:

$$\hat{y}_{ML}(X) = Xw = \begin{bmatrix} 1 & x^1 \\ \vdots & \vdots \\ 1 & x^m \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

Loss function: $\mathscr{L} = \|\hat{y}_{ML} - y\|_2^2 = \|Xw - y\|_2^2$

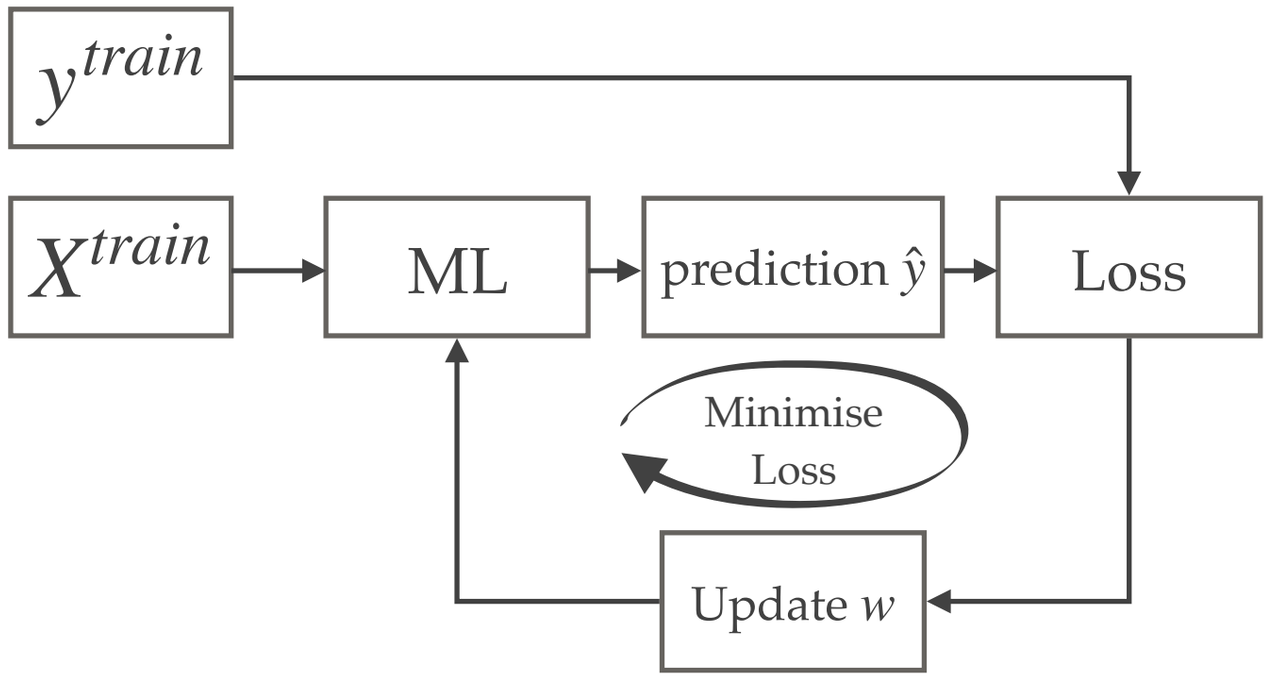
New system: $\hat{y}_{ML}(x^*) = x^* w^{opt}$

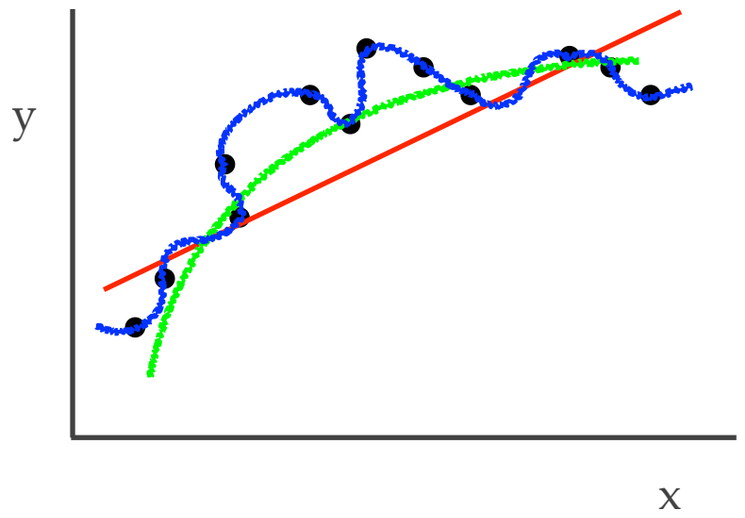
Loss



Optimisation steps

Nonlinear regression





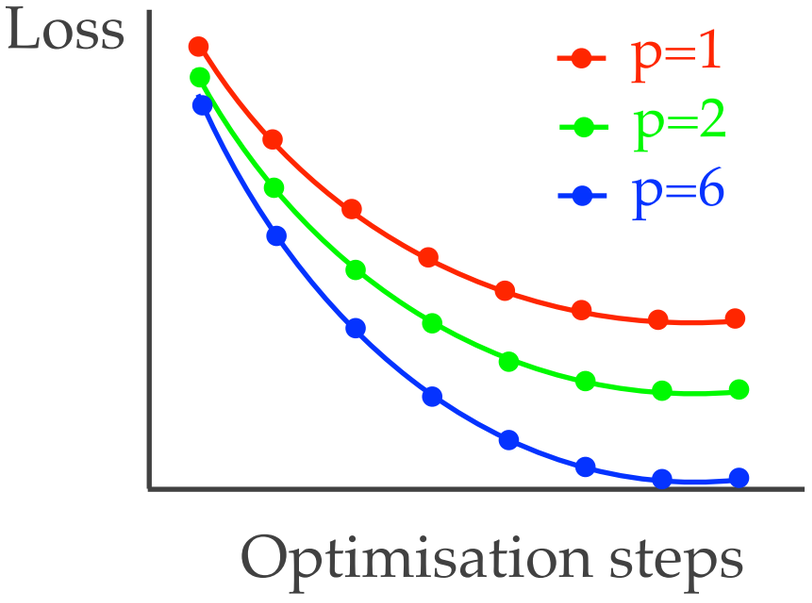
Adding nonlinear terms: polynomial order p

$$\hat{y}_{ML}(x^*) = x^* w = \begin{bmatrix} 1 & x^* & (x^*)^2 & \dots & (x^*)^p \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}$$

Matrix form:

$$\hat{y}_{ML}(X) = Xw = \begin{bmatrix} 1 & x^1 & (x^1)^2 & \dots (x^1)^p \\ \vdots & \vdots & & \ddots \quad \vdots \\ 1 & x^m & (x^m)^2 & \dots (x^m)^p \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_p \end{bmatrix}$$

Loss function: $\mathscr{L} = \|\hat{y}_{ML} - y\|_2^2 = \|Xw - y\|_2^2$



Training the model (finding parameters *w*)

1

System

$x^* \in \mathbb{R}^{n+1}$

Features describing the system

ML model

$f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$

3

Parameters

$w \in \mathbb{R}^p$

2

Properties

$y \in \mathbb{R}$

Energy, Gas uptake, …

For *m* systems

x^*, y

X, *Y*