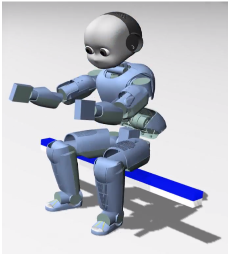
Regression

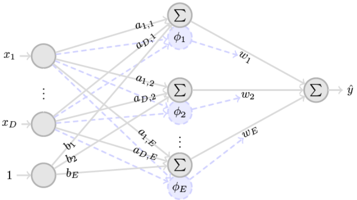
1. Introduction

Olivier Sigaud

Sorbonne Université
http://people.isir.upmc.fr/sigaud

Motivations





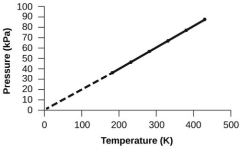
Two main motivations for this class:

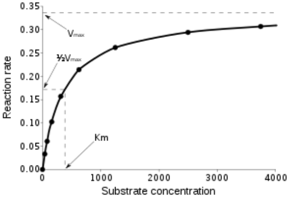
- Regression for robotics: approximating mechanical models
- Function approximation with deep neural networks, particularly deep RL

Sigaud, O., Salaün, C., & Padois, V. (2011) On-line regression algorithms for learning mechanical models of robots: a survey. *Robotics and Autonomous Systems*, 59(12):1115–1129

Regression is function approximation

Temperature (°C)	Temperature (K)	Pressure (kPa)
–150	173	36.0
–100	223	46.4
–50	273	56.7
0	323	67.1
50	373	77.5
100	423	88.0





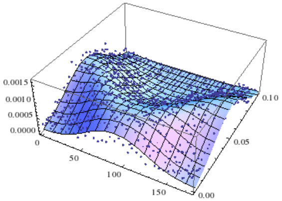
In regression:

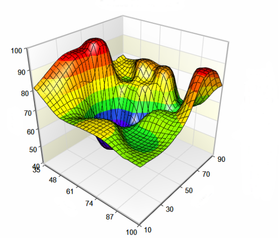
- We have a set of related measurements
- We look for a latent function that relates the corresponding variables
- We choose a model over a class of (parametric) functions
- Here an example of linear (center) and nonlinear (right) regression

Nonlinear regression

Sample Size(n)	λ	k	$BIAS$	CV
120	0 (without ridge)	30	6.12	6.01
100	1.15e-1	50	0.0030	0.0090
100	1.15e-5	20	2.52e-6	0.008
100	7.15e-5	20	1.57e-5	0.0085
100	1.15e-2	20	0.0023	0.0098
100	1.15e-3	20	2.50e-4	0.0083
100	1.15e-7	20	2.88e-8	0.009
100	8.15e-10	10	3.61e-6	0.0082
100	1.15e-3	10	0.0058	0.0096
100	7.15e-5	5	0.0179	0.0182
100	8.15e-10	5	0.0165	0.0177
120	1.15e-3	30	1.68e-5	0.0086
120	7.15e-5	30	0.0014	0.0093

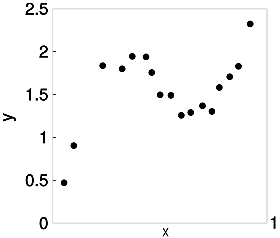
Table 2 – The simulation results for two-dimensional dataset obtained by the proposed method.





- The function between variables can be arbitrarily complicated
- It can be multidimensional

Regression: basic process and notations

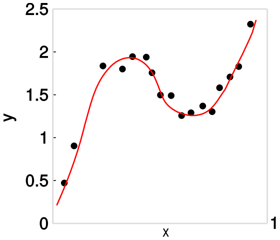


$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,D} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N,1} & \mathbf{x}_{N,2} & \cdots & \mathbf{x}_{N,D} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

(design matrix)

- The set of measurements is a batch of datapoints
- Input: N samples $\mathbf{x}_n \in \mathbb{R}^D$, $y_n \in \mathbb{R}$,
- Stored in $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$ (design matrix), $\mathbf{y} = [y_1, \cdots, y_N]$

Regression: latent function

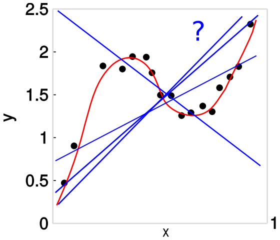


$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,D} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N,1} & \mathbf{x}_{N,2} & \cdots & \mathbf{x}_{N,D} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

(design matrix)

- The measurements correspond to an unknown function
- They can be subject to some noise
- We want to approximate the latent function (in red)

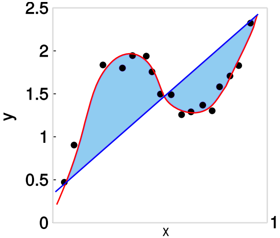
Regression: the output is a model



$$\mathbf{X} = \underbrace{\begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,D} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N,1} & \mathbf{x}_{N,2} & \cdots & \mathbf{x}_{N,D} \end{pmatrix}}_{\text{(design matrix)}} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

- We choose a parametric model over a class of functions
- Here we consider linear models (in blue)
- Output: a model $\hat{f}$ of the latent function f such that $\mathbf{y} \sim \hat{f}(\mathbf{X})$

Regression: minimizing the expectation of the error

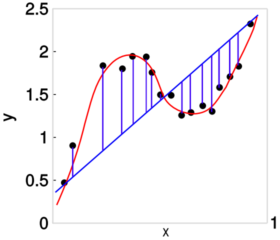


$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,D} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N,1} & \mathbf{x}_{N,2} & \cdots & \mathbf{x}_{N,D} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

(design matrix)

- We want to minimize some difference between f and $\hat{f}$
- An option would be the expectation of the error between f and $\hat{f}$ over the input space (light blue area)
- In practice, since we do not know f, we cannot compute this expectation

Regression: over a batch of samples

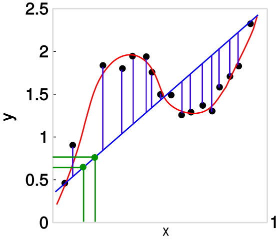


$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,D} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N,1} & \mathbf{x}_{N,2} & \cdots & \mathbf{x}_{N,D} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

(design matrix)

- We rather compute the error at the datapoints that we know
- We will formalize this in the next class

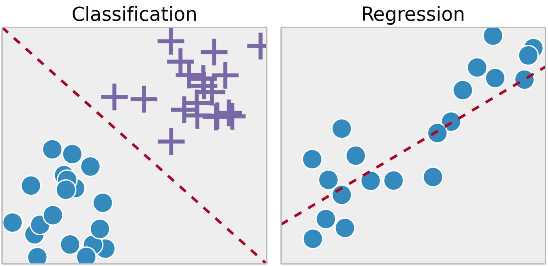
Regression: generalisation



$$\mathbf{X} = \underbrace{\begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,D} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N,1} & \mathbf{x}_{N,2} & \cdots & \mathbf{x}_{N,D} \end{pmatrix}}_{\text{(design matrix)}} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

- If the error is low, the model will generalize well to unseen data (new $\mathbf{x}$)

Regression and classification

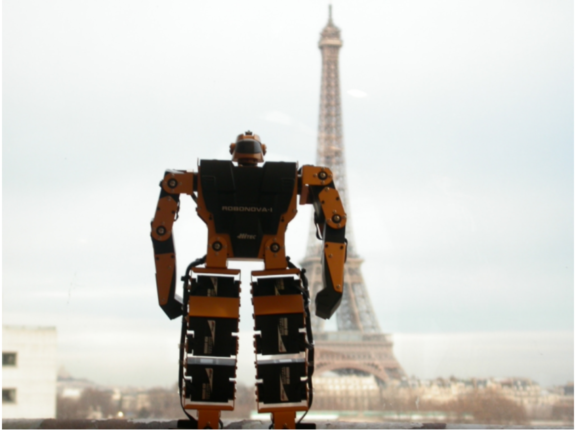


- Classification and regression are two forms of supervised learning
- In classification, the output space is a discrete set of classes, in regression it is continuous
- To perform classification, we usually rely on labelled databases (e.g. Imagenet)
- To perform regression, we rather rely on related measurements

Outline of next videos

1. Introduction (this video)
2. Linear Least Squares
3. Batch non-linear locally weighted regression
4. Batch non-linear regression through projection
5. Iterative and incremental methods
6. Gradient descent
7. Advanced gradient descent

Any question?



Send mail to: Olivier.Sigaud@upmc.fr

ISIR
INSTITUT
DES SYSTÈMES
INTELLIGENTS
ET DE ROBOTIQUE

Olivier Sigaud, Camille Salaün, and Vincent Padois.
On-line regression algorithms for learning mechanical models of robots: a survey.
Robotics and Autonomous Systems, 59(12):1115–1129, December 2011.