ML2 : The landscape of machine learning

Pierre CHAINAIS

The way it goes

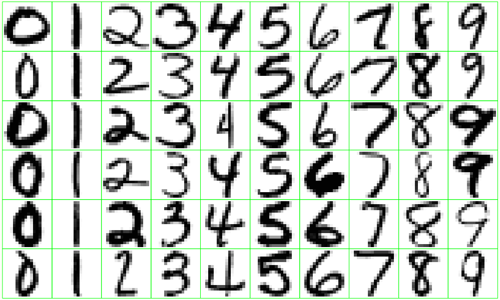
- Books :
  - **T. Hastie, R. Tibshirani et J. Friedman (2009)**
    The Elements of Statistical Learning
    Springer Series in Statistics, disponible en ligne.
  - **C. M. Bishop (2009)**
    Pattern Recognition and Machine Learning, *Springer*.
  - **K.P. Murphy (2012)**
    Machine Learning : a probabilistic perspective
    The MIT Press.
- Slides, TP... online on Moodle .
- Evaluation : TP + short exams + final test

The way it goes

1. Linear models for regression
2. Classification and theory of decision
3. Linear models for supervised classification
4. Dimension reduction
5. Unsupervised classification, clustering
6. Evaluation of performances
7. Decision trees
8. Boosting : AdaBoost (clever learning)
9. Support vector machines (kernel approaches)
10. Neural networks (deep learning)

Examples

Character recognition



- to identify the numbers on images (0,..., 9) from a 16x16 gray level image (0 to 255).
- **Supervised classification**

SPAM

```
WINNING NOTIFICATION! We are
pleased to inform you of the
result of the Lottery Winners
International Program held on
the 4th november 2013.
You have been approved for a
lump sum pay out of 175,000
euros.
CONGRATULATIONS!!!
```

NON SPAM

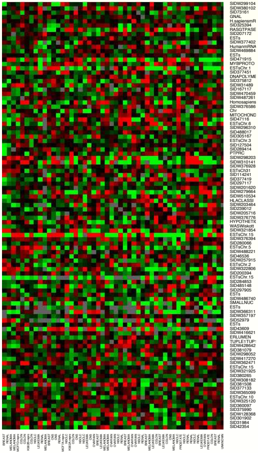
```
Dear Pascal,

Could you please send me the
report #1234 on the project
advancement?

Thanks in advance.


Regards, Clara
```

- Basis of reference mails, identified as SPAM or NOT
- Purpose : predict whether a new mail is SPAM or NOT
- Avoid to delete important messages! (false alarm)
- **Supervised classification**

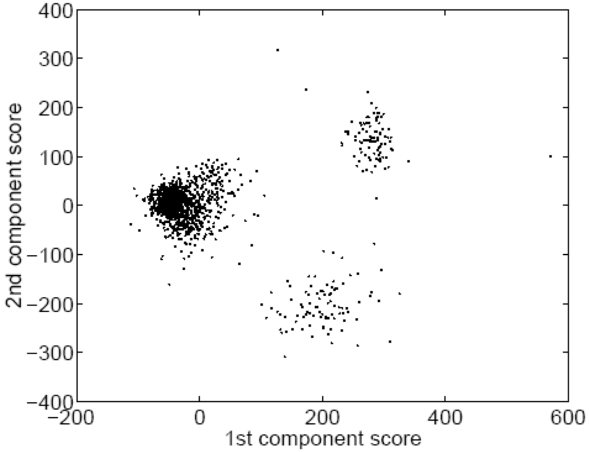


64 sequences (individuals) of 6830 genetical responses to a reference

- Which are the groups of similar samples ? (similarities and links between individuals)
- Which are the genes with similar expressions ? (similarities and links between genes)
- Are there genes which are more characteristic than others ? (extraction of characteristics)
- **Clustering**

Examples

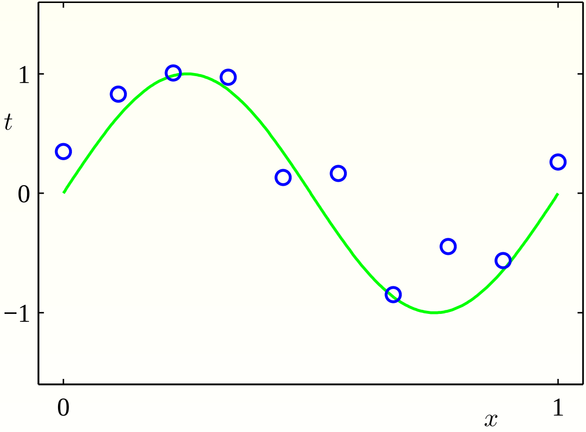
Clouds of points : identification of groups (clusters)



clustering : unsupervised classification

Examples

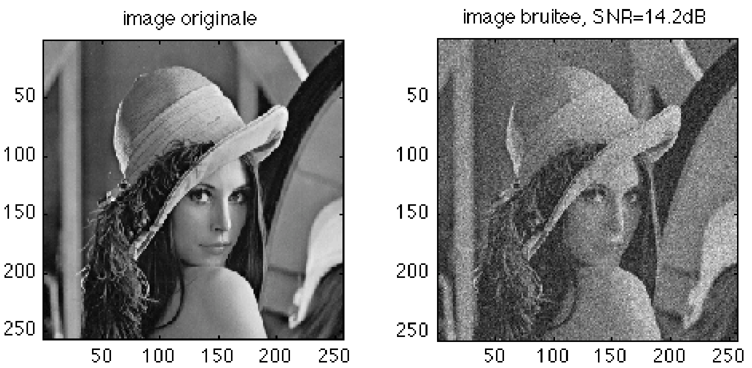
Approximation of functions, modelling, denoising, inpainting...



Regression

Examples

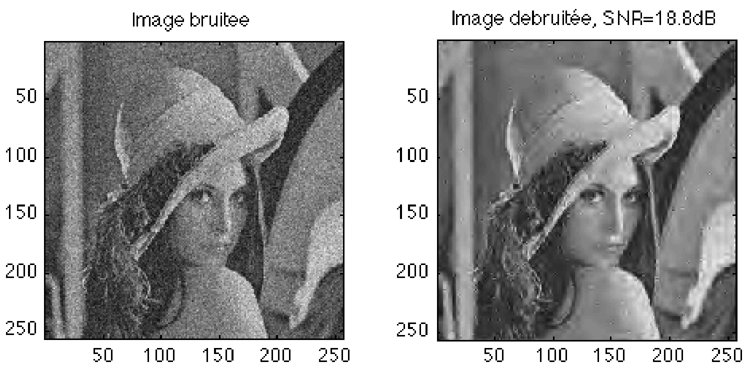
Approximation of functions, modelling, denoising, inpainting...



Regression

Examples

Approximation of functions, modelling, denoising, inpainting...



Regression

Examples

Approximation of functions, modelling, denoising, inpainting...

Regression

Examples

Approximation of functions, modelling, denoising, inpainting...

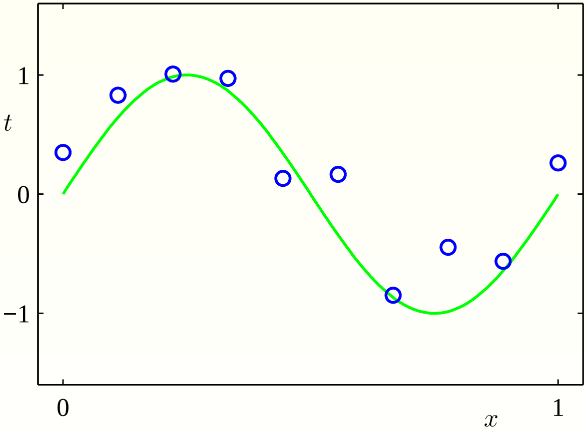
Regression

ML2 : The landscape of machine learning
1. Linear models for regression

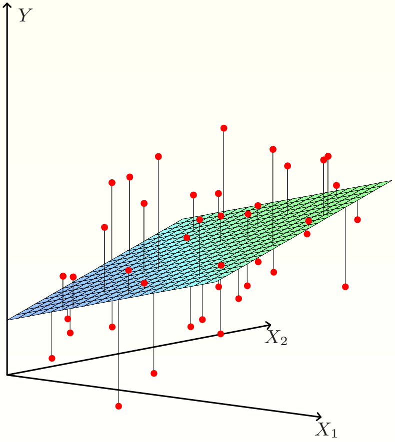
Pierre CHAINAIS

Trajectory

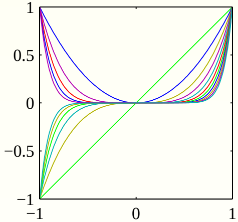
Regression : least square estimation



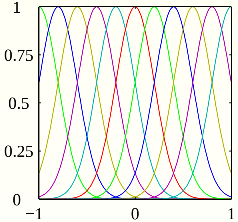
Regression : least square estimation



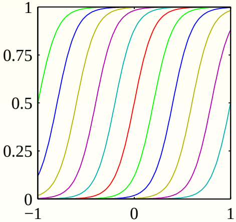
Examples of bases of functions



polynomials



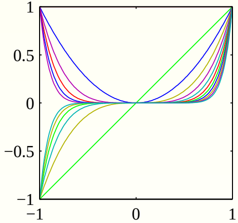
gaussian kernels



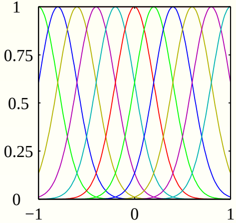
sigmoids

Term	Coefficient	Std. Error	Z Score
`Intercept`	2.46	0.09	27.60
`lcavol`	0.68	0.13	5.37
`lweight`	0.26	0.10	2.75
`age`	−0.14	0.10	−1.40
`lbph`	0.21	0.10	2.06
`svi`	0.31	0.12	2.47
`lcp`	−0.29	0.15	−1.87
`gleason`	−0.02	0.15	−0.15
`pgg45`	0.27	0.15	1.74

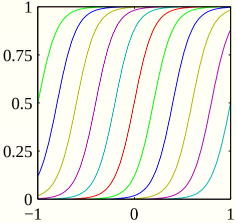
Examples of bases of functions



polynomials

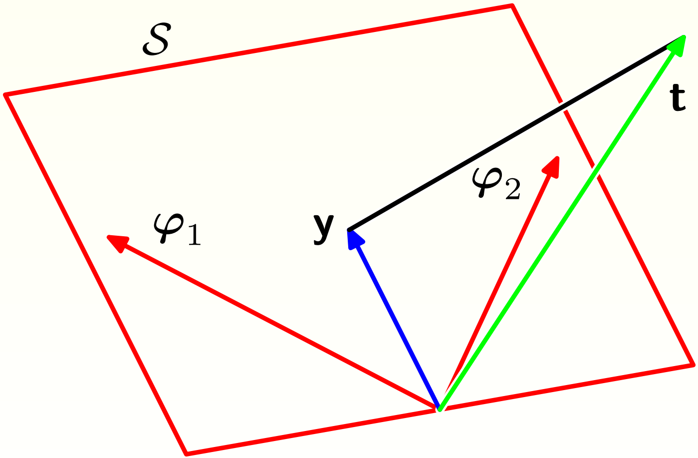


gaussian kernels



sigmoids

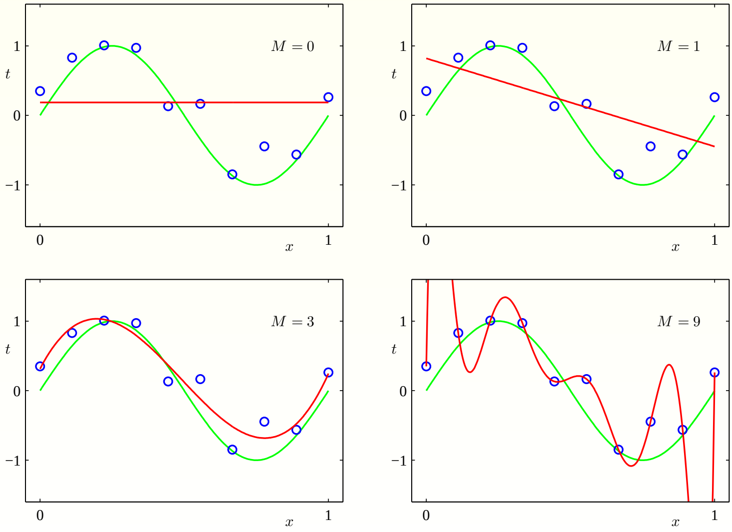
and also local cosines, wavelets...



$\mathcal{S}$
$\mathbf{t}$
φ_2
$\mathbf{y}$
φ_1

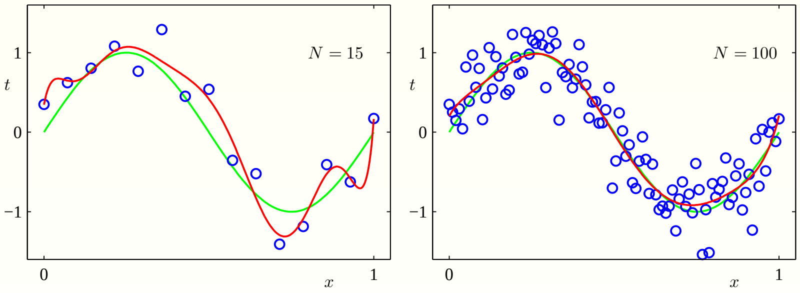
Regularized least squares

Polynomial approximation : importance of the order

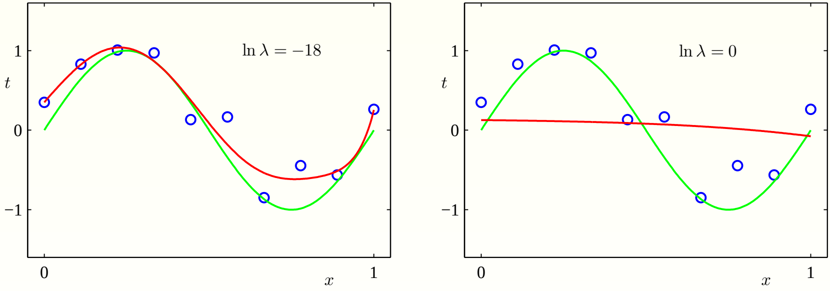


Regularized least squares

Polynomial approximation : importance of the order



Polynomial approximation + regularization : minimizes higher order coeff.



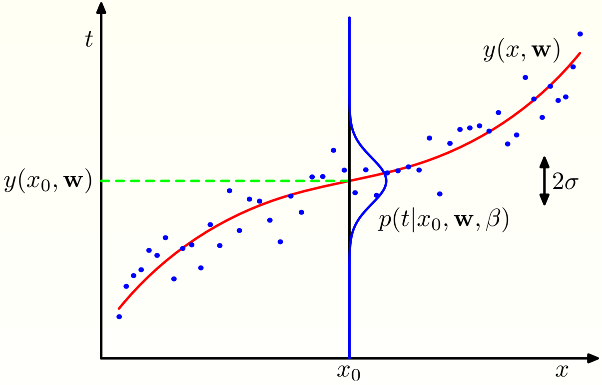
average reg. strong reg.

Regularized least squares

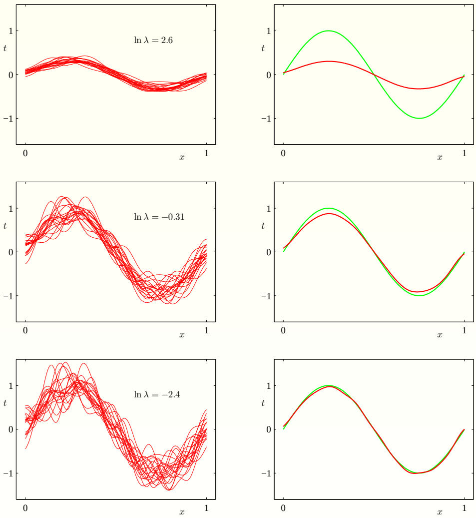
Example : prediction of the level of prostate specific antigen

Term	LS	Best Subset	Ridge	Lasso	PCR	PLS
`Intercept`	2.465	2.477	2.452	2.468	2.497	2.452
`lcavol`	0.680	0.740	0.420	0.533	0.543	0.419
`lweight`	0.263	0.316	0.238	0.169	0.289	0.344
`age`	−0.141		−0.046		−0.152	−0.026
`lbph`	0.210		0.162	0.002	0.214	0.220
`svi`	0.305		0.227	0.094	0.315	0.243
`lcp`	−0.288		0.000		−0.051	0.079
`gleason`	−0.021		0.040		0.232	0.011
`pgg45`	0.267		0.133		−0.056	0.084
Test Error	0.521	0.492	0.492	0.479	0.449	0.528
Std Error	0.179	0.143	0.165	0.164	0.105	0.152

Bayesian approach



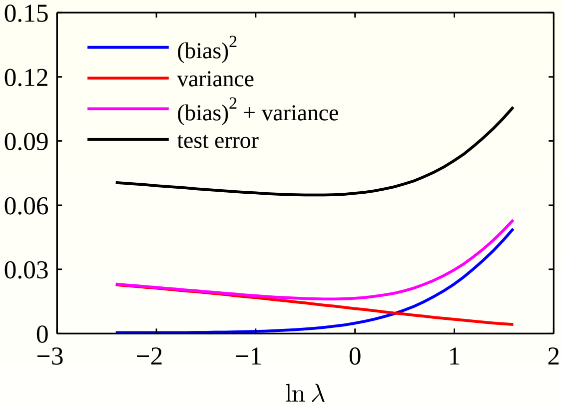
The bias-variance compromise



different data averaged estimates

M = 25 Gaussian kernels, N = 25 points,

ridge regression (λ)



M = 25 Gaussian kernels, N = 25 points, ridge regression (λ)