



Spectroscopic Data in R and Validation of Soft Classifiers: Classifying Cells and Tissues by Raman Spectroscopy

Claudia Beleites^{1,2} (Claudia.Beleites@ipht-jena.de),
Christoph Krafft², Jürgen Popp^{2,3}, and Valter Sergo¹

¹ CENMAT and Dept. of Industrial and Information Engineering,
University of Trieste, Trieste/Italy

² Institute of Photonic Technology, Jena/Germany

³ Abbe Center of Photonics, University Jena/Germany



UseR! 2011

Gliomas

- Most common primary brain tumors
- Astrocytomas most frequent subgroup
- WHO grades:

(Normal)



Astro. °II



Astro. °III

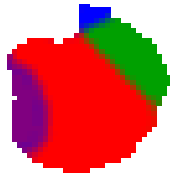
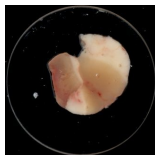
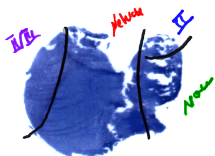


Glioblastoma (°IV)



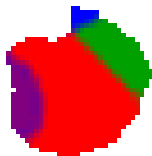
[wikipedia:Astrozytom.jpg]

Classification of Tumour Tissues



- De-differentiate over time
 - Mixture of tumour grades
 - 37 % of tumour sections mainly tissue between grades
- Are polymorphous / heterogeneous:
 - One tumour has different cell populations
 - Infiltrative growth
 - Areas with mixtures of cells

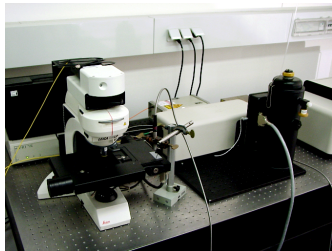
Soft Classification



- class membership as fraction of 0 - 100%
- interpretation:
 - mixture
 - probability

- soft prediction: very common
- soft reference: less common, but available
- soft test: topic of this talk

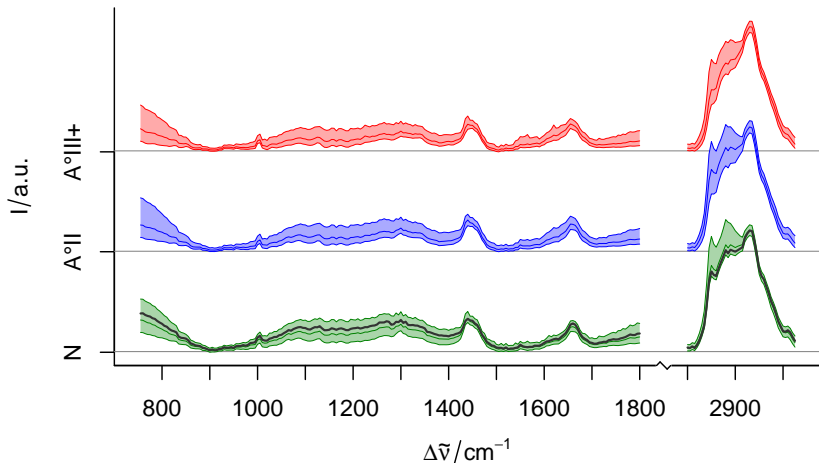
Raman Spectroscopy



- Molecular vibrations:
atoms oscillate against each other
- Vibration has particular energetic level
- Characteristic frequencies
 - ⇒ **biochemical composition**
 - ⇒ **(lipids, proteins, carbohydrates, ...)**
- Fingerprint region
 - ⇒ **identification of cell/tissue type**



Spectra

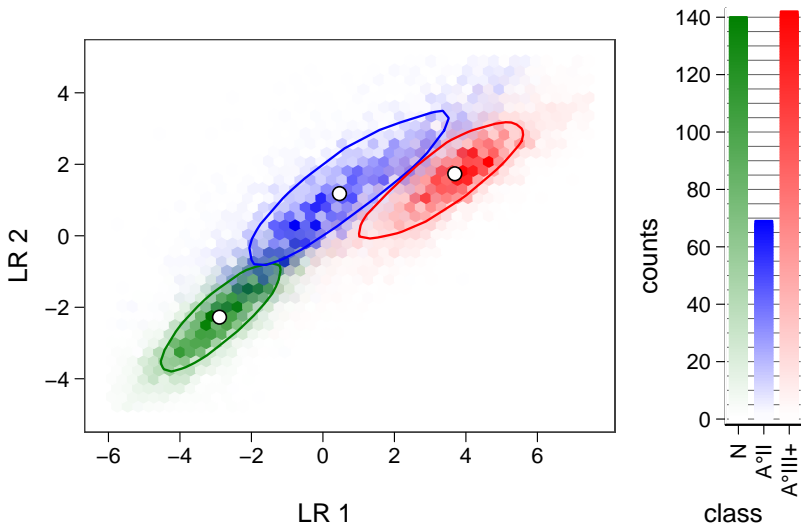


weighted mean ± 1 standard deviation spectra

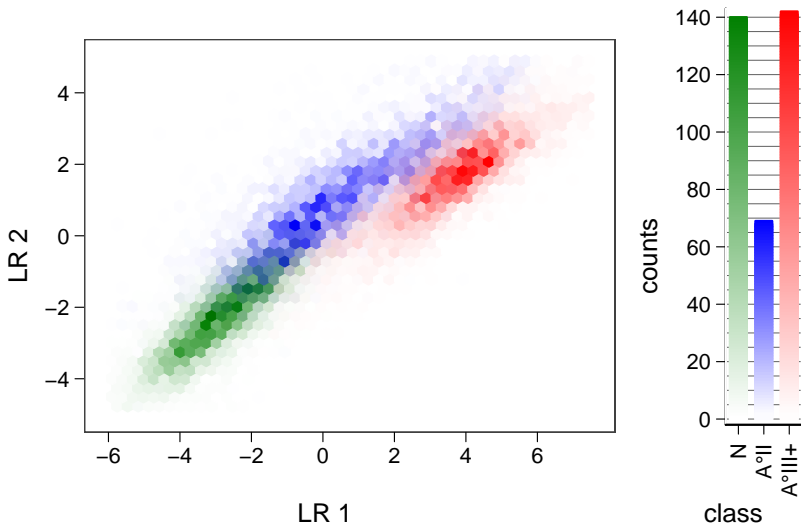
Data Set Composition

class	crisp reference		soft reference	
	patients	spectra	patients	spectra
Normal	16	7 456	35	15 747
thereof controls	9	4 902	9	4 902
Astrocytoma °II	17	4 171	47	19 128
Astrocytoma °III+	27	8 279	53	21 617
total	53	19 906	80	37 015

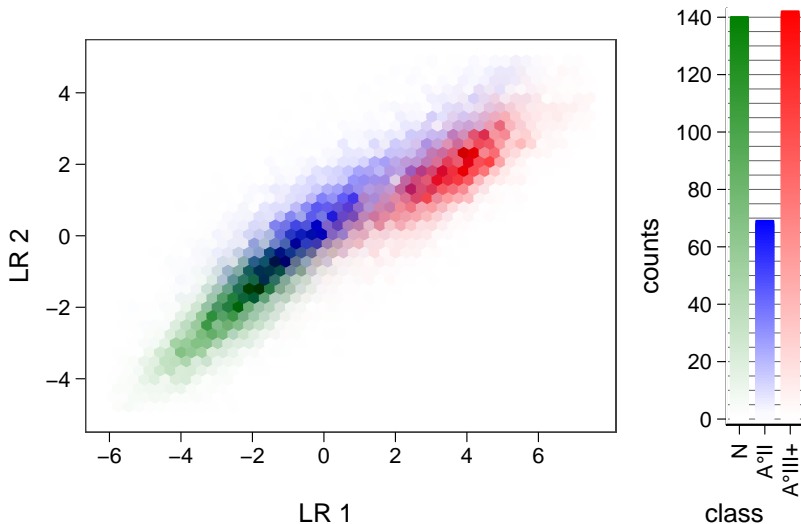
LR Projection



LR Projection



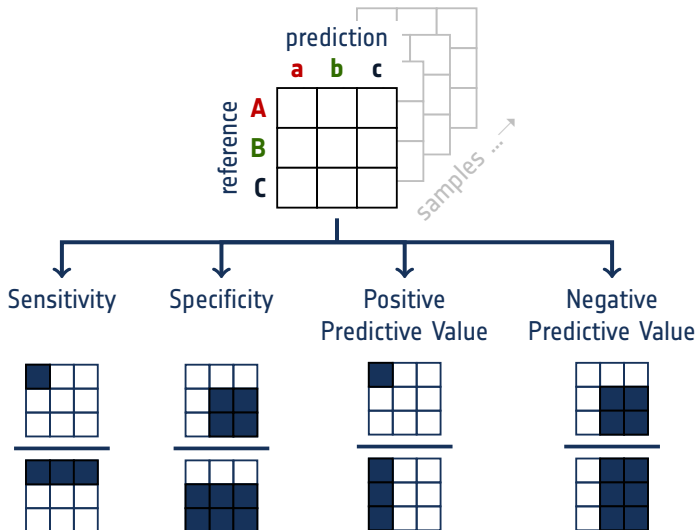
LR Projection



Model setup

- **No data-driven optimization**
- Intensity calibration
- Baseline correction (linear + quadratic)
- Normalization: area 2900 – 3025 cm^{-1}
- “Centering”: subtract mean spectrum of normal gray matter
- Classification: Logistic regression (`nnet::multinom`)
- 125×8 -fold cross validation
- Splitting patient-wise
spectra of one patient are not statistically independent

Classifier Performance Measures



“Classical” Confusion Matrix

		Prediction \hat{Y}							
		1		0		0.6		0.4	
Reference Y	1	1	0	0	1	<i>harden prediction</i>		<i>harden prediction</i>	
	0	0	0	0	0	<i>not defined</i>		<i>not defined</i>	
	0.6	<i>not defined</i>		<i>not defined</i>		<i>not defined</i>		<i>not defined</i>	
	0.4	<i>not defined</i>		<i>not defined</i>		<i>not defined</i>		<i>not defined</i>	

$$Z_{i,j} = \begin{cases} 1 & \text{if } Y_i = \hat{Y}_j = 1 \\ 0 & \text{else} \end{cases}$$

Continuous AND Operators

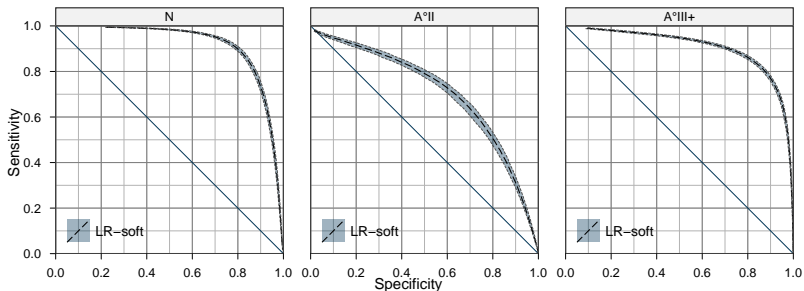
weak AND: $\min(\text{ref}_i, \text{pred}_j)$
highest possible overlap \leadsto best case performance

product: $\text{ref}_i \cdot \text{pred}_j$ expected overlap for uniform mixture \leadsto
expected performance

strong AND: $\max(\text{ref}_i + \text{pred}_j - 1, 0)$
lowest possible overlap \leadsto worst case performance

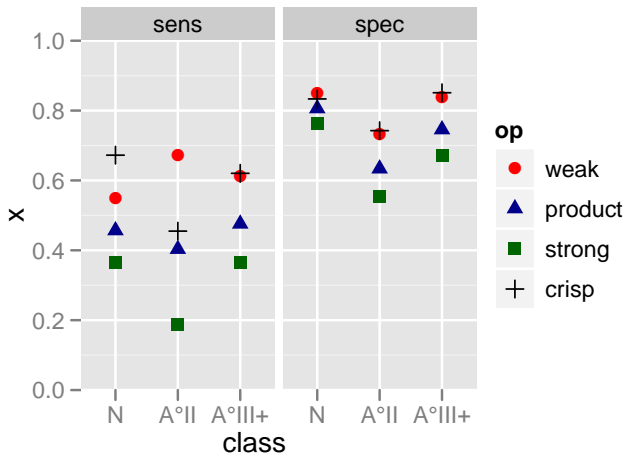
- crisp pred and ref : all coincide with “classical” AND
- strong- and product-AND: performance $< 100\%$ for $\text{pred} == \text{ref}$
 \leadsto use difference to performance for $\text{pred} == \text{ref}$
 \leadsto product-AND: \leadsto weighted MAE and RMSE

Results for Astrocytoma Grading

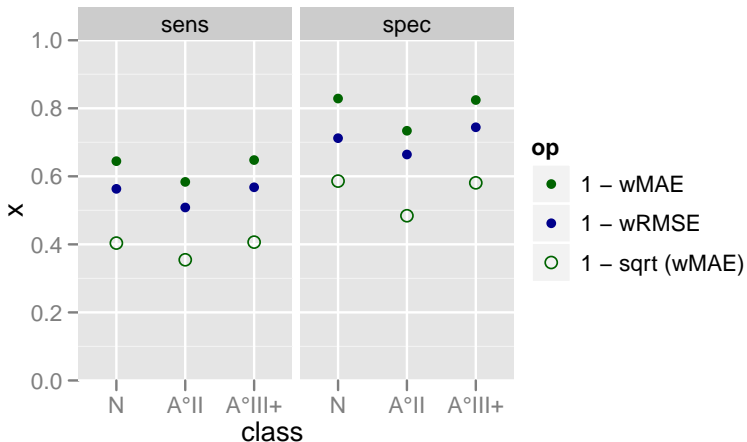


median and inter quartile range over 125 iterations crossval

Results for Astrocytoma Grading



Results for Astrocytoma Grading



Acknowledgements



A.B.C. Burlo

- Homepage:
softclassval.r-forge.r-project.org
hyperSpec.r-forge.r-project.org
- Contact:
Claudia.Beleites@ipht-jena.de
- Installation:
`install.packages ("softclassval")`