Calculating Sensitivity and Specificity of Soft Classification Models

C. Beleites[a,b], K. Geiger[c], G. Schackert[c], R. Salzer[b], and V. Sergo[a]

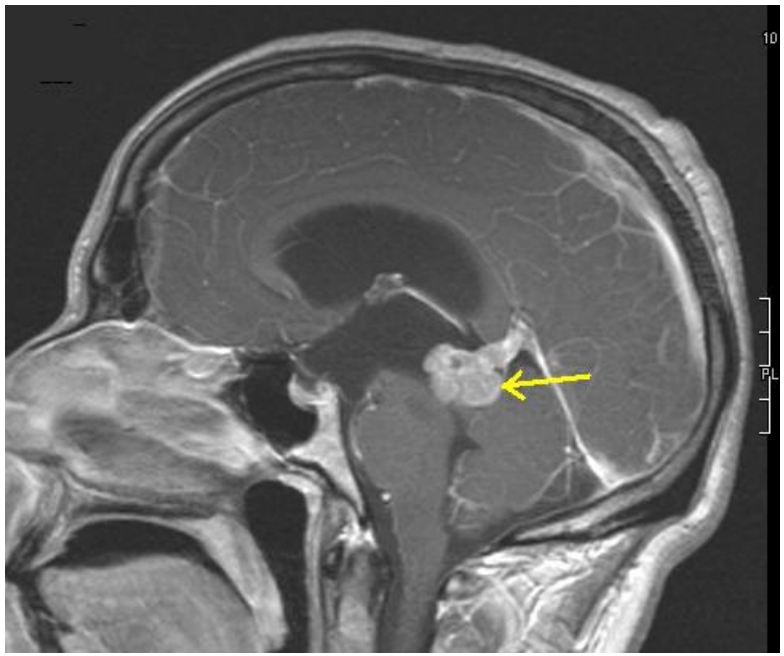
[a]CENMAT & DMRN, Università degli Studi di Trieste/Italy
[b]Analytical Chemistry, Dresden University of Technology, Dresden/Germany
[c] University Hospital, Dresden University of Technology, Dresden/Germany

Afrodata 2010, Rabat
September 24th, 2010

Outline

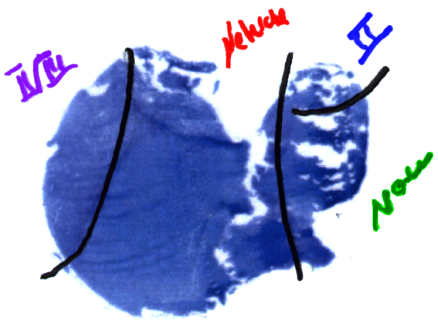
TECHNISCHE UNIVERSITÄT DRESDEN

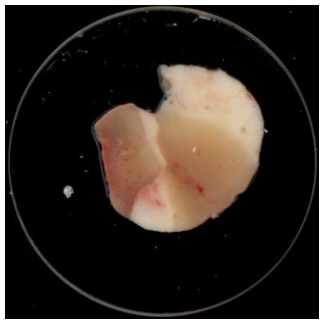
Gliomas

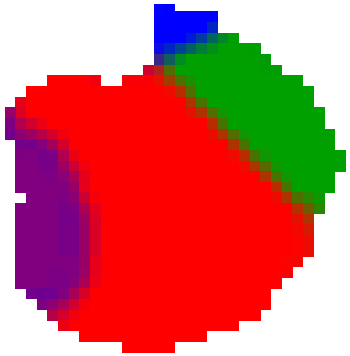


- Most common primary brain tumors
- Astrocytomas most frequent subgroup
- Astro. °II → Astro. °III → Glioblastoma (°IV)

Tumour Tissues

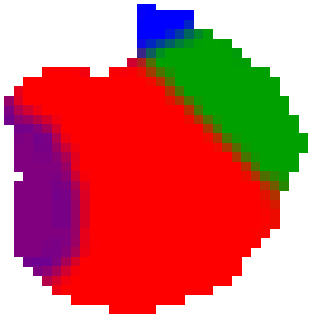






- De-differentiate
  - Mixture of tumour grades
  - 37 % of tumour sections mainly tissue between grades
- Polymorphous / Heterogeneous:
  - One tumour has different cell populations
  - Infiltrative growth
  - Areas with mixtures of cells

Soft Classification



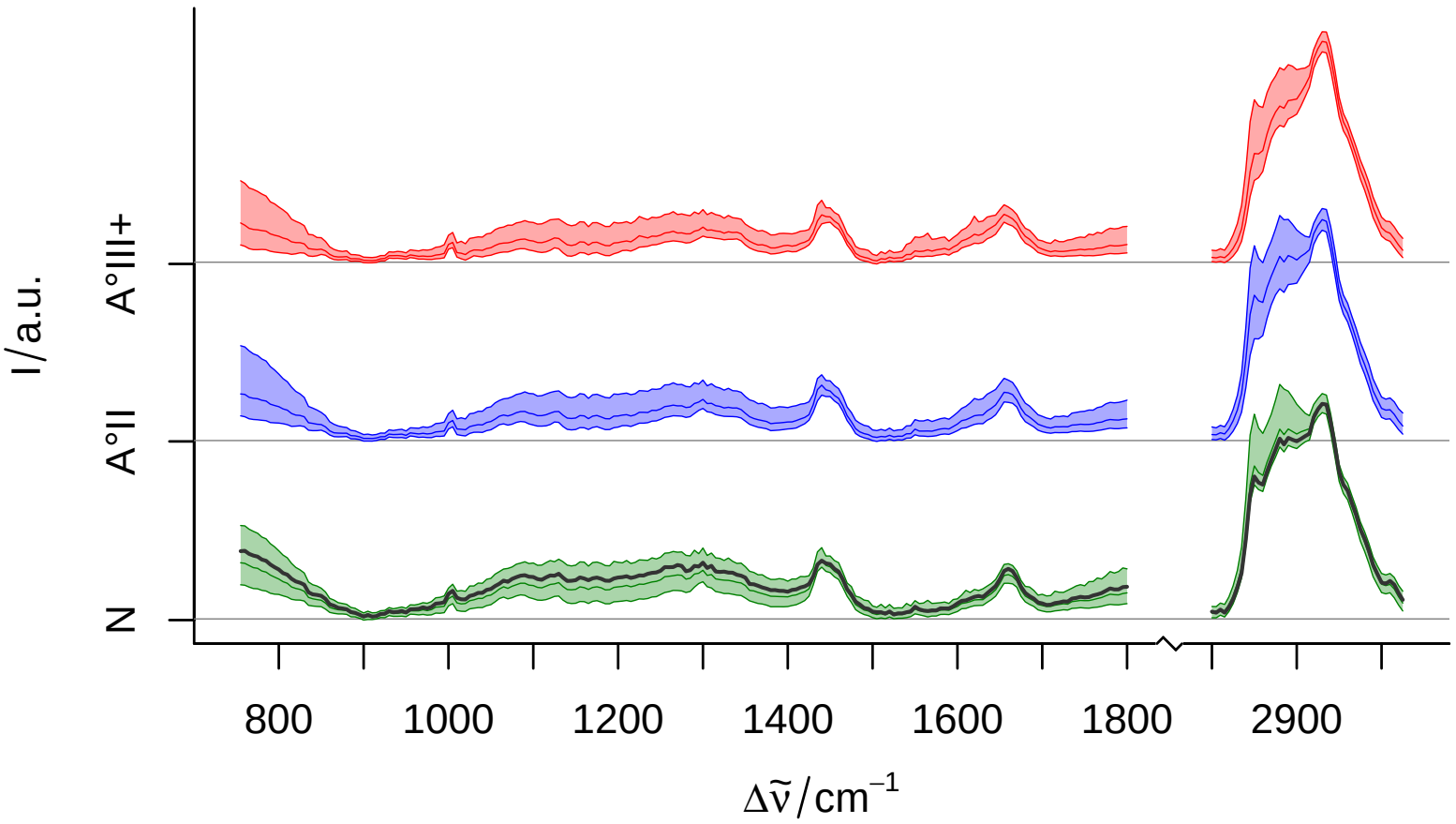
- class membership as fraction of 0 - 100%
- interpretation:
  - mixture
  - probability

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

Data Set Composition

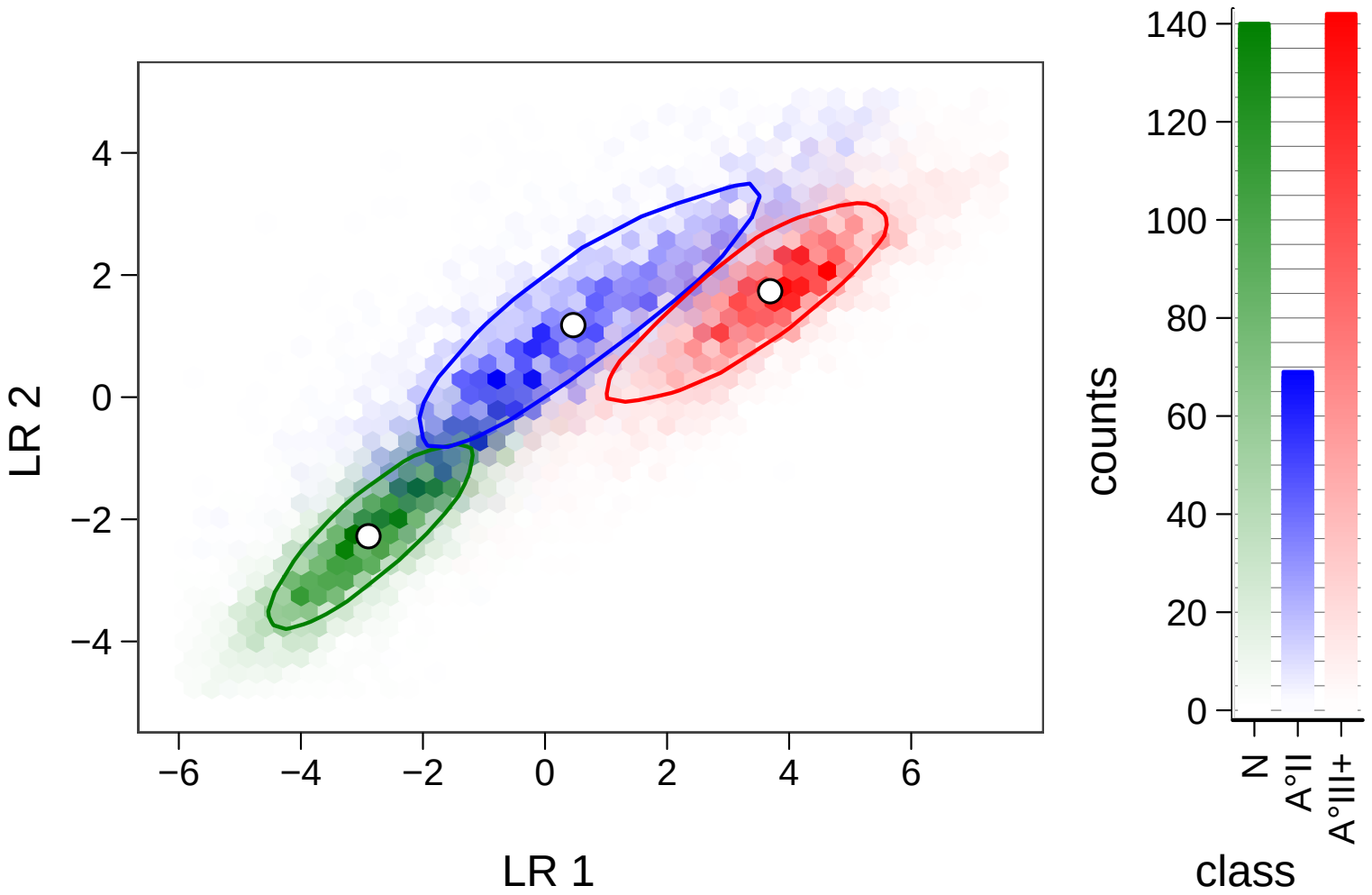
class	crisp reference patients	crisp reference spectra	soft reference patients	soft reference spectra
Normal	16	7 456	35	15 747
thereof controls	9	4 902	9	4 902
Astrocytoma °II	17	4 171	47	19 128
Astrocytoma °III+	27	8 279	53	21 617
total	53	19 906	80	37 015

Spectra



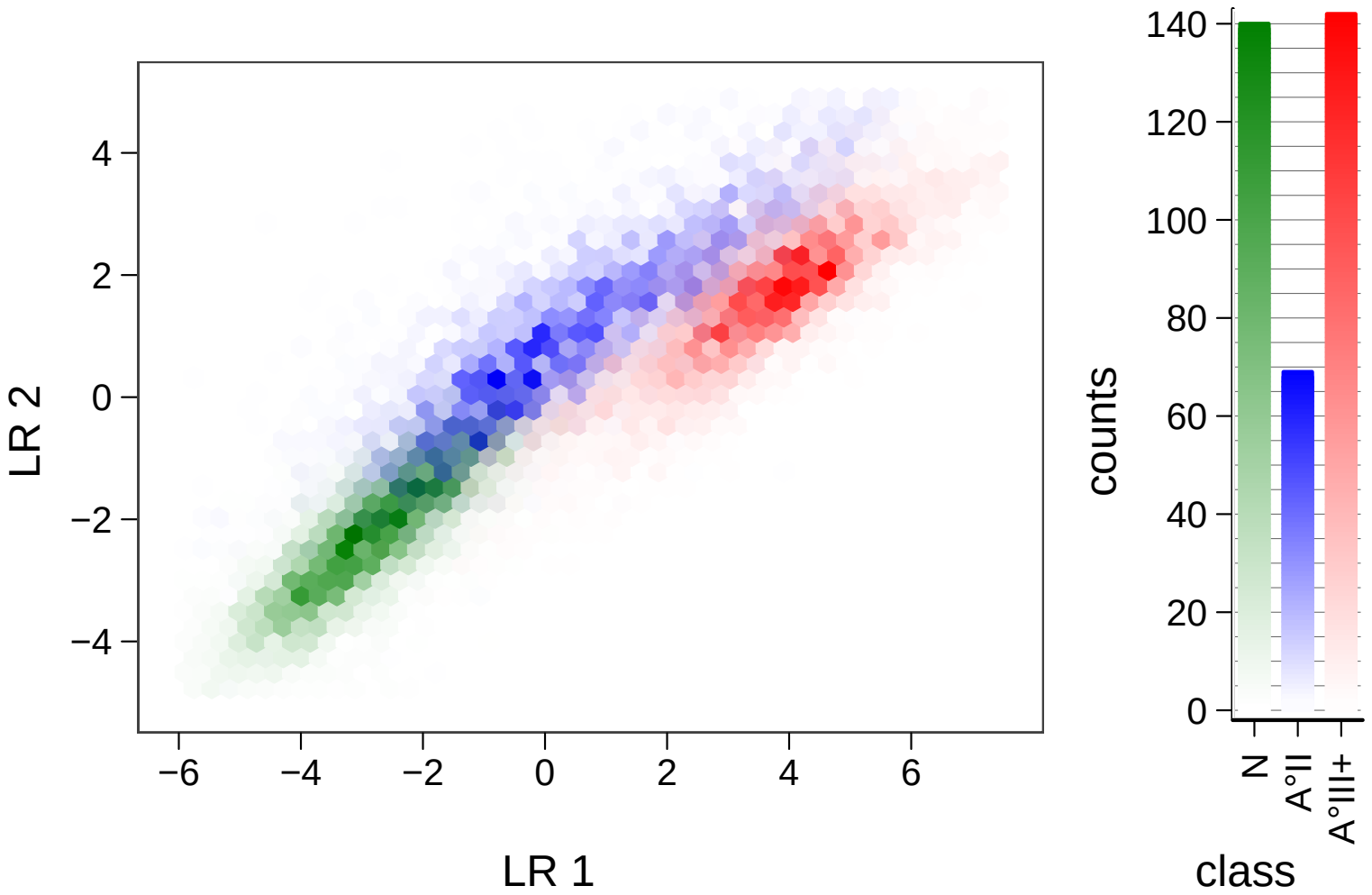
weighted median, 16$^{\text{th}}$ and 84$^{\text{th}}$ percentile spectra

LR Projection



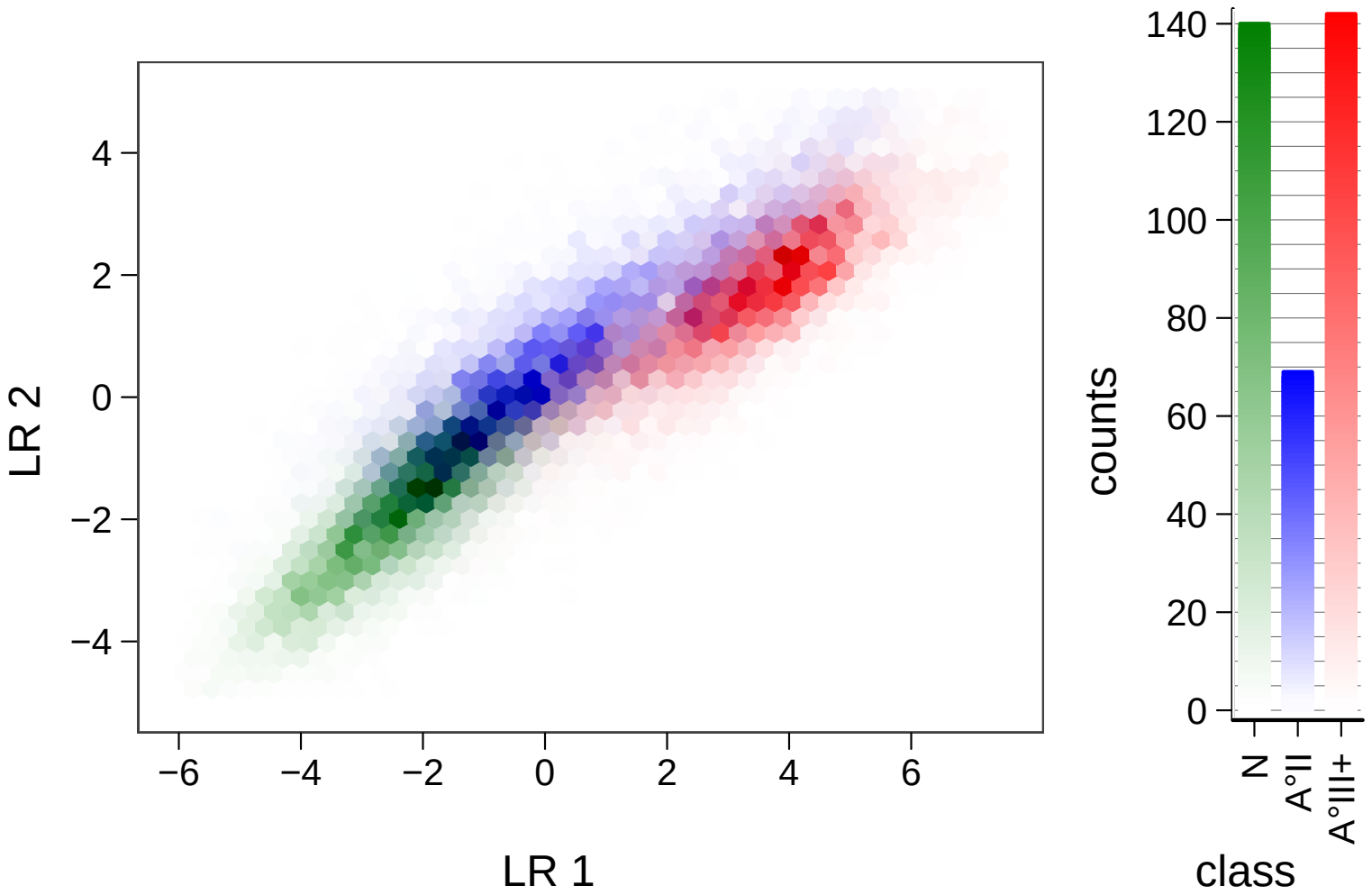
Contour contains 50 % of spectra, dot “2d median”

LR Projection



Contour contains 50 % of spectra, dot “2d median”

LR Projection



Contour contains 50 % of spectra, dot “2d median”

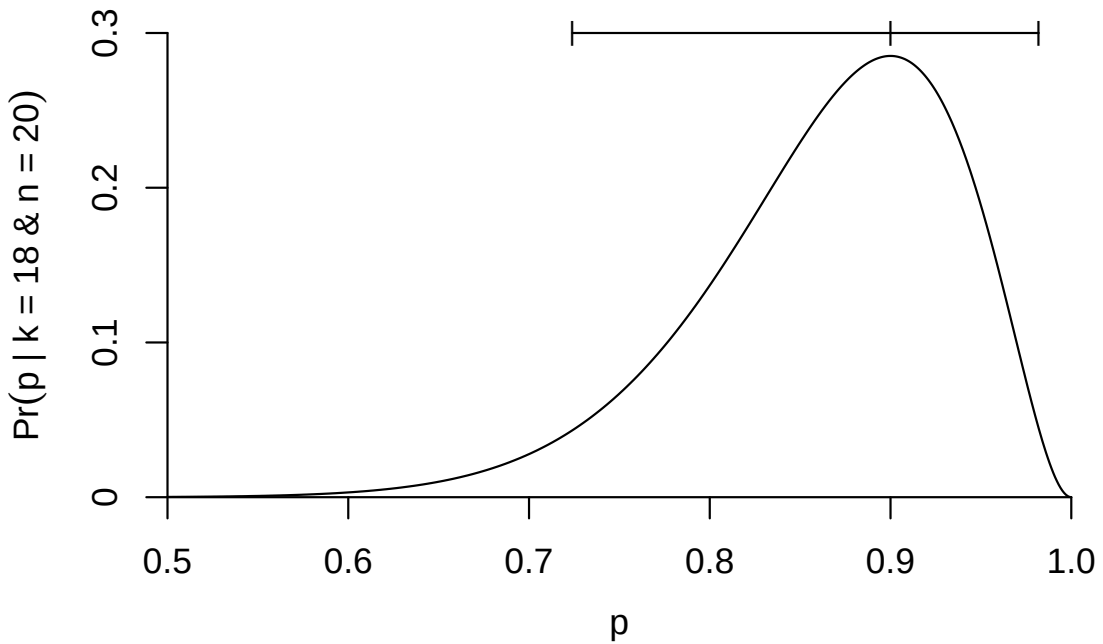
Model setup

Not optimized, **no data-driven optimization**

- intensity calibration
- baseline correction (linear + quadratic)
- normalization: area 2900 – 3025 cm^{-1}
- "centering": substract mean spectrum of normal gray matter
- Classification: Logistic regression

Optimization of Classification Models

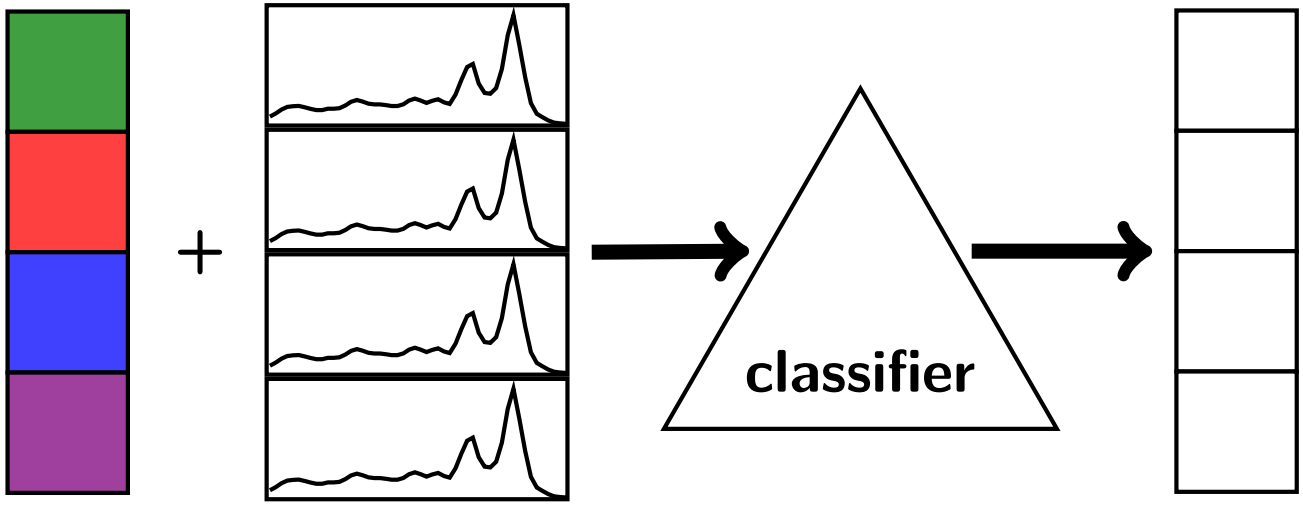
- Comparison of models: statistical test
- Sensitivity & Co.: high variance $\sigma^2(p) = \frac{1}{n} p(1-p)$
- Observation: 18 correct predictions of 20 test samples
  95 % confidence interval for true sensitivity: 0.72 – 0.98



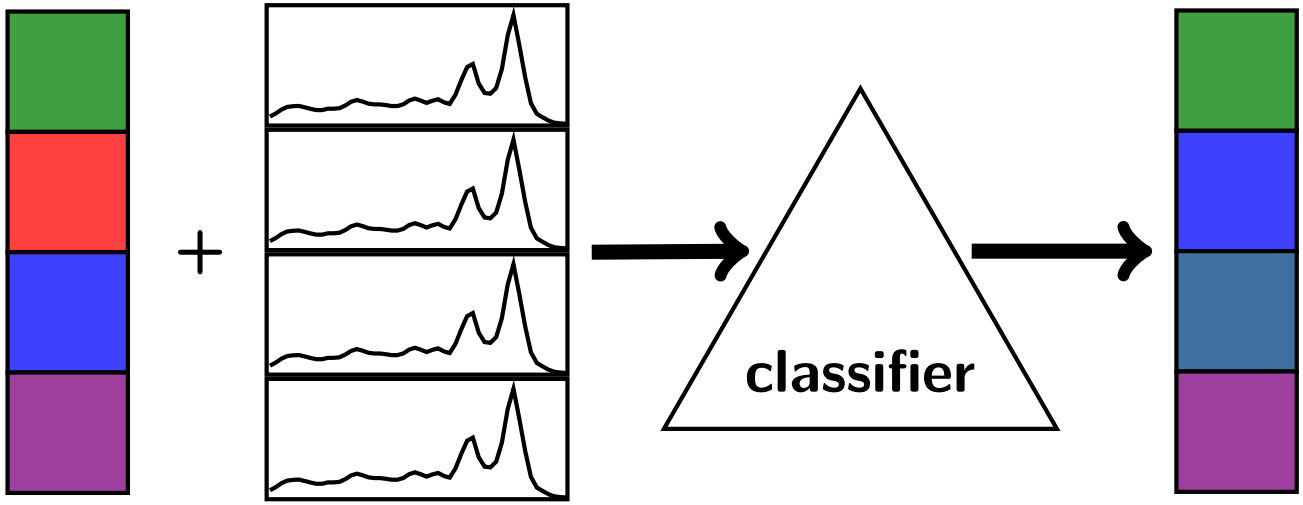
Validation Setup

- 125× 8-fold cross validation
- splitting patient-wise
  spectra of one patient are not statistically independent
- No outer loop
  - Characterization of these models possible
  - Decisions needed from surgeons

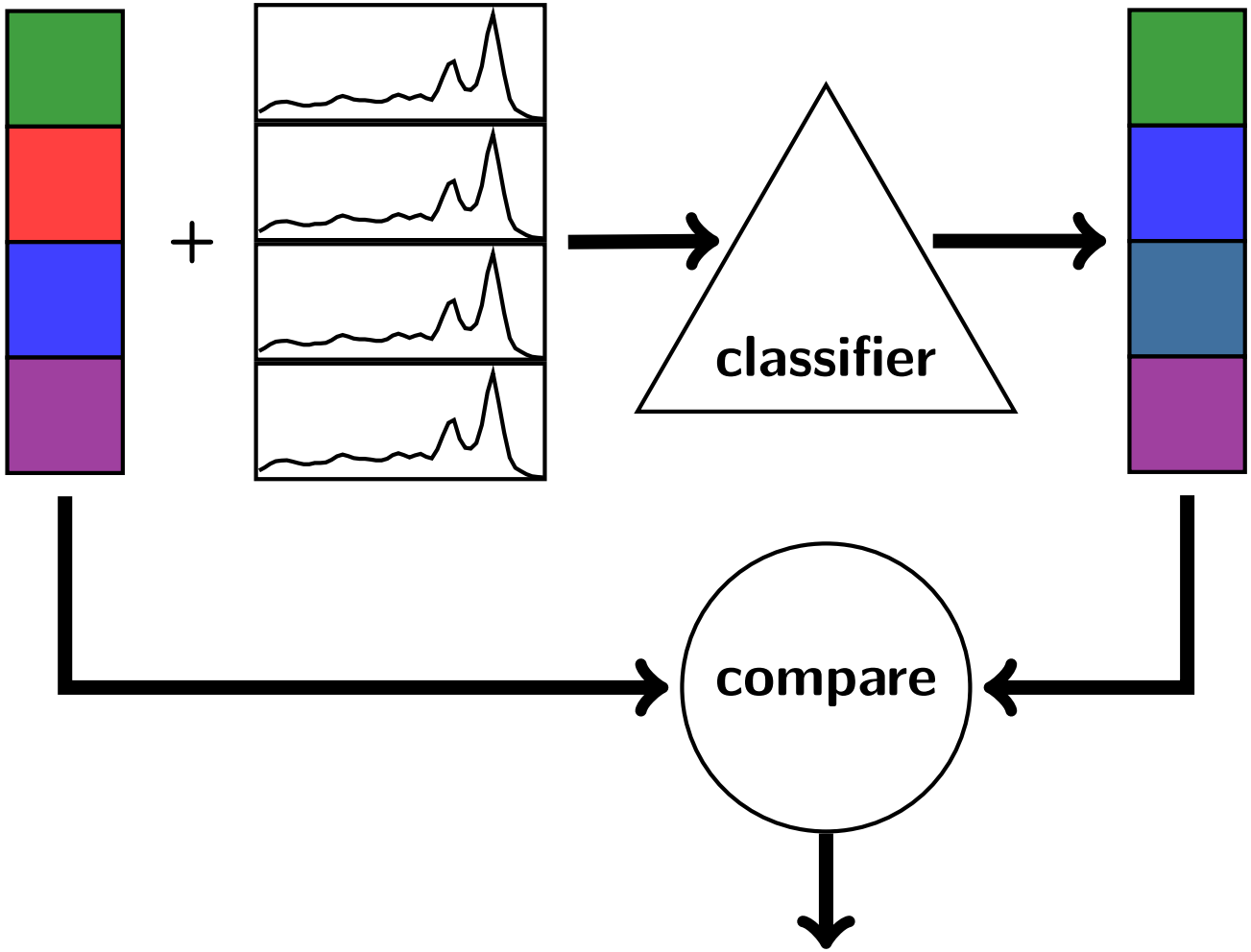
Validation



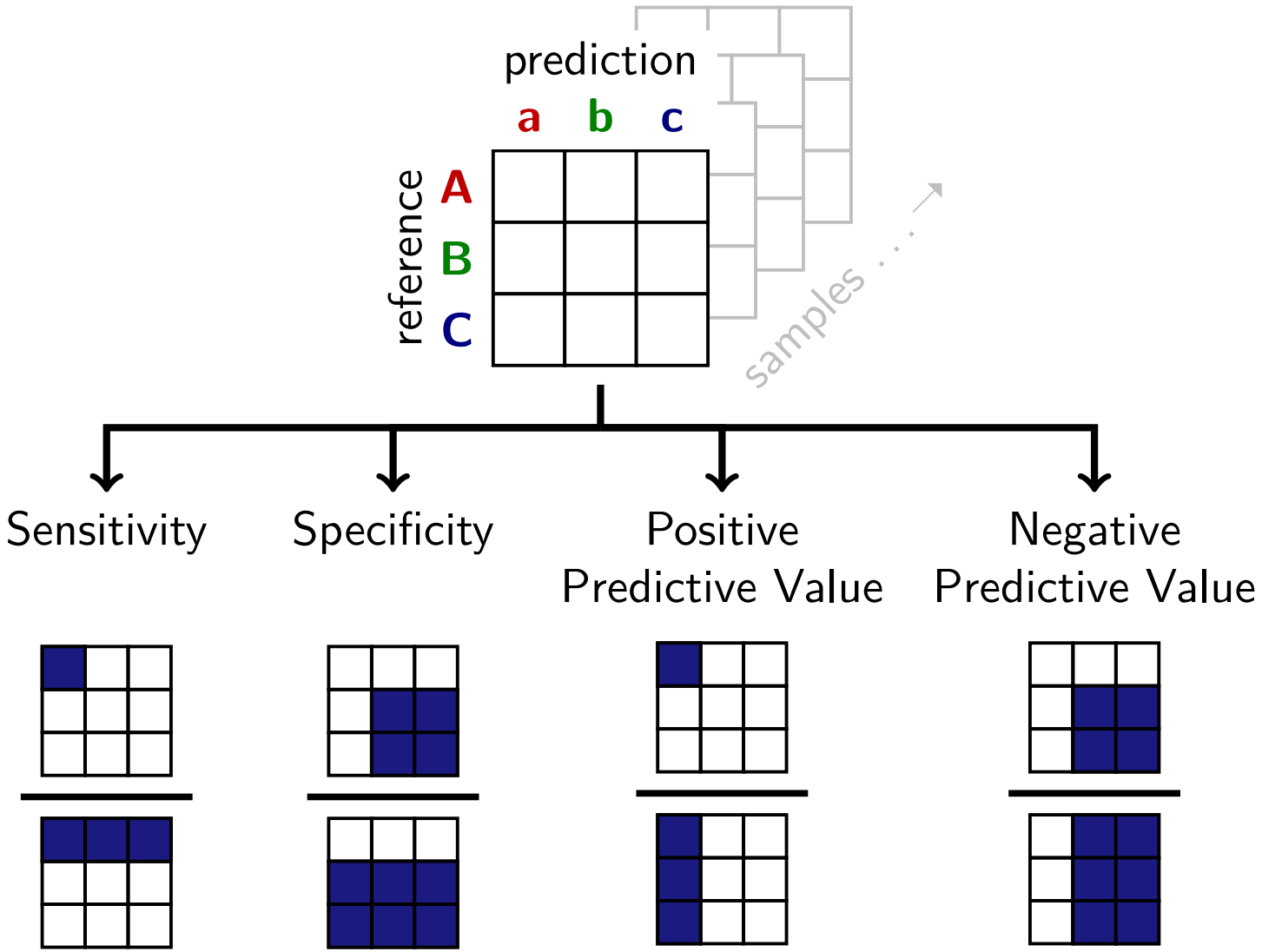
Validation



Validation



Classifier Performance Measures



"Classical" Confusion Matrix

Reference $\mathbf{Y}$ \ Prediction $\hat{\mathbf{Y}}$	1 0	0 1	0.6 0.4	0.7 0.3
1 0	1 0 0 0	0 1 0 0	*harden prediction*	*harden prediction*
0.6 0.4	*not defined*	*not defined*	*not defined*	*not defined*

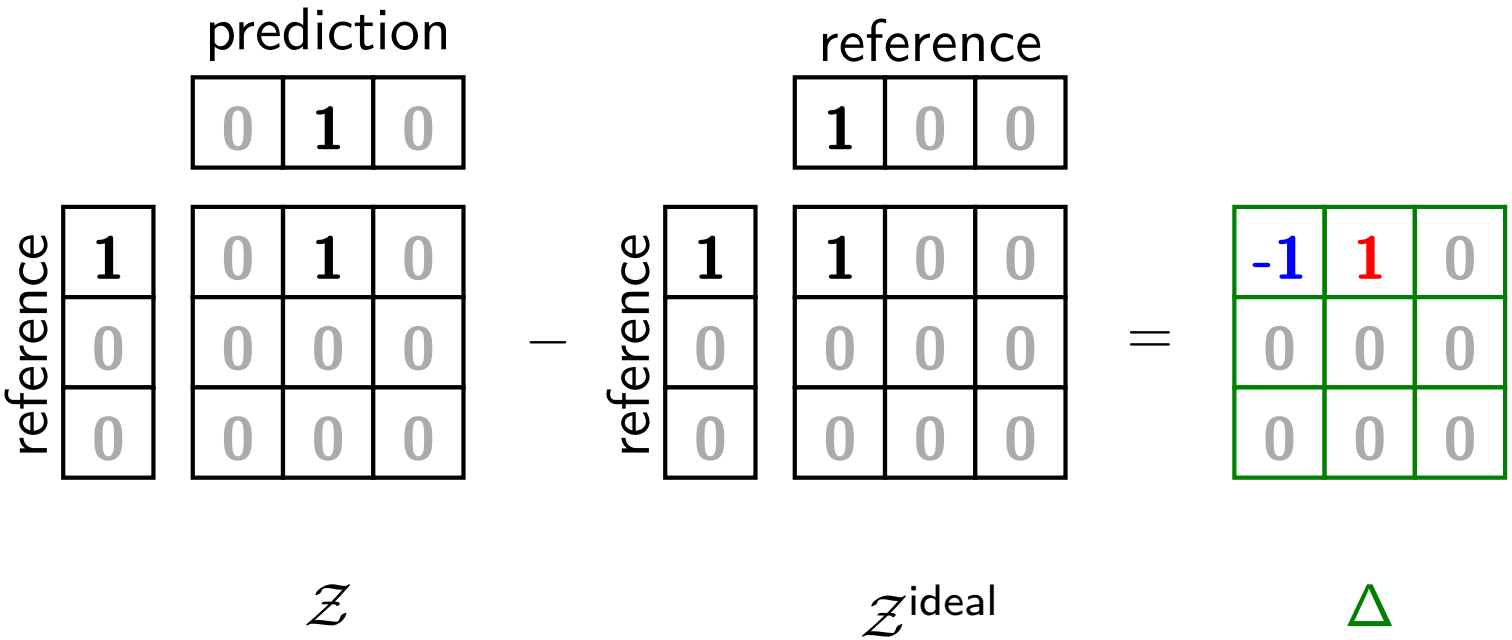
$$\mathcal{Z}_{i,j} = \begin{cases} 1 & \text{if } \mathbf{Y}_i = \hat{\mathbf{Y}}_j = 1 \\ 0 & \text{else} \end{cases}$$

Confusion Matrix: Multiplication

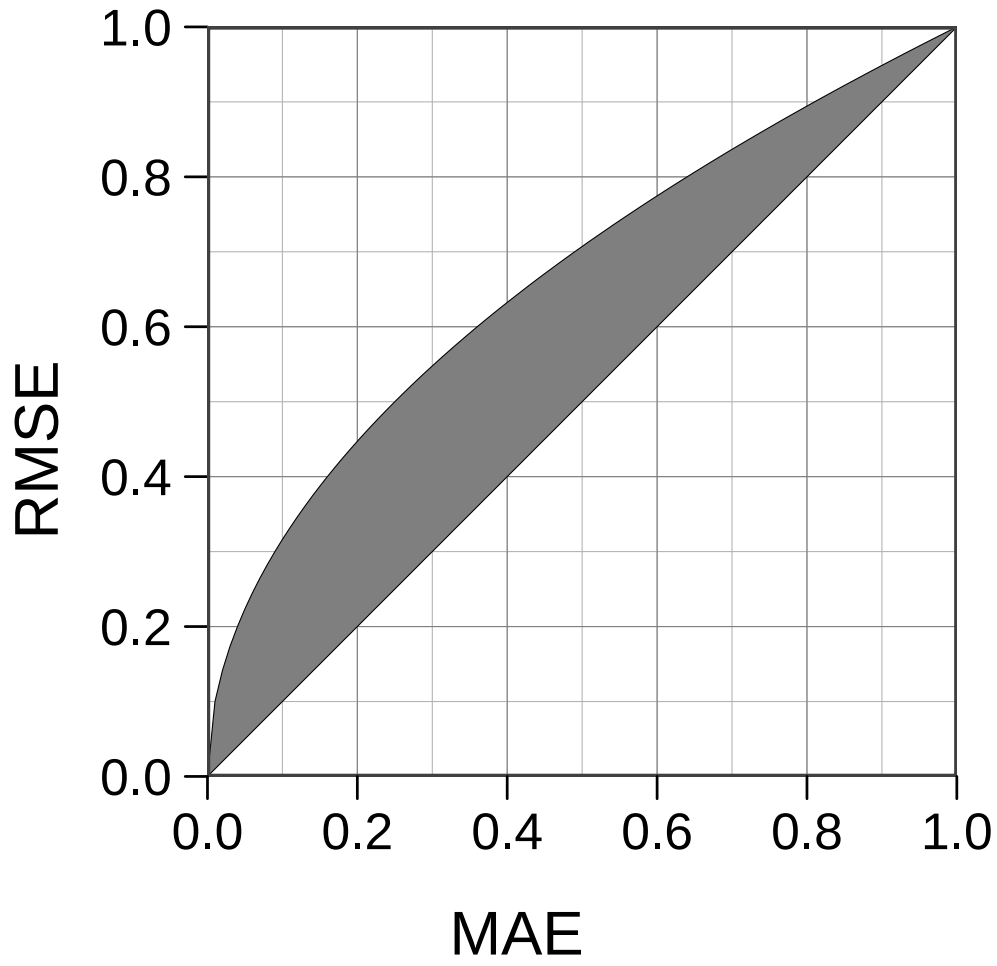
Reference $\mathbf{Y}$ \ Prediction $\hat{\mathbf{Y}}$	1 0	0 1	0.6 0.4	0.7 0.3
1	1 0	0 1	0.6 0.4	0.7 0.3
0	0 0	0 0	0 0	0 0
0.6	0.6 0	0 0.6	0.36 0.24	0.42 0.18
0.4	0.4 0	0 0.4	0.24 0.16	0.28 0.12

$$\mathcal{Z}_{i,j} = \mathbf{Y}_i \cdot \hat{\mathbf{Y}}_j$$

Residual Matrix Δ



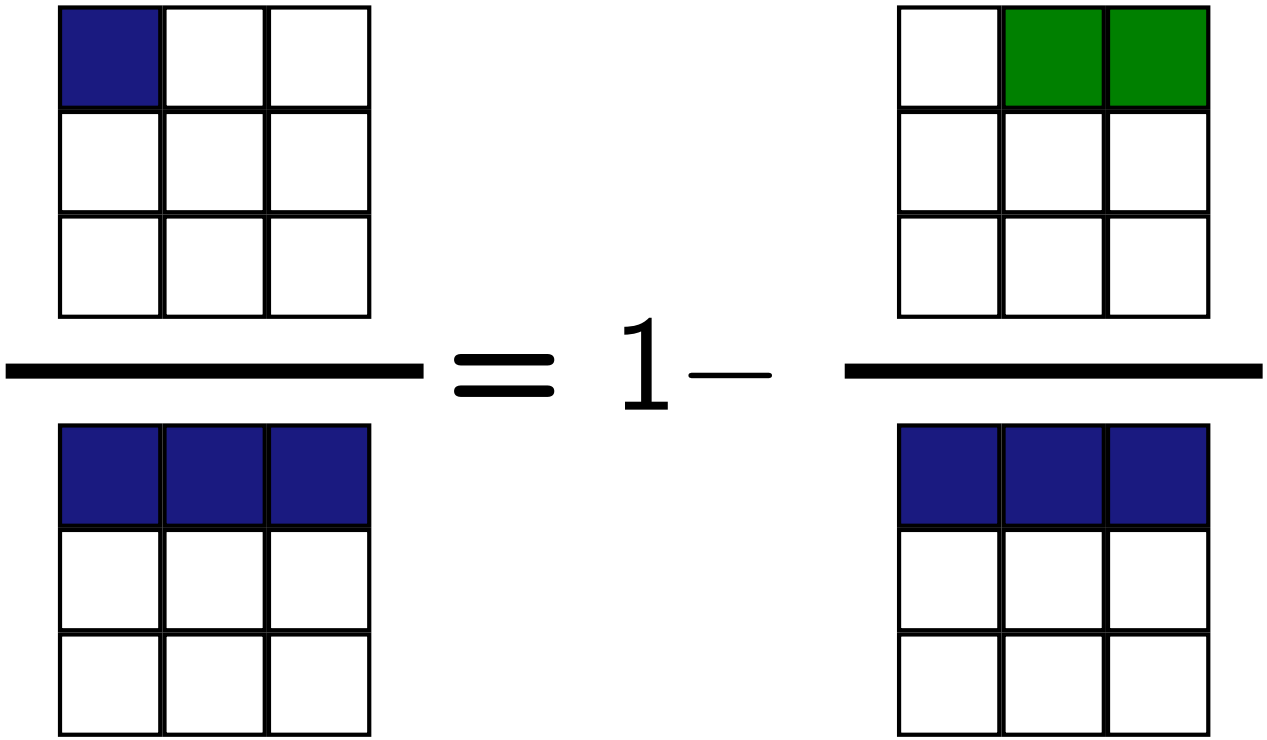
Absolute vs. Squared Error



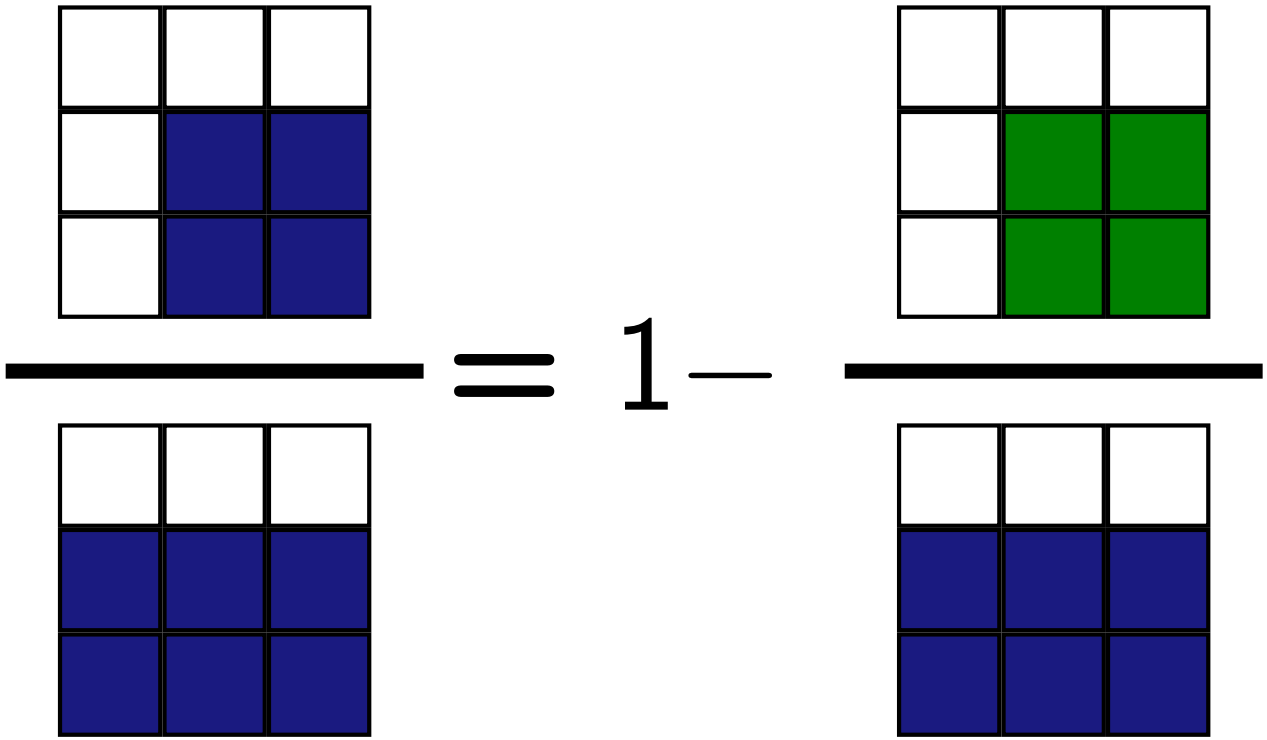
$$MAE = \frac{1}{n} \sum |\hat{y}_i - y_i| \qquad RMSE = \sqrt{\frac{1}{n} \sum (\hat{y}_i - y_i)^2}$$

Putting Things Together: Sensitivity

C. Beleites



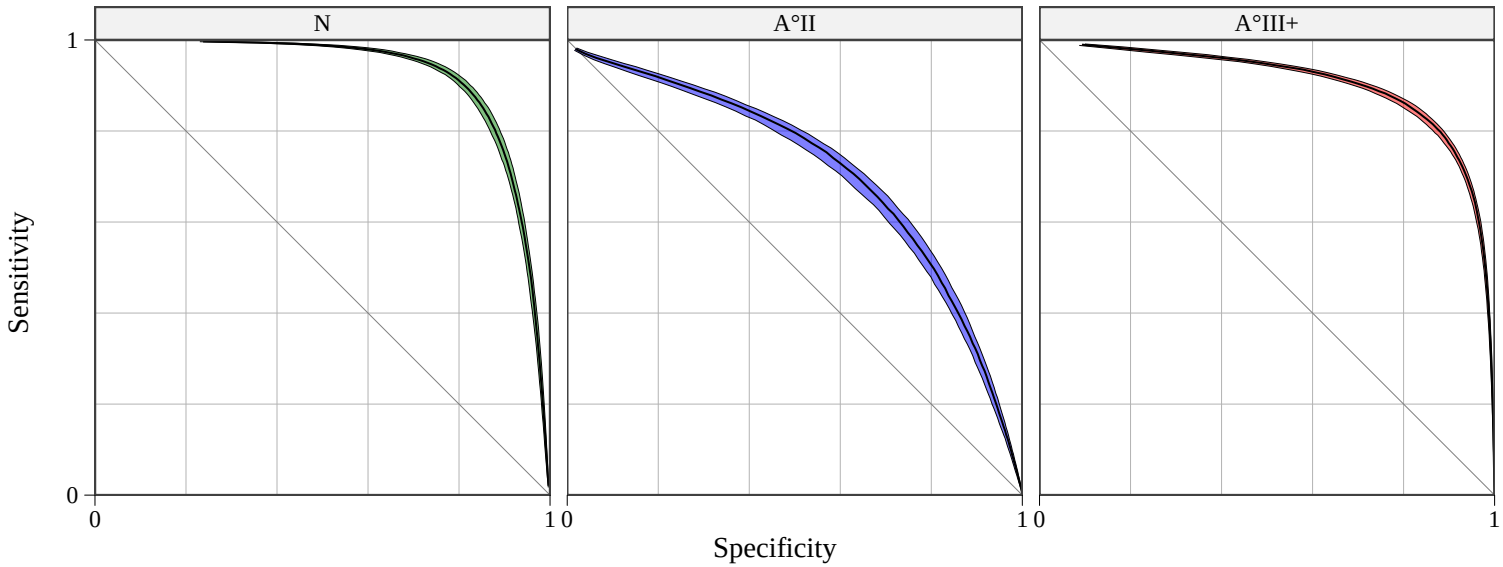
Putting Things Together: Specificity



C. Beleites

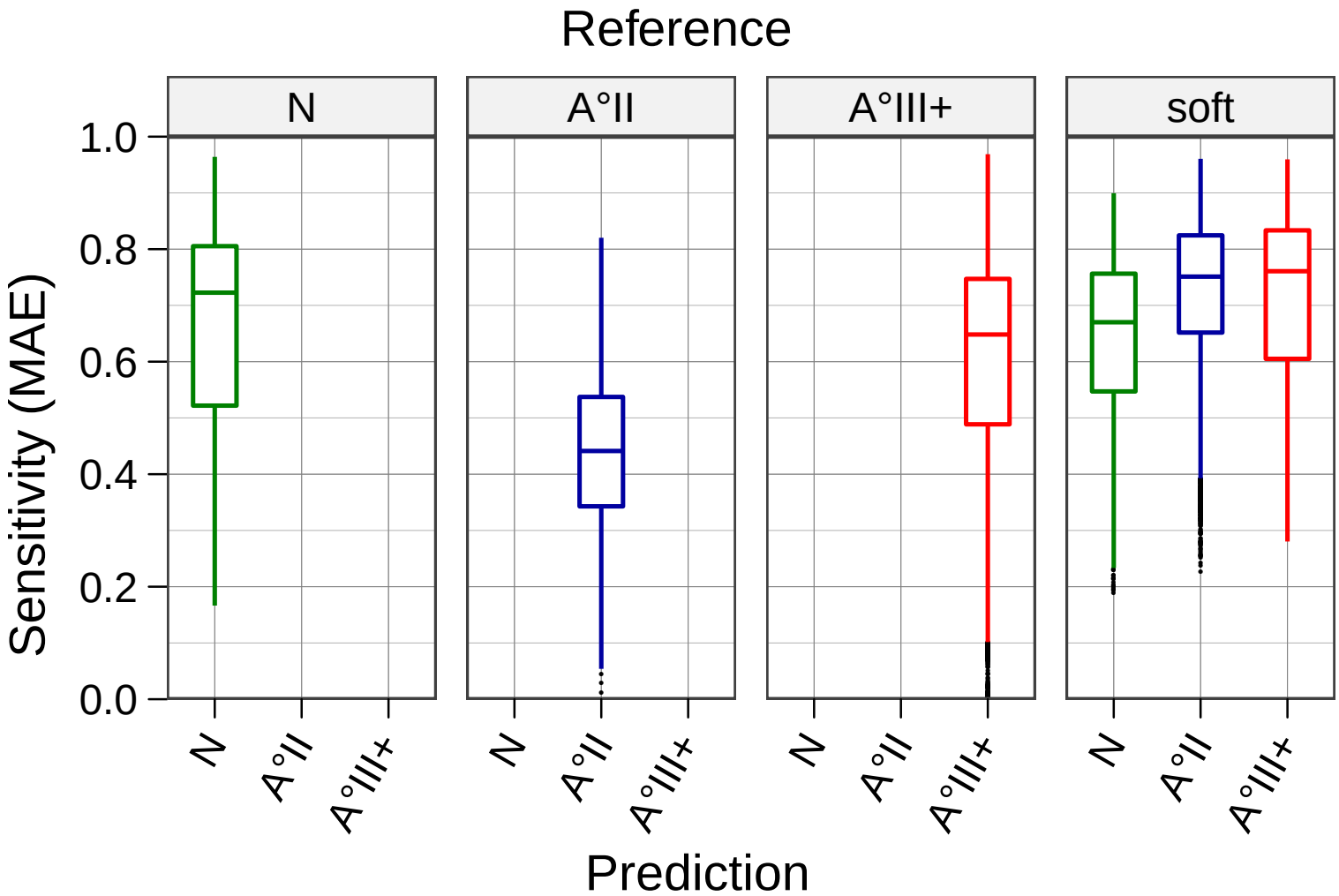
UNIVERSITÀ DEGLI STUDI DI TRIESTE
TECHNISCHE UNIVERSITÄT DRESDEN

“Classical” Confusion Matrix: hardening



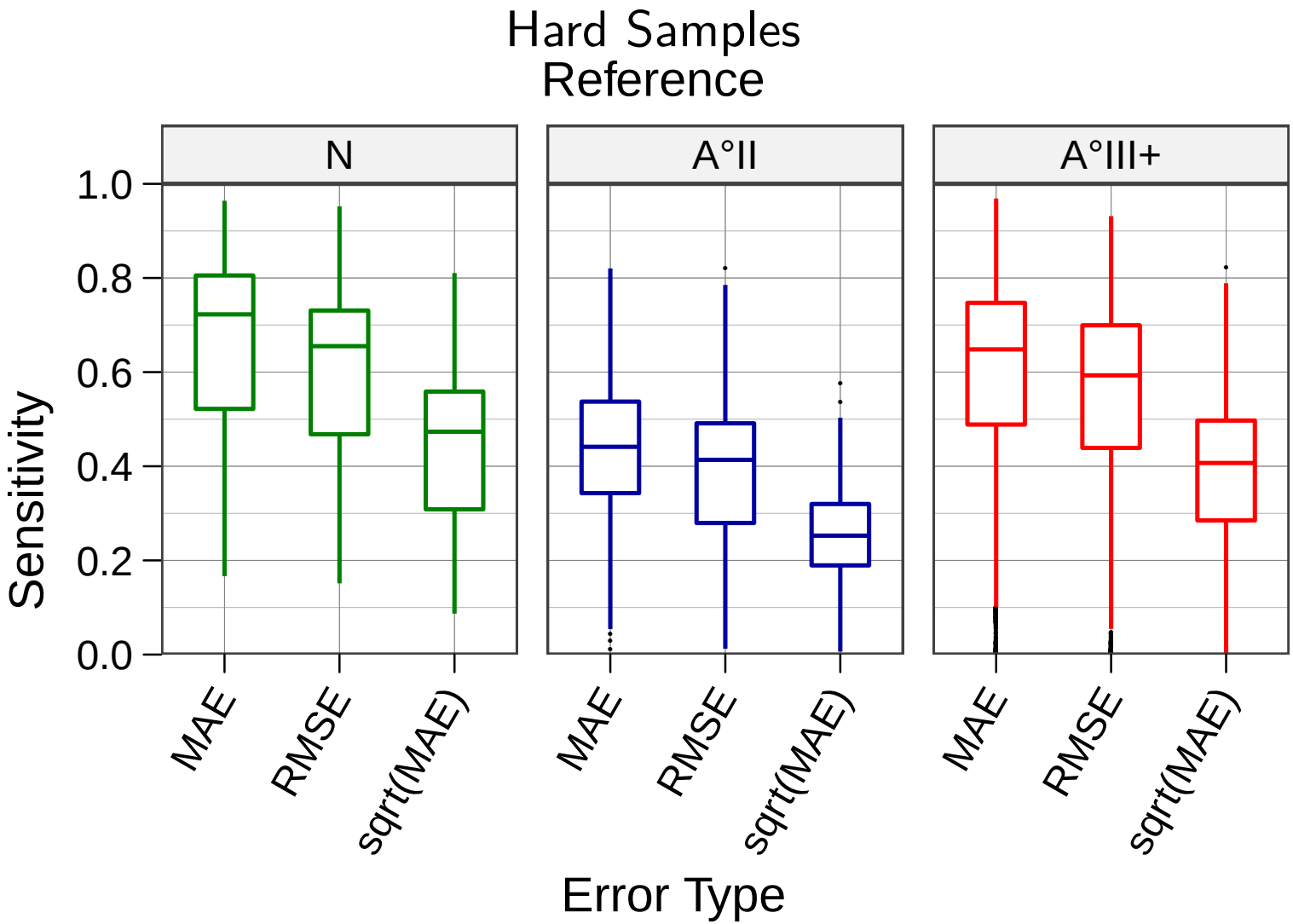
Median, 16th and 84th percentile over 125 iterations.

Application: Sensitivity



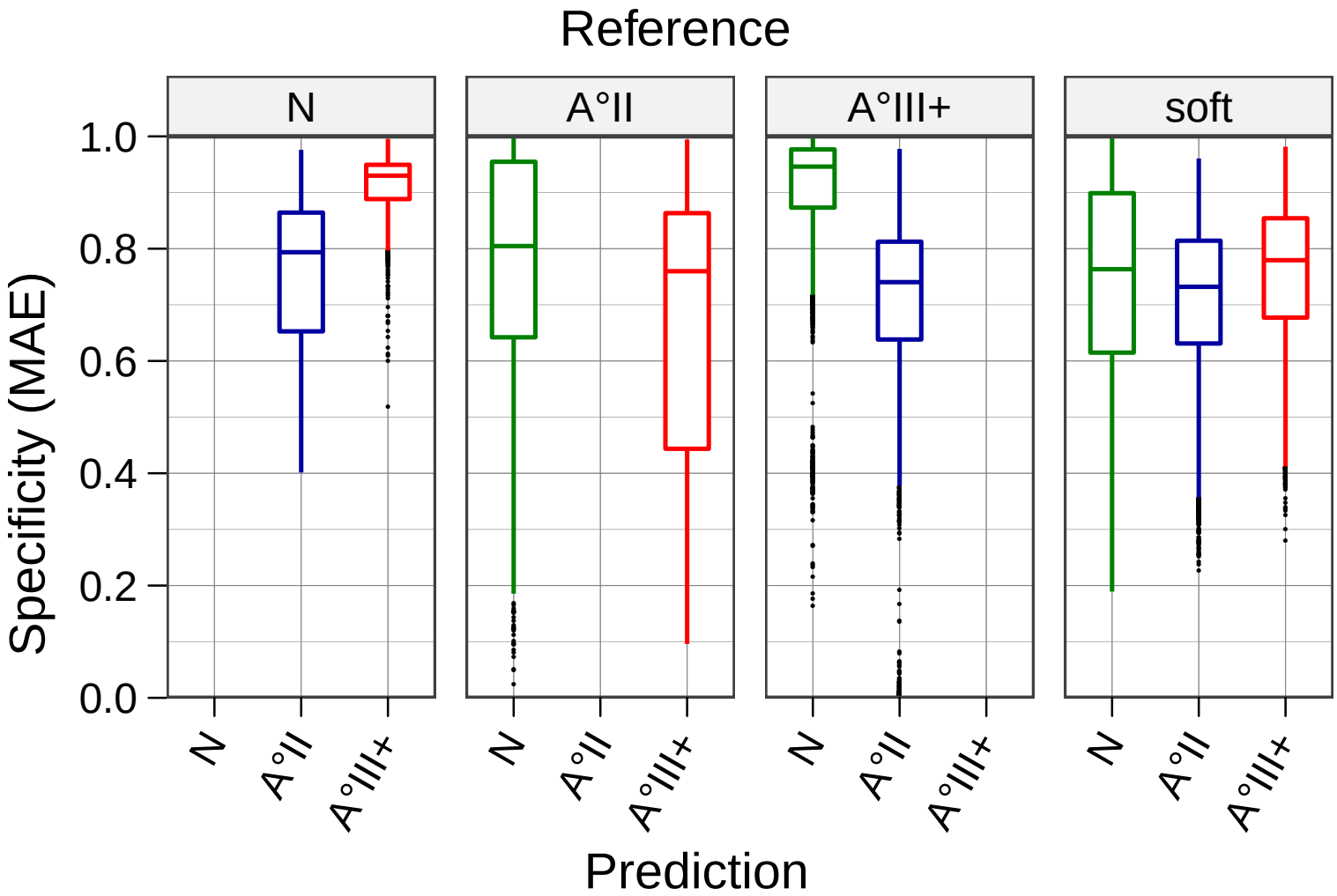
Observed sensitivities for all iterations and samples.

Application: Sensitivity MAE vs. RMSE



Observed sensitivities for all iterations and samples.

Application: Specificity



Observed specificities for all iterations and samples.

Summary

- Soft or partial class membership describes
  - class mixture
  - probability
- Sensitivity, specificity, predictive values for soft classication

✔ *RMSE* or *MAE*

✔ No hardening of soft classifier output needed

✔ More sensitive measures of model performance

✘ Careful comparing classifier performance

- Borderline cases most important

Acknowledgements

Financial Support

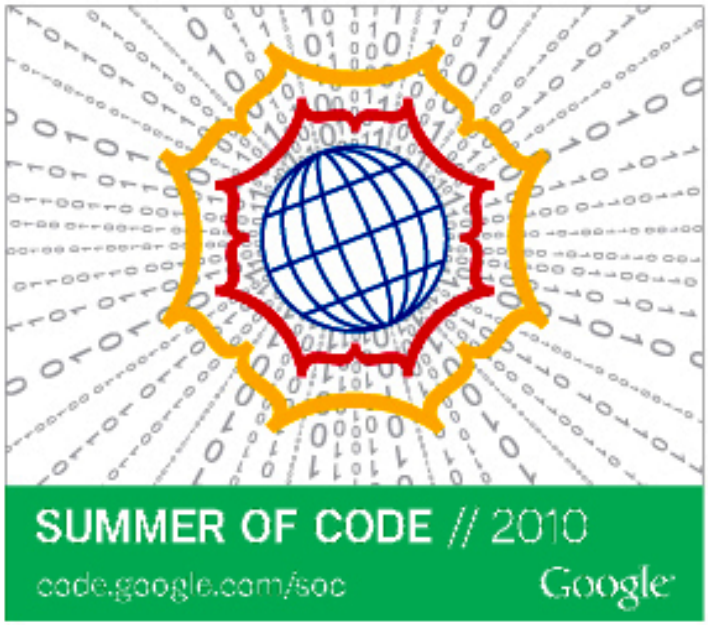
- Deutsche Telekom Stiftung (scholarship)
- Associazione per i Bambini Chirurgici del Burlo
- Fondo Trieste

TECHNISCHE UNIVERSITÄT DRESDEN

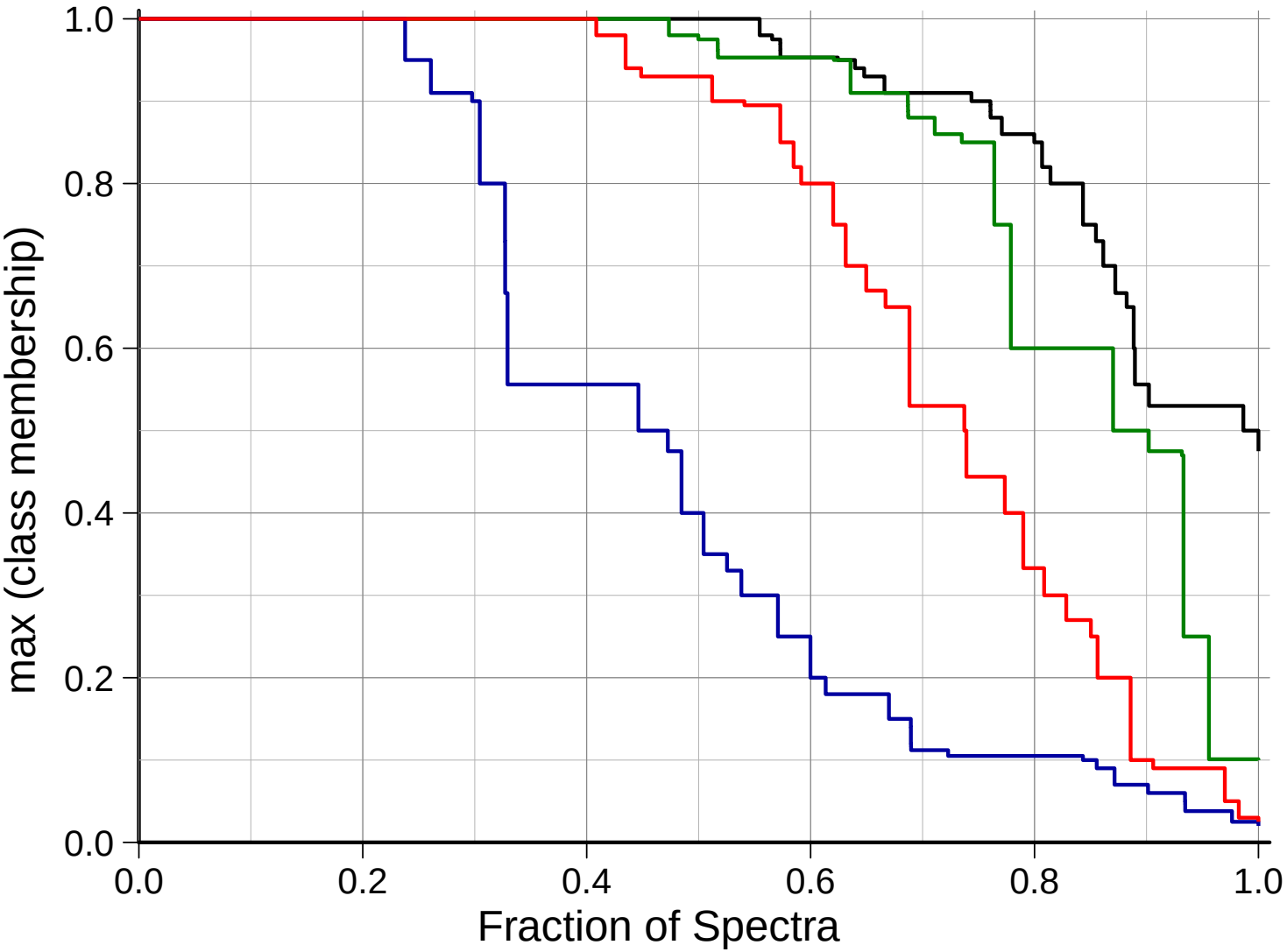
Software

- `softclassval.r-forge.r-project.org`
  The discussed classification performance measures
- `hyperspec.r-forge.r-project.org`:
  Handling of spectroscopic data sets

- Google Summer of Code 2011



Maximum Class Membership



Maximum Class Membership