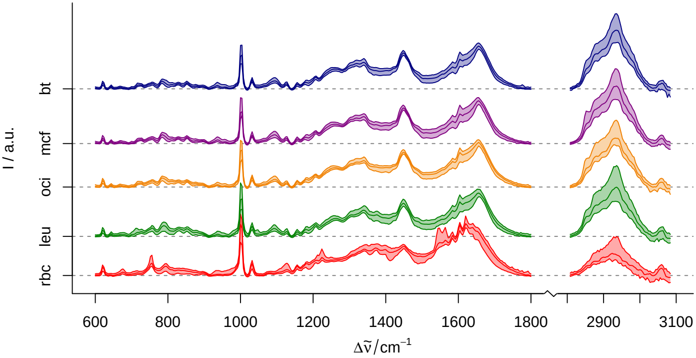
Sample Size Considerations for Raman Spectroscopic Cell Identification

Claudia Beleites[1] (Claudia.Beleites@ipht-jena.de),
Ute Neugebauer[1,2], Thomas Bocklitz[3], Christoph Krafft[1],
and Jürgen Popp[1,3]

[1]Institute of Photonic Technology, Jena/Germany
[2]Center for Sepsis Control and Care, University Hospital Jena, Jena/Germany
[3]Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich-Schiller-University Jena/Germany

Chemometrics in Analytical Chemistry 2012

Raman Spectra of Single Cells



rbc	normal red blood cells	5 donors	372 spectra
leu	normal leukocytes	5 donors	569 spectra
oci	acute myelotic leukemia cell line OCI-AML	5 batches	518 spectra
mcf	breast cancer cell line MCF-7	5 batches	558 spectra
bt	breast cancer cell line BT-20	5 batches	532 spectra
total			2549 spectra

Small Sample Size Problems

- Samples are ever too few...
- But: how many samples (per class) do we **really** need?
- ...to train a good classifier:
  $\Rightarrow$ **learning curve**
- ...to **precisely** measure the classifier's performance:
  $\Rightarrow$ **confidence interval for test results**

Data Analysis Set Up

- PLS-LDA, 25 latent variables (for n_{train} / class < 10: $\frac{1}{2}$ n_{train})
- 50$\times$ iterated 5-fold cross validation
- 100 growing “small” data sets
- “large” test with 320 – 520 spectra / class
  $\rightsquigarrow$ 95 % confidence interval: $p = 0.5 \pm 0.055$
  $\approx$ 1 : 9 split for “small” : “large” sets

Measuring a Learning Curve

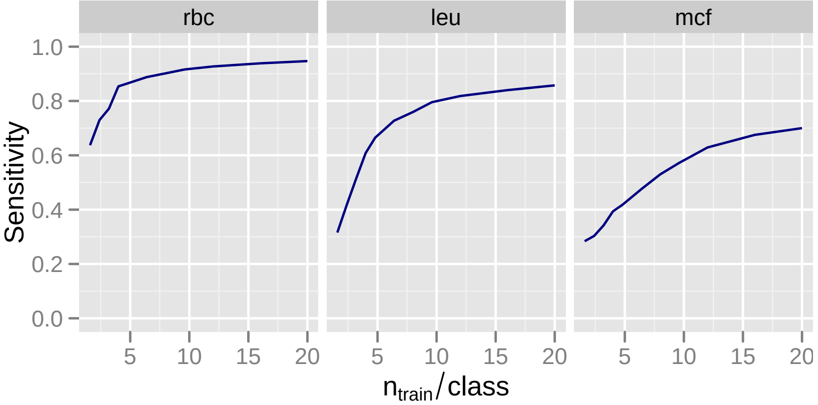
- Learning curve:
- true performance p of model trained with n_{train} samples:

$$\text{error}^2 = \underbrace{\text{Bayes-error}^2 + \text{bias}^2(n_{\text{train}})}_{\text{learning curve}} + \underbrace{\text{var}(n_{\text{train}})}_{\text{model instability}}$$

- observed performance $\hat{p}$ = true performace p

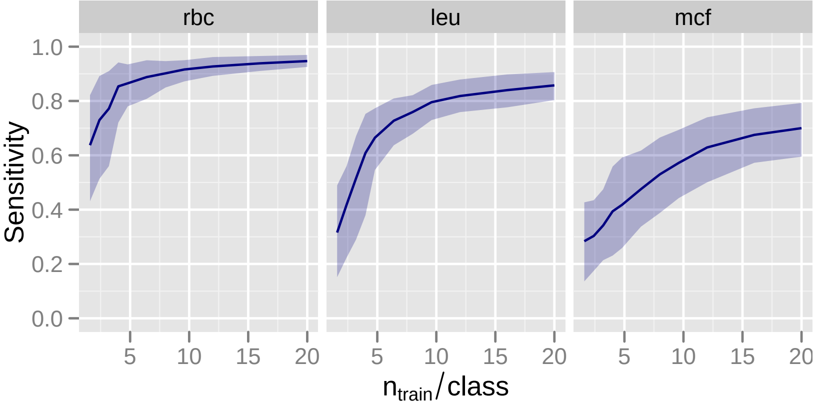
$$+ \text{systematic test error}(n, \text{method})$$
$$+ \text{random test error}(n_{\text{test}})$$

Learning Curve



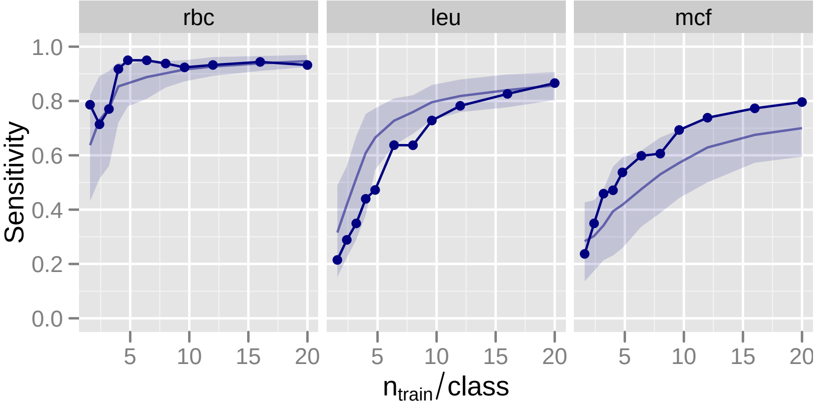
- Confidence band: 5th – 95th percentile of observations
- 100 repetitions
- tested with large test set

Learning Curve



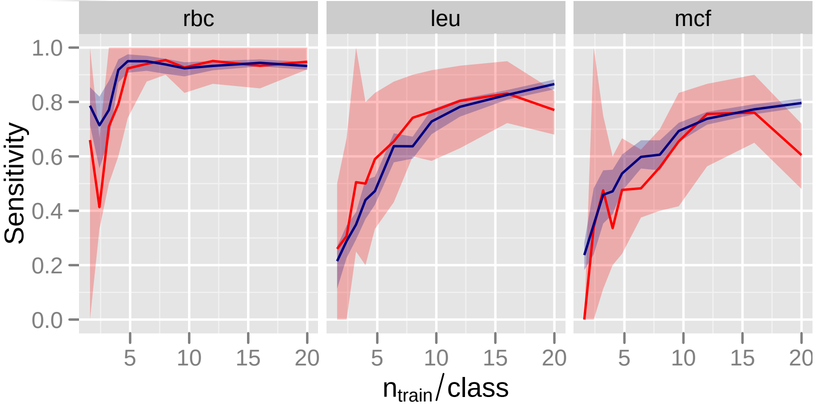
- Confidence band: 5th – 95th percentile of observations
- 100 repetitions
- tested with large test set

Learning Curve



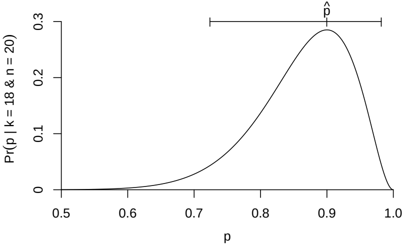
- Confidence band: 5th – 95th percentile of observations
- 100 repetitions
- tested with large test set

Learning Curve: small set



- Confidence band: $5^{th} - 95^{th}$ percentile of observations
- single, growing data set: iteration no. 17
- blue: tested with large test set
- red: 50$\times$ 5-fold cross validation

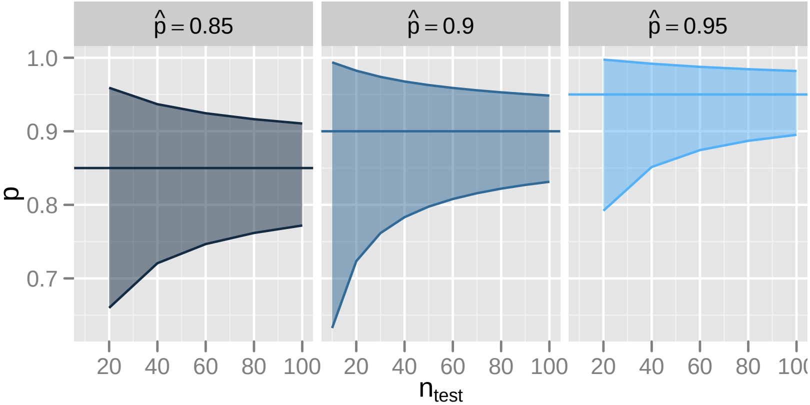
Confidence Intervals for Proportion



- Classifier performance: proportions
- Statistical description: Bernoulli trial
- Uncertainty on proportion: $\text{var}(\hat{p}) = \frac{p(1-p)}{n_{test}}$

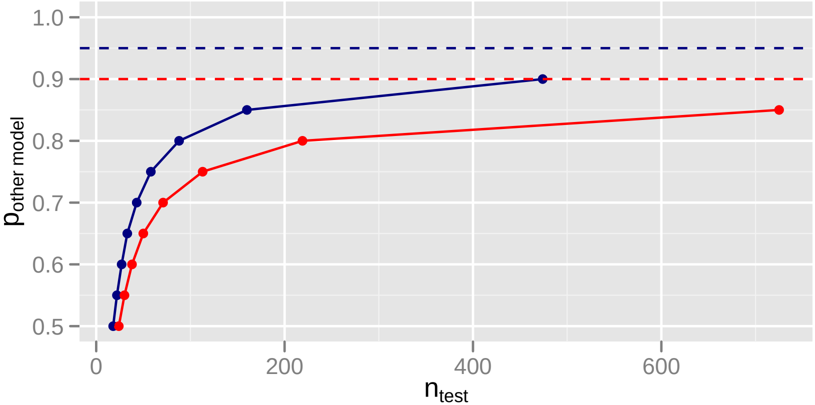
$\rightsquigarrow$ Estimate necessary n_{test}

Confidence Interval Width



- 95 % confidence intervals
- “Bayesian” method

Proving Advantage over Other Model



$\alpha = 5\,\%, \beta = 20\,\%$

from: Fleiss "Statistical Methods for Rates and Proportions"

- More powerful tests available for **paired** test

Summary

- Training a good classifier is not enough, you actually need to demonstrate the performance.
- Learning curve: expected performance + variance
- Distinguish: data set of size n vs. given data set
- Learning curve is difficult to measure from small sample set: **Testing** uncertainty dominates.
- Calculating necessary test sample size
- Necessary n_{test} often $\gg$ necessary n_{train}

EFRE
EUROPA FÜR THÜRINGEN
EUROPÄISCHER FONDS FÜR REGIONALE ENTWICKLUNG

Observing at least $\hat{p}$