

Pattern Recognition VII

Classification Performance Assessment

Michal Haindl

Faculty of Information Technology, KTI
Czech Technical University in Prague
Institute of Information Theory and Automation
Czech Academy of Sciences
Prague, Czechia



<http://ro.utia.cz/>



MI/MIE-ROZ 2023-2024

© M. Haindl

MI-ROZ - 07

1/35

Outline

- 1 Postprocessing
 - Thematic Map Filtration
- 2 Performance Assessment
 - Bayesian error
 - Contingency Table
 - ALL - kNN Classifier
 - Error Probability Estimation
 - Experimental Comparison of Classifiers



© M. Haindl

MI-ROZ - 07

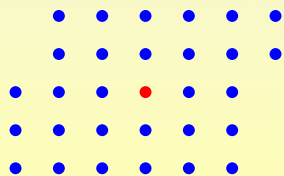
2/35

Thematic Map Filtration

thematic map - matrix whose elements are class indicators
objective to increase TM homogeneity, $P(\text{error}) \searrow$ using a prior knowledge

Requirements

- $P(\text{error}) \searrow$
- $f : \Omega \rightarrow \Omega$, not injective



Dome London



© M. Haindl

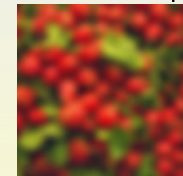
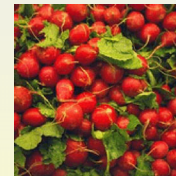
MI-ROZ - 07

3/35

Thematic Map Filtration

Examples

- an alternative low-pass filtering of a feature space



7×7

13×13

- ⊖ between-class object generation (mixels)
- ⊖ localization precision reduction

- majority filter
- median filter
- HFIL, controllable 2D



© M. Haindl

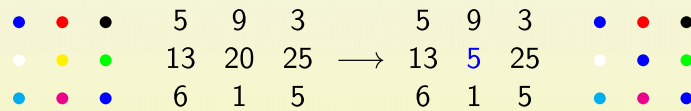
MI-ROZ - 07

3/35

Thematic Map Filtration

Examples

- an alternative low-pass filtering of a feature space
- majority filter



- median filter
- HFIL, controllable 2D



Thematic Map Filtration

Examples

- an alternative low-pass filtering of a feature space
- majority filter

- median filter



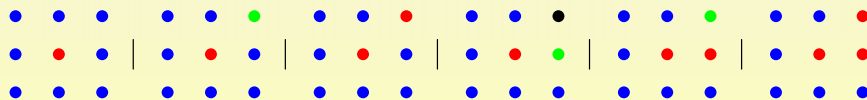
- HFIL, controllable 2D



Thematic Map Filtration

Examples

- an alternative low-pass filtering of a feature space
- majority filter
- median filter
- HFIL, controllable 2D



Classification Performance Assessment

How good is our classifier?

Can we improve our recognition rate?

(Preprocessing, postprocessing, trainee set, data model, features, learning, classifier, context, ensemble of classifiers, etc.?)

- Subjective Evaluation (human evaluator)
- Objective Evaluation
 - Analytical Methods
 - Empirical Methods



Classification Performance Assessment

How good is our classifier?
Can we improve our recognition rate?

- Subjective Evaluation (human evaluator)
- Objective Evaluation
 - Analytical Methods
 - Empirical Methods



Classification Performance Assessment 2

- lack of inherent **superiority** of any classifier (context & usage dependent)
- important → prior information, data distribution, training set size, cost function
- off-training set error estimation
- no assumptions → no preferred feature representation



Classification Performance Assessment 2

- lack of inherent **superiority** of any classifier (context & usage dependent)
- important → prior information, data distribution, training set size, cost function
- off-training set error estimation
- no assumptions → no preferred feature representation



Classification Performance Assessment 2

- lack of inherent **superiority** of any classifier (context & usage dependent)
- important → prior information, data distribution, training set size, cost function
- off-training set error estimation
- no assumptions → no preferred feature representation



Classification Performance Assessment 2

- lack of inherent **superiority** of any classifier (context & usage dependent)
- important → prior information, data distribution, training set size, cost function
- off-training set error estimation
- no assumptions → no preferred feature representation



Classification Error Probability

Bayesian error probability - only numerical solution

$$P(\text{error}) = 1 - \int_S \max_{j=1,\dots,K} \{P(\omega_j)p(X|\omega_j)\} dX$$



Test Set Selection

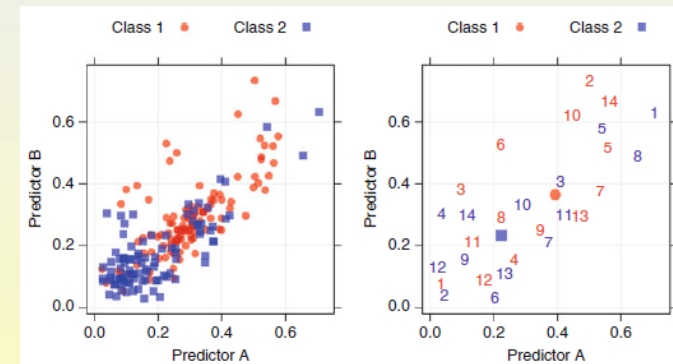
Required properties:

- representative
- random data selection
- $\text{card}(\tilde{T}_i) \rightarrow \infty$



Test Set Selection

maximum dissimilarity sampling



the most dissimilar samples → $\tilde{T}_i$

$$\tilde{T}_1 = \{2, 3, \dots\}$$



Errors Types

- ① type I error, omission error (misdetection)
occurs for true ω_i if assignment is made for ω_j $j \neq i$
- ② type II error, commission error (false identification)
occurs for class ω_i if true class is ω_j $j \neq i$, but assignment is made to ω_i

Mikes, S. - Haindl, M. Texture Segmentation Benchmark, IEEE Transactions on Pattern Analysis and Machine Intelligence ,vol. 44, no. 9, pp. 5647-5663, DOI 10.1109/TPAMI.2021.3075916, ISSN 0162-8828, September, 2022.



Errors Types

- ① type I error, omission error (misdetection)
occurs for true ω_i if assignment is made for ω_j $j \neq i$
- ② type II error, commission error (false identification)
occurs for class ω_i if true class is ω_j $j \neq i$, but assignment is made to ω_i

Mikes, S. - Haindl, M. Texture Segmentation Benchmark, IEEE Transactions on Pattern Analysis and Machine Intelligence ,vol. 44, no. 9, pp. 5647-5663, DOI 10.1109/TPAMI.2021.3075916, ISSN 0162-8828, September, 2022.



Contingency Table

Contingency table (a confusion matrix, error matrix) $C = \{C_{i,j}\}$
where $p_{i,j} = \frac{n_{ji}}{n_{\bullet,i}}$ is the probability that a unit has true class ω_i
and is assigned by the decision rule to ω_j

interpretation	reference data				
	1	2	...	K	
1	$n_{1,1}$	$n_{1,2}$	...	$n_{1,K}$	$n_{1,\bullet}$
2	$n_{2,1}$	$n_{2,2}$	...	$n_{2,K}$	$n_{2,\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
K	$n_{K,1}$	$n_{K,2}$	...	$n_{K,K}$	$n_{K,\bullet}$
	$n_{\bullet,1}$	$n_{\bullet,2}$	...	$n_{\bullet,K}$	

$$n_{i,\bullet} = \sum_{j=1}^K n_{ij}$$

$$n_{\bullet,i} = \sum_{j=1}^K n_{ji}$$

$$\text{card}\{T\} = \sum_{j=1}^K n_{j,\bullet} = \sum_{j=1}^K n_{\bullet,j}$$



Contingency Table

Contingency table (a confusion matrix, error matrix) $C = \{C_{i,j}\}$
where $p_{i,j} = \frac{n_{ji}}{n_{\bullet,i}}$ is the probability that a unit has true class ω_i
and is assigned by the decision rule to ω_j

interpretation	reference data				
	1	2	...	K	
1	$n_{1,1}$	$n_{1,2}$	...	$n_{1,K}$	$n_{1,\bullet}$
2	$n_{2,1}$	$n_{2,2}$	...	$n_{2,K}$	$n_{2,\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
K	$n_{K,1}$	$n_{K,2}$	...	$n_{K,K}$	$n_{K,\bullet}$
	$n_{\bullet,1}$	$n_{\bullet,2}$	...	$n_{\bullet,K}$	

$$n_{i,\bullet} = \sum_{j=1}^K n_{ij}$$

$$n_{\bullet,i} = \sum_{j=1}^K n_{ji}$$

$$\text{card}\{T\} = \sum_{j=1}^K n_{j,\bullet} = \sum_{j=1}^K n_{\bullet,j}$$



Contingency Table 2

number of omissions for ω_i $o_i = n_{\bullet,i} - n_{ii} \downarrow$

number of commissions for ω_i $c_i = n_{i,\bullet} - n_{ii} \downarrow$

number of erroneously classified pixels

$$n_e = \sum_{j=1}^K o_j = \sum_{j=1}^K c_j$$



Contingency Table - Example

interpretation	GT reference data				
	1	2	...	K	
1	$n_{1,1}$	$n_{1,2}$	...	$n_{1,K}$	$n_{1,\bullet}$
2	$n_{2,1}$	$n_{2,2}$	...	$n_{2,K}$	$n_{2,\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
K	$n_{K,1}$	$n_{K,2}$	...	$n_{K,K}$	$n_{K,\bullet}$
	$n_{\bullet,1}$	$n_{\bullet,2}$	...	$n_{\bullet,K}$	

	GT reference data				precision
	1	2	3	4	
i1	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,4}$	$\frac{n_{1,1}}{n_{1,\bullet}}$
i2	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	$n_{2,4}$	$\frac{n_{2,2}}{n_{2,\bullet}}$
i3	$n_{3,1}$	$n_{3,2}$	$n_{3,3}$	$n_{3,4}$	$\frac{n_{3,3}}{n_{3,\bullet}}$
i4	$n_{4,1}$	$n_{4,2}$	$n_{4,3}$	$n_{4,4}$	$\frac{n_{4,4}}{n_{4,\bullet}}$
recall	$\frac{n_{1,1}}{n_{\bullet,1}}$	$\frac{n_{2,2}}{n_{\bullet,2}}$	$\frac{n_{3,3}}{n_{\bullet,3}}$	$\frac{n_{4,4}}{n_{\bullet,4}}$	

	GT reference data				precision
	1	2	3	4	
i1	10	0	4	1	0.66
i2	1	12	1	2	0.75
i3	0	1	7	5	0.54
i4	2	3	0	9	0.64
recall	0.77	0.75	0.58	0.53	



Contingency Table - Example

interpretation	GT reference data				
	1	2	...	K	
1	$n_{1,1}$	$n_{1,2}$	...	$n_{1,K}$	$n_{1,\bullet}$
2	$n_{2,1}$	$n_{2,2}$	...	$n_{2,K}$	$n_{2,\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
K	$n_{K,1}$	$n_{K,2}$	...	$n_{K,K}$	$n_{K,\bullet}$
	$n_{\bullet,1}$	$n_{\bullet,2}$	...	$n_{\bullet,K}$	

	GT reference data				precision
	1	2	3	4	
i1	10	0	4	1	0.66
i2	1	12	1	2	0.75
i3	0	1	7	5	0.54
i4	2	3	0	9	0.64
recall	0.77	0.75	0.58	0.53	

	Recall matrix GT reference data			
	1	2	3	4
i1	0.77	0	0.33	0.06
i2	0.08	0.752	0.08	0.12
i3	0	0.06	0.58	0.29
i4	0.15	0.19	0	0.53

	Precision matrix GT reference data			
	1	2	3	4
i1	0.67	0	0.27	0.07
i2	0.06	0.75	0.06	0.13
i3	0	0.08	0.54	0.38
i4	0.14	0.21	0	0.64



Heuristic Accuracy Estimations

- class accuracy

$$ca_i = \frac{n_{ii}}{o_i + c_i + n_{ii}} \quad \langle 0; 1 \rangle \uparrow$$

- interpretation accuracy (recall, hit rate, sensitivity)

$$co_i = \frac{n_{ii}}{n_{\bullet,i}} \quad \langle 0; 1 \rangle \uparrow$$

- object accuracy (precision)

$$cc_i = \frac{n_{ii}}{n_{i,\bullet}} \quad \langle 0; 1 \rangle \uparrow$$

- overall classification accuracy

$$oa = \frac{1}{n} \sum_{j=1}^K n_{jj}$$



Heuristic Accuracy Estimations 2

- mean square root proportion estimation error

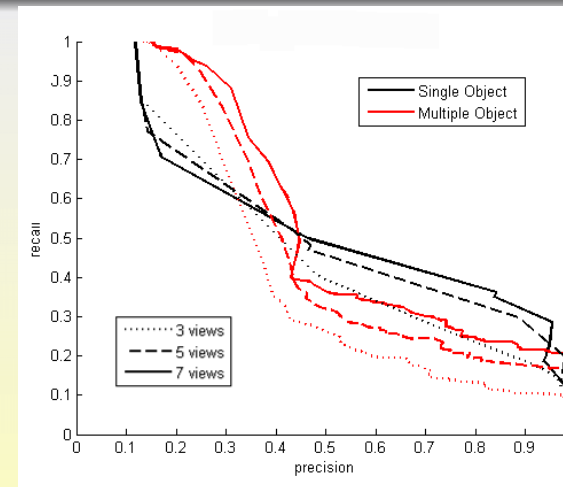
$$rm = \left[\frac{1}{K} \sum_{j=1}^K \left(\frac{n_{j,\bullet}}{n} - \frac{n_{ij}}{n} \right)^2 \right]^{\frac{1}{2}} 10^2 = \frac{1}{n} \left[\frac{1}{K} \sum_{j=1}^K (c_j - o_j)^2 \right]^{\frac{1}{2}} 10^2$$

- mapping score

$$ms_i = 1.5co_i - 0.5 \frac{n_{i,\bullet}}{n_{\bullet,i}} \leq 1 \uparrow$$



Precision Recall Curve



- if P-R curve₁ encloses P-R curve₂ ⇒ classifier₁ is better
- if P-R curve₁ intersects P-R curve₂ → P-R area / specific curve location only



Precision Recall Curve

Precision is the probability that a (randomly selected) retrieved object is relevant (correct interpretation).

$$Precision = \frac{TP}{TP + FP}$$

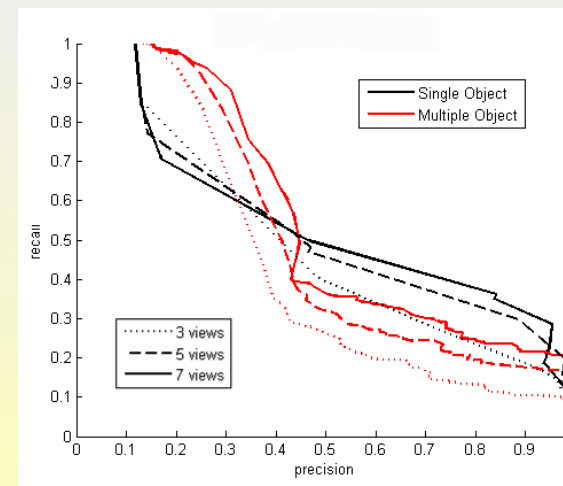
Recall (sensitivity) is the probability that a (randomly selected) relevant object (from the class) is retrieved in a search.

$$Recall = \frac{TP}{TP + FN}$$

(probabilistic interpretation for the uniform distribution over objects)



Precision Recall Curve



Cost-Sensitive Error

unequal costs to different errors:

healthy - diagnosis for cancerous patient $\hat{\omega}_{\text{healthy}}, \omega_{\text{ill}}$

ill - diagnosis for non-cancerous patient $\hat{\omega}_{\text{ill}}, \omega_{\text{healthy}}$

$$c_{\text{ill}, \text{healthy}} < c_{\text{healthy}, \text{ill}}$$

$$\text{error}_{\text{cost-sensitive}} = \frac{1}{n} \sum_{j=1}^K \sum_{f(X_i)=\hat{\omega}_j} (1 - \delta(f(X_i), \omega_i)) c_{j,i}$$

$$c_{j,i} = \begin{cases} 0, & i = j; \\ > 0, & \text{otherwise.} \end{cases}$$



Class Imbalance Problem

$$\text{card}\{T_i\} \gg \text{card}\{T_j\}$$

Example:

$$\begin{aligned} K &= 2 \\ \text{card}\{T_a\} &= 998 & \omega_{\text{apple}} \\ \text{card}\{T_p\} &= 2 & \omega_{\text{pear}} \\ g(X_i) &= \omega_{\text{apple}} \Rightarrow \text{error}_{\text{apple}} = 0,002 \end{aligned}$$



Class Imbalance Problem

$$\text{card}\{T_i\} \gg \text{card}\{T_j\}$$

if possible

undersampling T_i

oversampling T_j



ALI - kNN Classifier 2

$$k = 0.75$$

criteria $\times 100$ in [%] $\bar{R}$ GT, R MS

↑ **CS** (correct detection $\langle 0; 1 \rangle$): $[R_m; \bar{R}_n]$ iff

$$1 \quad \text{card}\{R_m \cap \bar{R}_n\} \geq k \text{ card}\{R_m\}$$

$$2 \quad \text{card}\{R_m \cap \bar{R}_n\} \geq k \text{ card}\{\bar{R}_n\}$$

$$\text{CS} = 67.74 \%$$

↓ **OS** (over-segmentation $\langle 0; 1 \rangle$):

$[R_{m1}, \dots, R_{mx}; \bar{R}_n]$, $2 \leq x \leq M$ iff

$$1 \quad \forall i \in \{1; x\}, \text{card}\{R_{mi} \cap \bar{R}_n\} \geq k \text{ card}\{R_{mi}\}$$

$$2 \quad \sum_{i=1}^x \text{card}\{R_{mi} \cap \bar{R}_n\} \geq k \text{ card}\{\bar{R}_n\}$$

$$\text{OS} = 0.0 \%$$



ALI - kNN Classifier 2

 $k = 0.75$ criteria $\times 100$ in [%] $\bar{R}$ GT, R MS↓ **US** (under-segmentation $\langle 0; 1 \rangle$): $[R_m; \bar{R}_{n1}, \dots, \bar{R}_{nx}], 2 \leq x \leq N$ iff

- 1 $\sum_{i=1}^x \text{card}\{R_m \cap \bar{R}_{ni}\} \geq k \text{card}\{R_m\}$
 - 2 $\forall i \in \langle 1; x \rangle, \text{card}\{R_m \cap \bar{R}_{ni}\} \geq k \text{card}\{\bar{R}_{ni}\}$
- US** = 0.0 %

↓ **ME** (missed $\langle 0; 1 \rangle$): $[\bar{R}_n]$ iff

- 1 $\bar{R}_n \notin$ correct detection
 - 2 $\bar{R}_n \notin$ over-segmentation
 - 3 $\bar{R}_n \notin$ under-segmentation
- ME** = 31.61 %



ALI - kNN Classifier 2

 $k = 0.75$ criteria $\times 100$ in [%] $\bar{R}$ GT, R MS↓ **NE** (noise $\langle 0; 1 \rangle$): $[R_m]$ iff

- 1 $R_m \notin$ correct detection
 - 2 $R_m \notin$ over-segmentation
 - 3 $R_m \notin$ under-segmentation
- NE** = 28.15 %

↓ **O** (omission error – the overall ratio of wrongly interpreted pixels):

$$O = \text{median} \left\{ \frac{O_i}{n_{\bullet,i}} \right\}_{i=1}^N = \text{median} \left\{ \frac{(n_{\bullet,i} - n_{\hat{i},i})}{n_{\bullet,i}} \right\}_{i=1}^N \quad \langle 0; 1 \rangle,$$

where O_i is the i -th class omission error

$$O = 1.74 \%$$



ALI - kNN Classifier 3

↓ **C** (commission error – the overall ratio of wrongly assigned pixels):

$$C = \text{median} \left\{ \frac{C_i}{n_{\hat{i},\bullet}} \right\}_{\hat{i}=1}^M = \text{median} \left\{ \frac{(n_{\hat{i},\bullet} - n_{\hat{i},i})}{n_{\hat{i},\bullet}} \right\}_{\hat{i}=1}^M \quad \langle 0; 1 \rangle,$$

where C_i is the i -th class commission error

$$C = 6.67 \%$$

↑ **CA** (the weighted average class accuracy):

$$CA = \frac{1}{n} \sum_{i=1}^K \frac{n_{\hat{i},i} n_{\bullet,i}}{n_{\bullet,i} + n_{\hat{i},\bullet} - n_{\hat{i},i}} \quad \langle 0; 1 \rangle$$

$$CA = 83.44 \%$$



ALI - kNN Classifier 3

↑ **CO** (recall, the weighted average correct assignment):

$$CO = \frac{1}{n} \sum_{i=1}^K n_{\bullet,i} CO_i = \frac{1}{n} \sum_{i=1}^K n_{\hat{i},i} \quad \langle 0; 1 \rangle$$

$$CO = 90.10 \%$$

↑ **CC** (precision, object accuracy, overall accuracy):

$$CC = \frac{1}{n} \sum_{i=1}^K n_{\bullet,i} CC_i = \frac{1}{n} \sum_{i=1}^K \frac{n_{\hat{i},i} n_{\bullet,i}}{n_{\hat{i},\bullet}} \quad \langle 0; 1 \rangle$$

$$CC = 92.86 \%$$



ALI - kNN Classifier 3

↓ I. (type I error, the weighted probability of wrong assignment of classes pixels):

$$I = \frac{1}{n} \sum_{i=1}^K (n_{\bullet,i} - n_{\hat{i},i}) = 1 - CO \quad \langle 0; 1 \rangle$$

$$I. = 9.90 \%$$

↓ II. (type II error, the weighted probability of commission error):

$$II = \frac{1}{n} \sum_{i=1}^K \frac{n_{\hat{i},\bullet} n_{\bullet,i} - n_{\hat{i},i} n_{\bullet,i}}{n - n_{\bullet,i}} \quad \langle 0; 1 \rangle$$

$$II. = 1.62 \%$$



Classification Error Probability Estimation

methods

- resubstitution (S, apparent) estimate

$$T_i \equiv \tilde{T}_i \quad \forall i$$

$$E\{P_r(\text{error})\} \leq P(\text{error})$$

$$T_i = S_i \quad \forall i \Rightarrow P_r(\text{error}) = P(\text{error})$$

- hold-out estimate

$$T_i \cap \tilde{T}_i = \emptyset \quad \forall i \quad \frac{\text{card}(T_i)}{\text{card}(\tilde{T}_i)} \doteq 1$$

$$E\{P_h(\text{error})\} \geq P(\text{error})$$

⊖ inefficient data use



Classification Error Probability Estimation

methods

- resubstitution (S, apparent) estimate

$$T_i \equiv \tilde{T}_i \quad \forall i$$

$$E\{P_r(\text{error})\} \leq P(\text{error})$$

$$T_i = S_i \quad \forall i \Rightarrow P_r(\text{error}) = P(\text{error})$$

- hold-out estimate

$$T_i \cap \tilde{T}_i = \emptyset \quad \forall i \quad \frac{\text{card}(T_i)}{\text{card}(\tilde{T}_i)} \doteq 1$$

$$E\{P_h(\text{error})\} \geq P(\text{error})$$

⊖ inefficient data use



Classification Error Probability Estimation 2

- cross-validation (leave-one-out, U, deleted) estimate

$$\tilde{T}_i = \bigcup_j {}^j \tilde{T}_i \quad {}^k \tilde{T}_i \cap {}^j \tilde{T}_i = \emptyset \quad \forall k \neq j$$

$$\text{card}\{{}^k \tilde{T}_i\} \approx \text{card}\{{}^j \tilde{T}_i\} \quad \forall k, j$$

leave-one-out

$${}^j \tilde{T}_i = T_i - \{X_j\} \quad \forall X_j \in T_i$$

$${}^j \hat{\mu}_i, {}^j \hat{\Sigma}_i$$

$$E\{P_u(\text{error})\} = \frac{1}{\text{card}(\tilde{T}_i)} \sum_{j=1}^{\text{card}(\tilde{T}_i)} {}^j P_u(\text{error})$$



Classification Error Probability Estimation 2

- cross-validation (leave-one-out, U, deleted) estimate
leave-one-out

$${}^j\tilde{T}_i = T_i - \{X_j\} \quad \forall X_j \in T_i$$

$${}^j\hat{\mu}_i, {}^j\hat{\Sigma}_i$$

$$E\{P_u(error)\} = \frac{1}{card(\tilde{T}_i)} \sum_{j=1}^{card(\tilde{T}_i)} {}^jP_u(error)$$

$$E\{P_u(error)\} \geq P(error)$$

- ⊕ Test sets are independent.
- ⊖ Training sets are interdependent (significant overlap).
- ⊖ Computationally demanding.
- ⊖ Which of the $card(\tilde{T}_i)$ classifiers to use?



Classification Error Probability Estimation 2

- bootstrap
 ${}^j\tilde{T}_i \subset T_i$ randomly generated with replacement
 ${}^j\tilde{T}_i \cap {}^l\tilde{T}_i \neq \emptyset \quad j \neq l \quad card\{{}^j\tilde{T}_i\} = n \quad \forall j \in J$

$$E\{P_b(error)\} = \frac{1}{card(J)} \sum_{j=1}^{card(J)} {}^jP_b(error)$$

- ⊕ Simplicity.
- ⊕ Asymptotically consistent (under some conditions).
- ⊖ Computationally demanding.
- ⊖ Which of the $card(J)$ classifiers to use (one / ensemble / complete T_i)?



Classification Error Probability Estimation 2

bootstrap - out-of-bag estimate

if $\hat{T}$ is created by $n \times$ one sample randomly picked from T , the chance for a training sample not being selected for $\hat{T}$:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^n = \frac{1}{e} \approx 0.368$$

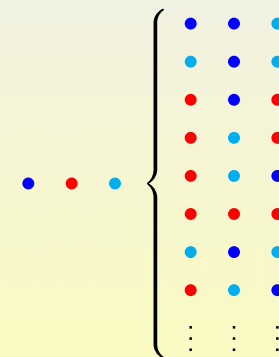
then $T_b = \hat{T}$ and $\tilde{T}_b = T \setminus \hat{T}$



Classification Error Probability Estimation 2

bootstrap

3^3



Classification Error Probability Estimation 3

- rotation π estimate

$${}^j\tilde{T}_i = \{{}^iX_1, \dots, {}^iX_j\} \quad 1 \leq j \leq \frac{\text{card}(T_i)}{2}$$

$$\forall j \tilde{T}_i : j \tilde{T}_i \cap k \tilde{T}_i = \emptyset \quad {}^jP_\pi(\text{error})$$

$$E\{P_\pi(\text{error})\} = \frac{1}{n_\pi} \sum_{j=1}^{n_\pi} {}^jP_\pi(\text{error})$$

$$E\{P_\pi(\text{error})\} \geq P(\text{error})$$

$j = 1 \rightarrow$ U method, $j = \frac{\text{card}(T_i)}{2} \rightarrow$ H method
usually more biased than U estimate but has smaller variance

- Toussaint's estimator



Classification Error Probability Estimation 3

- rotation π estimate
- Toussaint's estimator

$$P_t(\text{error}) = \alpha E\{P_\pi(\text{error})\} + (1-\alpha)P_r(\text{error}) \quad 0 \leq \alpha \leq 1$$

$$\alpha = \frac{1}{2} \Rightarrow P_t(\text{error}) \approx P_u(\text{error})$$



Error Estimation Example 2

MSS	LANDSAT	
MSS4	0.5 - 0.6 μm	(green)
MSS5	0.6 - 0.7 μm	(red)
MSS6	0.7 - 0.8 μm	(near infrared)
MSS7	0.8 - 1.1 μm	(near infrared)

$$d = 4 \quad K = 7$$

(clover, wheat, wood, water, maize, residential area, grass)

$$194 \leq \text{card}(T_i) \leq 1525 \quad p(\omega_i) = \frac{1}{7}$$



Error Estimation Example 3

Bayes classifier $\Sigma_i \neq \Sigma_j$	$P(\text{error})$ [%]
resubstitution	27.6
hold-out	33.9
rotation	30.4
Toussaint $\alpha = 0.5$	28.9



Error Estimation Example 3

Bayes classifier $\Sigma_i \neq \Sigma_j$	$P(error)$ [%]
resubstitution	27.6
hold-out	33.9
rotation	30.4
Toussaint $\alpha = 0.5$	28.9

classifier	$P_{resubstitution}(error)$ [%]
$\Sigma_i \neq \Sigma_j$	27.6
$\Sigma_i = \frac{1}{7} \sum_j \Sigma_j$	32.4
$\Sigma_i = diag[...]$	29.6
$\Sigma_i = I$	33.4



Contingency Table Example

TM	LANDSAT	
TM1	0.45 - 0.52 μm	(blue)
TM2	0.52 - 0.6 μm	(green)
TM3	0.63 - 0.69 μm	(orange)
TM4	0.76 - 0.9 μm	(near infrared)
TM5	1.55 - 1.75 μm	(short wave infrared)
TM7	2.08 - 2.35 μm	(infrared)
TM6	10.4 - 12.5 μm	(thermal infrared)

8 bits/pixel (256) , $30 \times 30 m^2$ ($120 \times 120 m^2$ TM6) 460 Mbit



Contingency Table Example

TM	LANDSAT	
TM1	0.45 - 0.52 μm	(blue)
TM2	0.52 - 0.6 μm	(green)
TM3	0.63 - 0.69 μm	(orange)
TM4	0.76 - 0.9 μm	(near infrared)
TM5	1.55 - 1.75 μm	(short wave infrared)
TM7	2.08 - 2.35 μm	(infrared)
TM6	10.4 - 12.5 μm	(thermal infrared)

$K = 15$, $d = 7$, Bayes classifier $\Sigma_i \neq \Sigma_j$, resubstitution

1. water, 2. red clover, 3. white clover, 4. wood, 5. wheat,
6. maize, 7. mixture, 8. millet, 9. grass, 10. harvested rapeseed,
11. rapessed, 12. sugarbeet, 13. wheat II, 14. mixed clover,
15. residential area



Contingency Table Example 2

class	Bayesian $\Sigma_i \neq \Sigma_j \forall i \neq j$	Euclidean $\Sigma_i = I \forall i$	linear $\Sigma_i = \Sigma \forall i$	linearized classifier
1	1	1	0.92	1
2	0.8	0.65	0.45	0.7
3	0.99	0.82	0.85	0.91
4	1	0.81	0.71	0.8
5	0.98	0.93	0.9	0.94
6	1	1	0.98	1
7	0.92	0.79	0.88	0.88
8	1	0.93	0.88	0.93
9	0.8	0.46	0.74	0.75
10	0.97	0.97	0.9	0.96
11	1	0.89	0.89	0.91
12	1	0.91	0.9	0.93
13	1	0.6	0.89	0.9
14	0.92	0.88	0.9	0.91
15	0.95	0.58	0.23	0.65
accuracy	0.96	0.81	0.79	0.88

$oa = 0.774$

$\# = 2337$

$khat = 0.772$



Contingency Table Example 2

K		Reference Data															η_i
.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
I	0	4	22	91	74	21	5	3	5	3	10	5	6	4	7	134	394
N	1	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49
T	2	0	66	0	0	1	0	5	0	16	0	0	0	0	0	0	88
E	3	0	0	527	0	1	0	0	0	0	0	0	0	0	0	0	528
R	4	0	0	0	194	0	0	0	0	0	0	0	0	0	0	5	199
P	5	0	2	3	0	235	0	0	0	0	0	0	0	0	0	3	243
E	6	0	0	0	0	0	197	0	0	1	0	0	0	0	0	0	198
T	7	0	1	0	0	0	0	64	0	4	0	0	0	0	0	0	69
A	8	0	0	0	0	0	0	0	50	0	1	0	0	0	0	2	53
T	9	0	57	0	0	4	0	1	0	70	0	0	0	0	0	0	132
I	10	0	0	0	1	0	0	2	0	103	0	0	0	0	0	10	116
T	11	0	0	0	4	0	0	0	0	0	0	40	0	0	0	1	45
O	12	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	52
N	13	0	0	0	0	0	0	0	0	0	0	0	0	31	1	2	34
	14	0	0	0	0	0	0	0	0	0	1	0	0	0	81	5	87
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	49	50
$n_{\bullet,i}$		53	148	621	273	262	202	73	57	94	115	45	58	35	90	211	1808
ca_i		92	39	85	70	87	97	82	83	45	80	80	90	82	84	23	[%]
co_i		92	45	85	71	90	98	88	88	74	90	89	90	89	90	23	[%]
cc_i		100	75	100	97	97	99	93	94	53	89	89	100	91	93	98	[%]

$$oa = 0.774$$

= 2337

$$\hat{k} = 0.772$$



Contingency Table Example 2

K		Reference Data															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$n_{i\bullet}$
I N T E R P R E T A T I O N	0	4	22	91	74	21	5	3	5	3	10	5	6	4	7	134	394
	1	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49
	2	0	66	0	0	1	0	5	0	16	0	0	0	0	0	0	88
	3	0	0	527	0	1	0	0	0	0	0	0	0	0	0	0	528
	4	0	0	0	194	0	0	0	0	0	0	0	0	0	0	5	199
	5	0	2	3	0	235	0	0	0	0	0	0	0	0	0	3	243
	6	0	0	0	0	0	197	0	0	1	0	0	0	0	0	0	198
	7	0	1	0	0	0	0	64	0	4	0	0	0	0	0	0	69
	8	0	0	0	0	0	0	0	50	0	1	0	0	0	0	2	53
	9	0	57	0	0	4	0	1	0	70	0	0	0	0	0	0	132
	10	0	0	0	1	0	0	0	2	0	103	0	0	0	0	10	116
	11	0	0	0	4	0	0	0	0	0	0	40	0	0	0	1	45
	12	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	52
	13	0	0	0	0	0	0	0	0	0	0	0	0	31	1	2	34
	14	0	0	0	0	0	0	0	0	0	1	0	0	0	81	5	87
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	49	50	
$n_{\bullet i}$		53	148	621	273	262	202	73	57	94	115	45	58	35	90	211	1808
ca_i		92	39	85	70	87	97	82	83	45	80	80	90	82	84	23	[%]
co_i		92	45	85	71	90	98	88	88	74	90	89	90	89	90	23	[%]
cc_i		100	75	100	97	97	99	93	94	53	89	89	100	91	93	98	[%]

2. red clover $\times$ 9. grass

15. residential area

$$oa = 0.774$$

$$\# = 2337$$

$$\hat{k} = 0.772$$



Contingency Table Example 3



2. red clover
0.45



9. grass
0.39



Contingency Table Example 3



15. residential area
0.23



?

0. unclassified

0.64



Experimental Comparison of Classifiers

Testing set choice - different testing sets may rank differently classifiers. Thus conclusions drawn from a single test set can be unreliable.

Training set choice - some classifiers (e.g. decision trees, neural networks) are unstable. Small changes in the training set can cause substantial changes of the classifier trained on this set.

Learning randomness - some learning methods have a random component (e.g. initialization in the backpropagation algorithm, genetic algorithms). Thus the same T_i can produce different classifiers.



McNemar Test

C_1	C_2	N_{ab}
1	1	N_{11}
1	0	N_{10}
0	1	N_{01}
0	0	N_{00}

two classifiers C_1, C_2

$$\chi^2 = \frac{(|N_{01} - N_{10}| - 1)^2}{N_{01} + N_{10}}$$

if $\chi^2 > 3.841$ then H_0 (no difference) is rejected on the level of significance 0.05

N_{01} number of objects classified using C_2 as correct, C_1 wrong

N_{10} number of C_1 correct, C_2 wrong



McNemar Test

$$\chi^2 = \frac{(|N_{01} - N_{10}| - 1)^2}{N_{01} + N_{10}}$$

if $\chi^2 > 3.841$ then H_0 (no difference) is rejected on the level of significance 0.05

N_{01} number of objects - C_2 as correct, C_1 wrong

N_{10} number of objects - C_1 correct, C_2 wrong

Example:

C_1	C_2	N_{ab}
1	1	203
1	0	14
0	1	5
0	0	27

$$\chi^2 = \frac{(|5 - 14| - 1)^2}{5 + 14} = \frac{64}{19} = 3.368 < 3.841$$

H_0 accepted - both classifiers are similar



Friedman Test

multiple data sets $D_1, D_2, \dots, D_N$ & $K > 2 \rightarrow$ Friedman test

$$\tau_{\chi^2} = \frac{12N}{K(K-1)} \left(\sum_{i=1}^K r_i^2 - \frac{K(K+1)^2}{4} \right)$$

r_i averaged rank of the i -th algorithm

small K, N :

$$\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(K-1) - \tau_{\chi^2}}$$



Friedman Test - Example

<i>Data set</i>	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$
D_1	1	2	3
D_2	1	2.5	2.5
D_3	1	2	3
D_4	2	1	3
D_5	1	2	3
r_i	1.2	1.9	2.9

$$\tau_{\chi^2} = \frac{60}{6} \left(\sum_{i=1}^3 r_i^2 - \frac{48}{4} \right) = 14.6$$

$\left. \begin{array}{l} 4.6 > 4.459_{\alpha=0.01} \\ 4.6 > 3.113_{\alpha=0.05} \end{array} \right\} \Rightarrow \text{hypothesis that } \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3 \text{ have}$
the same performance is **rejected**

