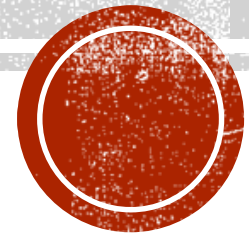


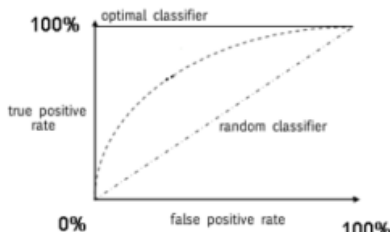
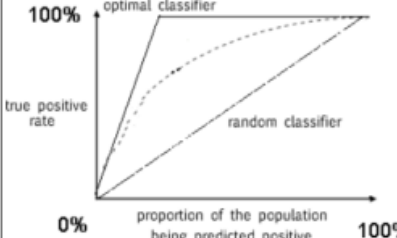
# MODEL SELECTION

Lectures 5,6



Binary classification performances measure cheat sheet

Damien François – v1.0 - 2009 (damien.francois@uclouvain.be)

<b>Confusion matrix for two possible outcomes <math>p</math> (positive) and <math>n</math> (negative)</b>																									
<table><tr><td colspan="2"></td><th colspan="2">Actual</th><td></td></tr><tr><td colspan="2"></td><th><math>p</math></th><th><math>n</math></th><th>Total</th></tr><tr><th rowspan="2">Predicted</th><th><math>p'</math></th><td>true positive</td><td>false positive</td><td><math>P</math></td></tr><tr><th><math>n'</math></th><td>false negative</td><td>true negative</td><td><math>N</math></td></tr><tr><td colspan="2"></td><th>total</th><th><math>P'</math></th><th><math>N'</math></th></tr></table>				Actual					$p$	$n$	Total	Predicted	$p'$	true positive	false positive	$P$	$n'$	false negative	true negative	$N$			total	$P'$	$N'$
		Actual																							
		$p$	$n$	Total																					
Predicted	$p'$	true positive	false positive	$P$																					
	$n'$	false negative	true negative	$N$																					
		total	$P'$	$N'$																					
Classification accuracy $(TP + TN) / (TP + TN + FP + FN)$ Error rate $(FP + FN) / (TP + TN + FP + FN)$																									
<b>Paired criteria</b>																									
<b>Precision:</b> (or Positive predictive value) proportion of predicted positives which are actual positive $TP / (TP + FP)$	<b>True positive rate:</b> proportion of actual positives which are predicted positive $TP / (TP + FN)$																								
<b>Recall:</b> proportion of actual positives which are predicted positive $TP / (TP + FN)$	<b>True negative rate:</b> proportion of actual negative which are predicted negative $TN / (TN + FP)$																								
<b>Sensitivity:</b> proportion of actual positives which are predicted positive $TP / (TP + FN)$	<b>Positive likelihood:</b> likelihood that a predicted positive is an actual positive sensitivity / (1 - specificity)																								
<b>Specificity:</b> proportion of actual negative which are predicted negative $TN / (TN + FP)$	<b>Negative likelihood:</b> likelihood that a predicted negative is an actual negative specificity / (1 - sensitivity)																								
<b>Combined criteria</b>																									
<b>BCR:</b> Balanced Classification Rate $\frac{1}{2} (TP / (TP + FN) + TN / (TN + FP))$ <b>BER:</b> Balanced Error Rate, or <b>HTER:</b> Half Total Error Rate: 1 - BCR <b>F-measure</b> harmonic mean between precision and recall $2 (precision \cdot recall) / (precision + recall)$ <b><math>F_\beta</math>-measure</b> weighted harmonic mean between precision and recall $(1+\beta)^2 TP / ((1+\beta)^2 TP + \beta^2 FN + FP)$  The harmonic mean between specificity and sensitivity is also often used and sometimes referred to as F-measure.																									
<b>Youden's index:</b> arithmetic mean between sensitivity and specificity sensitivity - (1 - specificity)  <b>Matthews correlation</b> correlation between the actual and predicted $(TP \cdot TN - FP \cdot FN) / ((TP+FP)(TP+FN)(TP + FP)(TN+FN))^{1/2}$ comprised between -1 and 1  <b>Discriminant power</b> normalised likelihood index $\sqrt{3} / \pi \cdot (\log(sensitivity / (1 - specificity)) + \log(specificity / (1 - sensitivity)))$ $<1 = \text{poor}, >3 = \text{good}, \text{fair otherwise}$																									
<b>Graphical tools</b>																									
<b>ROC curve</b> receiver operating characteristic curve : 2-D curve parametrized by one parameter of the classification algorithm, e.g. some threshold in the « true positive rate / false positive rate » space <b>AUC</b> The area under the ROC is between 0 and 1																									
																									
<b>(Cumulative) Lift chart</b> plot of the true positive rate as a function of the proportion of the population being predicted positive, controlled by some classifier parameter (e.g. a threshold)																									
																									
<b>Relationships</b>																									
sensitivity = recall = true positive rate specificity = true negative rate $BCR = \frac{1}{2} \cdot (sensitivity + specificity)$ $BCR = 2 \cdot \text{Youden's index} - 1$ $F\text{-measure} = F_1\text{measure}$ Accuracy = 1 - error rate																									
<b>References</b>																									
Sokolova, M. and Lapalme, G. 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manage. 45, 4 (Jul. 2009), 427-437. Demsar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7 (2006) 1-30																									

# PERFORMANCE MEASURES

- Confusion matrix is basis for many
- Useful for skewed class priors
- Properties of set rather than instance



# LEAVE ONE OUT VALIDATION

**Theorem:** LOO error is an almost unbiased estimate [Luntz&Brailovsky, 1969]:

$$E\left[err\left(f_{m-1}(X)\right)\right] = E\left[LOO_m\right]$$

