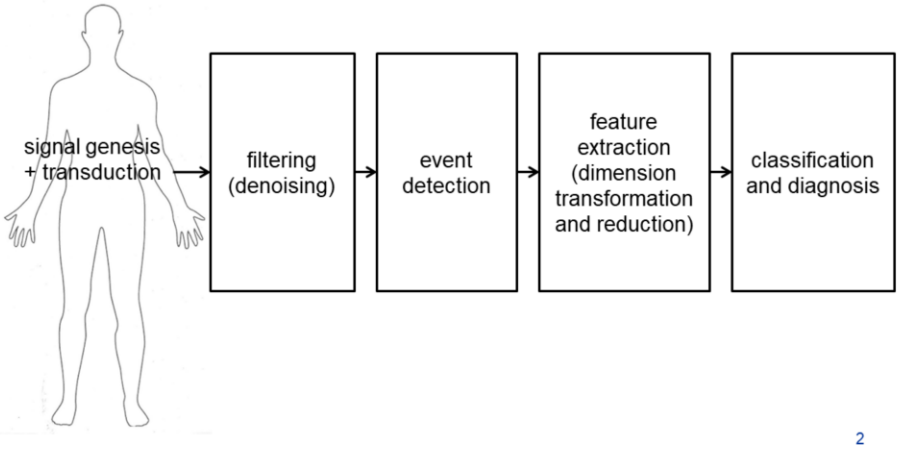
Pázmány Péter Catholic University
Faculty of Information Technology and Bionics

Biomedical Signal Processing
2018-2019 Autumn

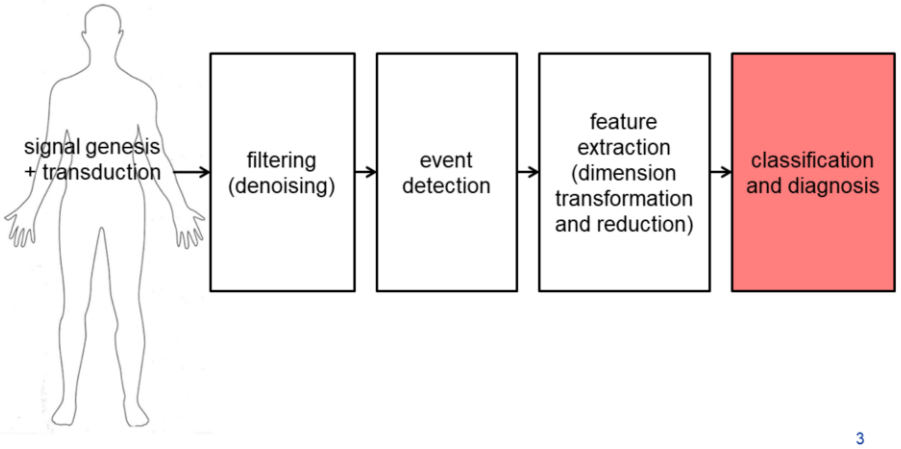
Classification

Lecturer: Janka Hatvani
Responsible lecturer: dr. Miklós Gyöngy

The BSP Flow Chart



The BSP Flow Chart

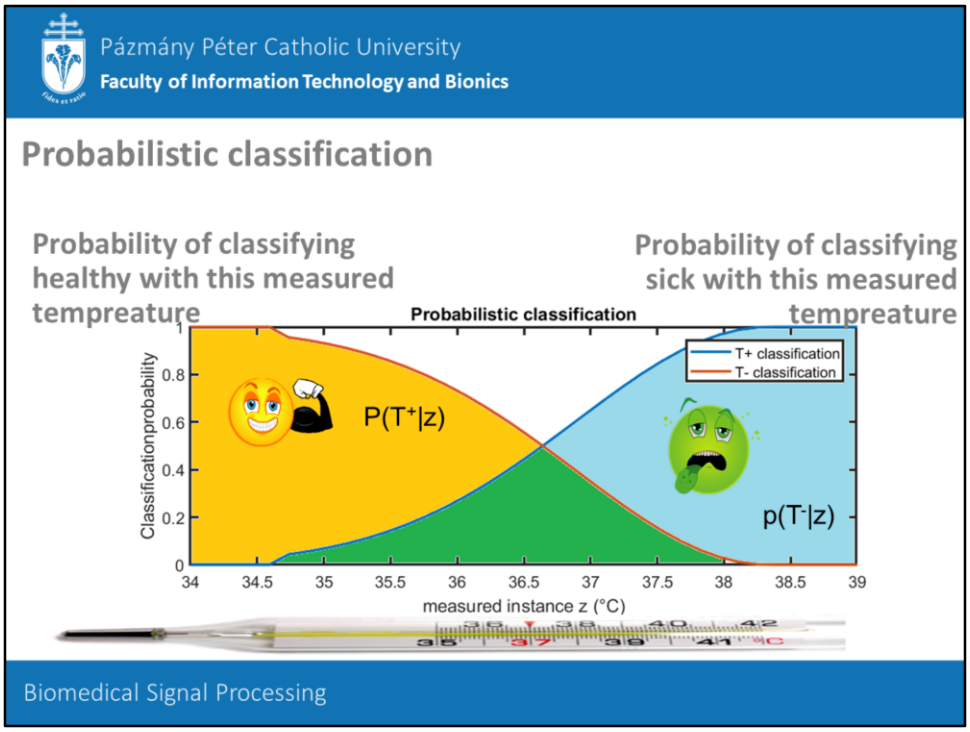


Motivation

- **Need to make classification of patient state to**
  - make a diagnosis
  - decide on a course of treatment based on the diagnosis
- **For more complex classification, computer-aided diagnosis (CAD) may be preferable**
  - does not get overwhelmed with number of variables
  - its performance is easier to evaluate
  - however: may not take into account all circumstances. This is why final diagnosis (and responsibility for it) still falls on doctor

Overview

- The classification problem
- Examples of classification
- Types of classification
- Bayesian classification
- Cost matrix (implicit, explicit)
  - the inability to escape from having to exercise moral judgment



z: measured temperature

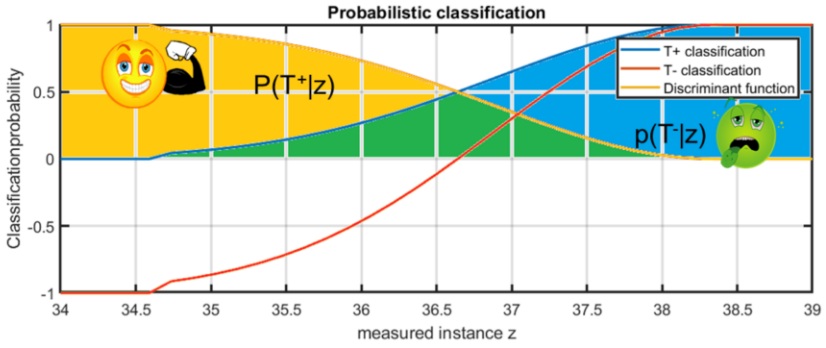
probabilistic classifier is a classifier that is able to predict, given an observation of an input, a probability distribution over a set of classes, rather than only outputting the most likely class that the observation should belong to. (Wikipedia)

$$\boldsymbol{P(T^{+}|z) + P(T^{-}|z) = 1 \quad \forall z}$$

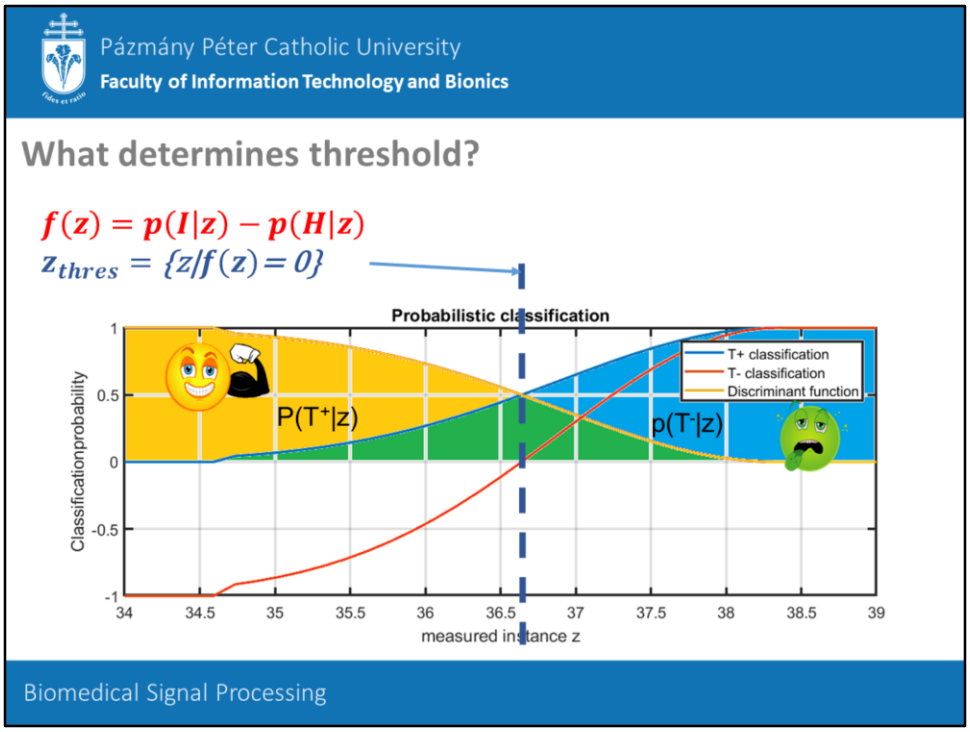
What determines threshold?

discriminant function:

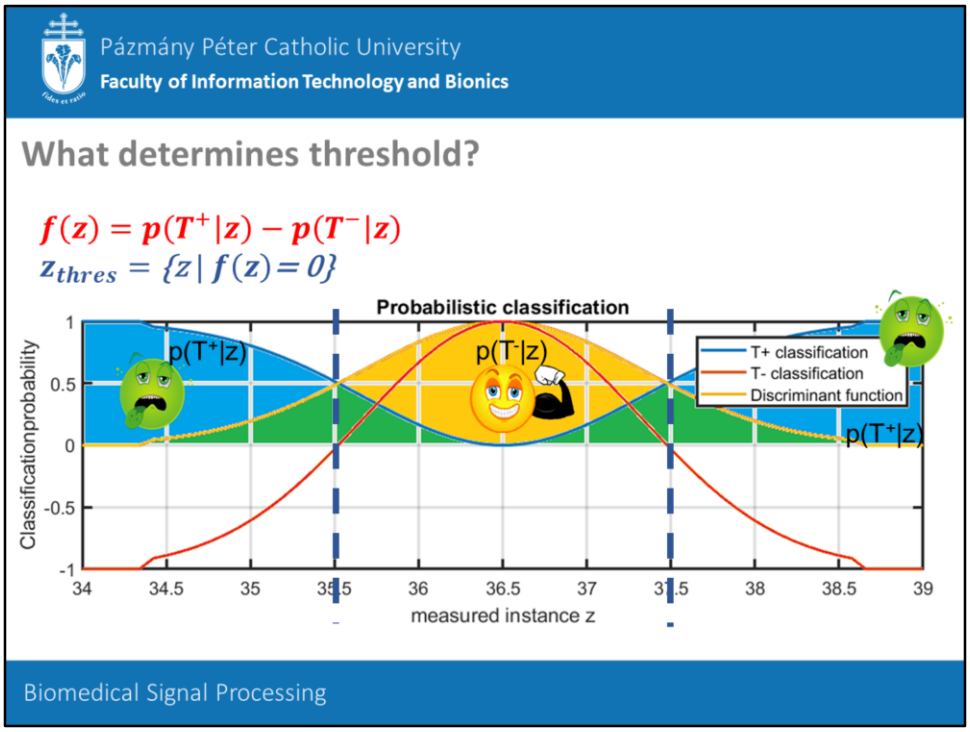
$$f(z) = p(T^+|z) - p(T^-|z)$$



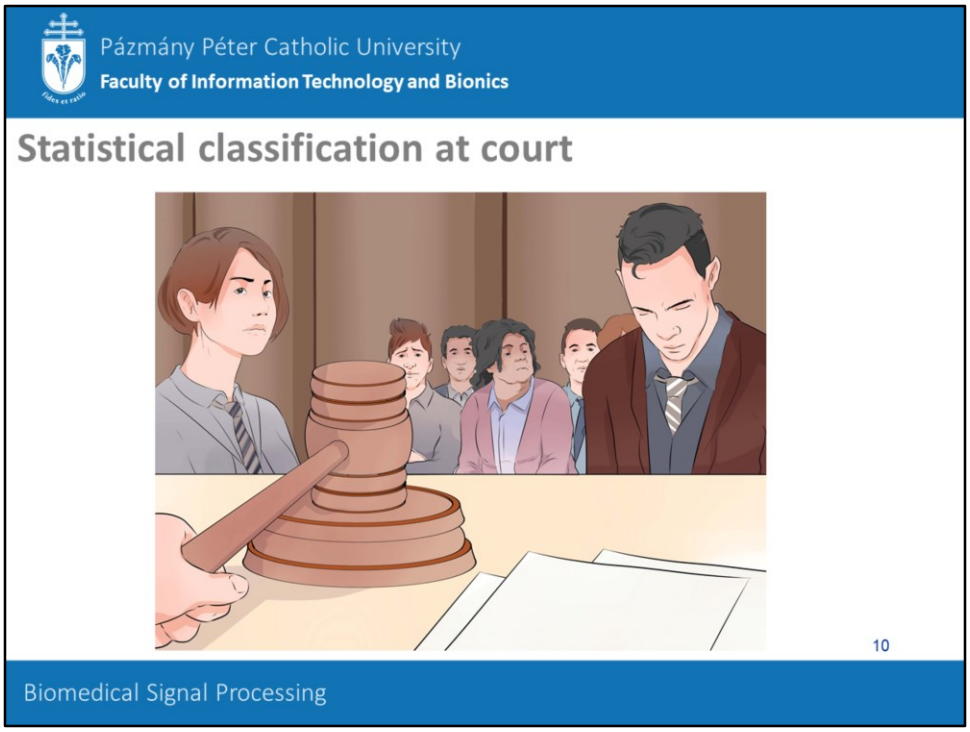
Let's difene a difference function f(z) for upcoming calculations.



Where the difference function f(z) is zero, I can define a threshold!
Temperature under 37 degrees indicate a healthy, above it a sick person.

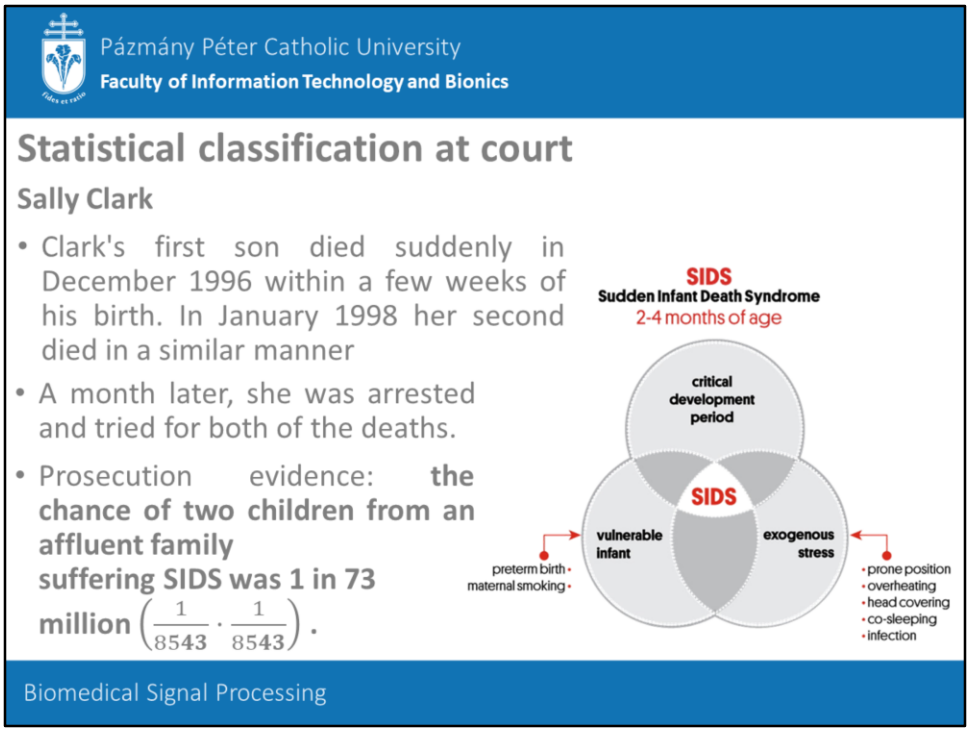


What if A low temperature indicates sickness too?
As the threshold is defined as a set of points, I might have multiple thresholds!



Fenton and Neil: Avoiding Probabilistic Reasoning Fallacies in Legal Practice using Bayesian Networks

Notable examples (of need for Bayesian reasoning):
Sally Clark: „1 in 73 million” (8543 x 8543) chance of her innocence
http://en.wikipedia.org/wiki/Sally_Clark
DNA-based conviction in rape case http://en.wikipedia.org/wiki/R_v_Adams



~700 000 newborns/year in the UK

Fenton and Neil: Avoiding Probabilistic Reasoning Fallacies in Legal Practice using Bayesian Networks

Notable examples (of need for Bayesian reasoning):
Sally Clark: „1 in 73 million" (8543 x 8543) chance of her innocence
http://en.wikipedia.org/wiki/Sally_Clark
DNA-based conviction in rape case http://en.wikipedia.org/wiki/R_v_Adams

Pázmány Péter Catholic University
Faculty of Information Technology and Bionics

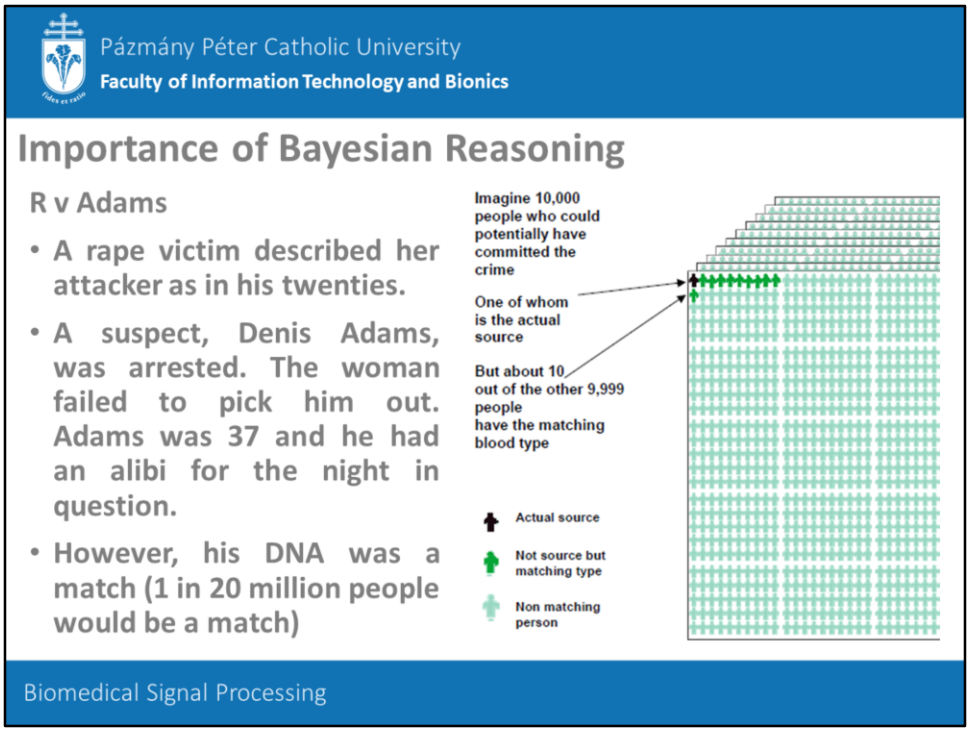
Importance of Bayesian Reasoning

Sally Clark

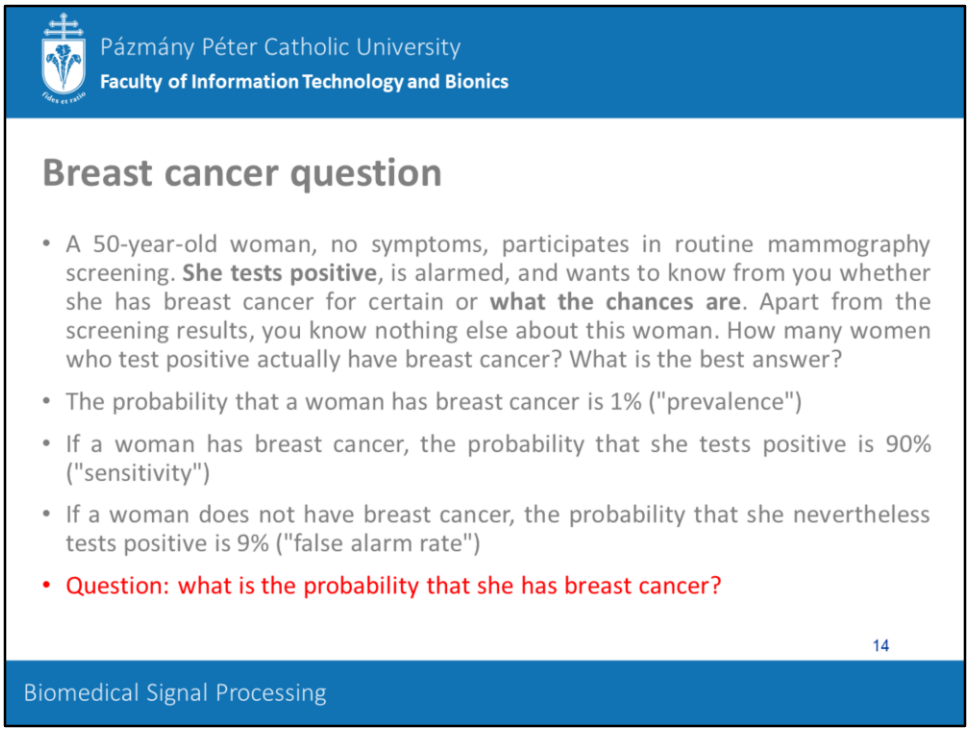
- She was **released from prison having served more than three years** of her sentence.
- Clark's experience caused her to develop serious psychiatric problems and she **died in her home in March 2007** from alcohol poisoning
- Journalists called Clark's experience **"one of the great miscarriages of justice in modern British legal history".**

Biomedical Signal Processing

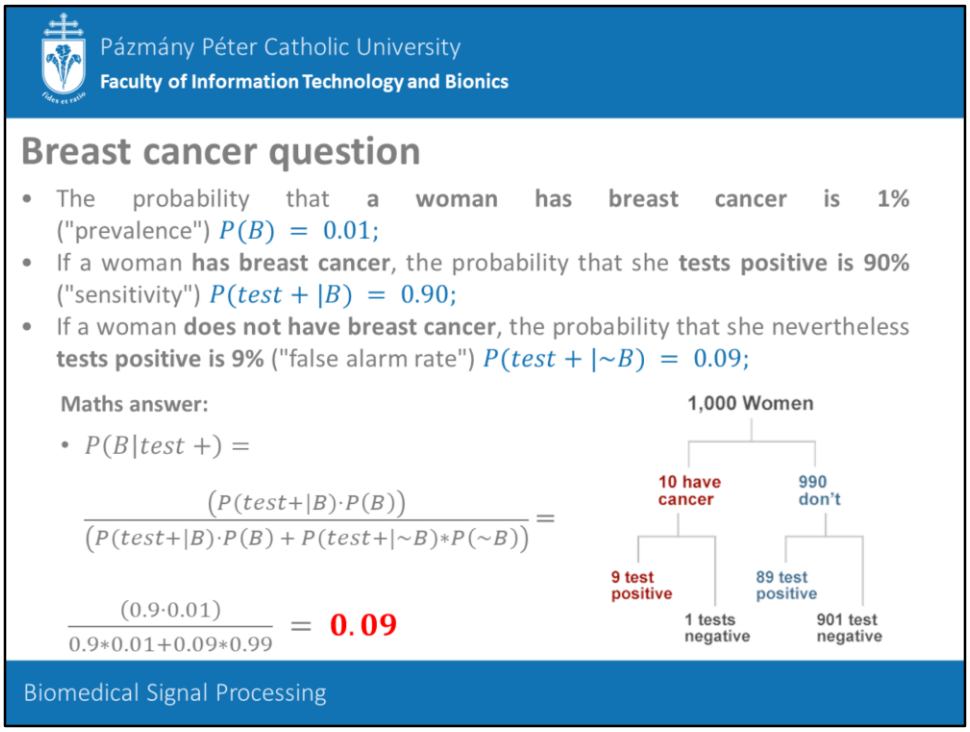
- The SIDSs within one family are not independent events, therefor their joint probability should not be calculated like this
- „double SIDS is very rare, double infant murder is likely to be rarer still" → Bayesian reasoning
- the very same factors which make a family low risk for cot death also make it low risk for murder.



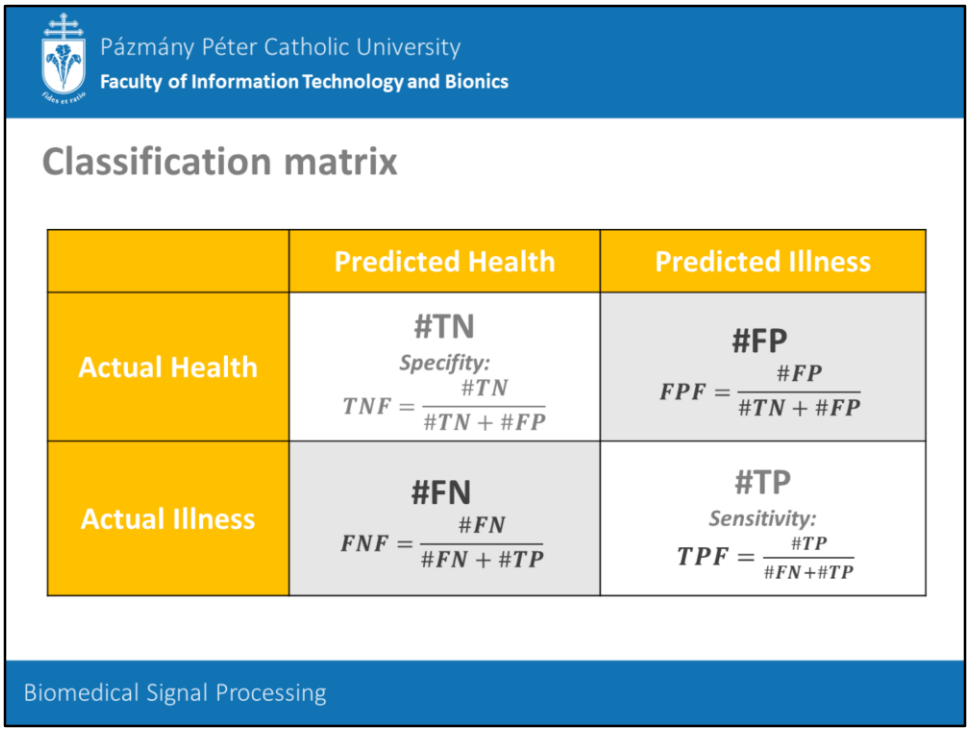
- The SIDSs within one family are not independent events, therefor their joint probability should not be calculated like this
- „double SIDS is very rare, double infant murder is likely to be rarer still" → Bayesian reasoning
- the very same factors which make a family low risk for cot death also make it low risk for murder.



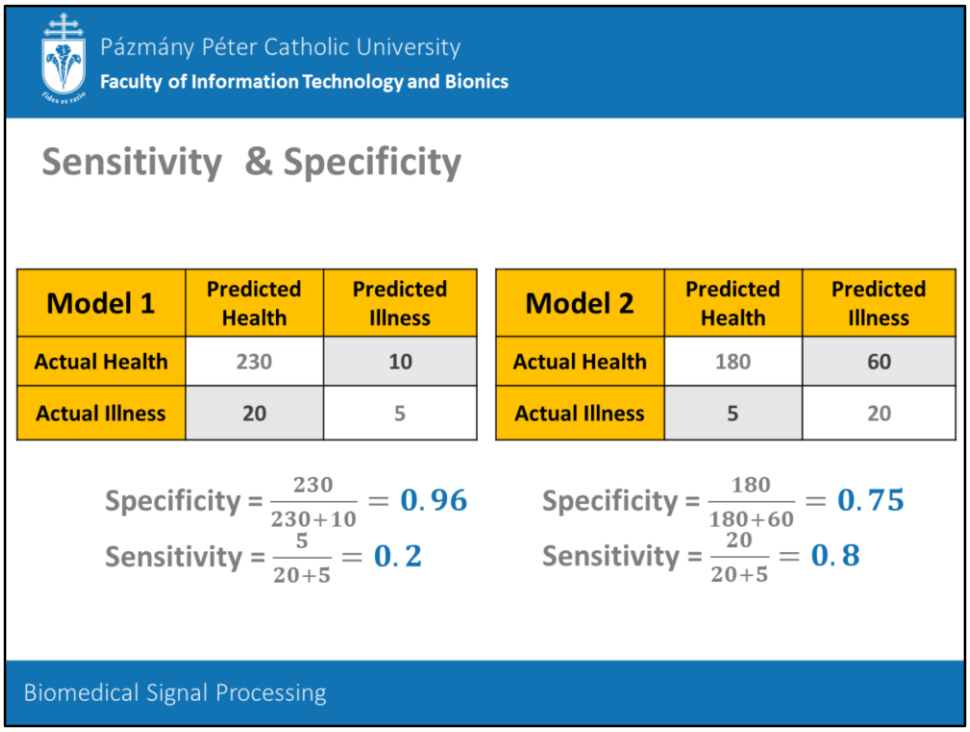
From http://www.bbc.com/news/magazine-28166019



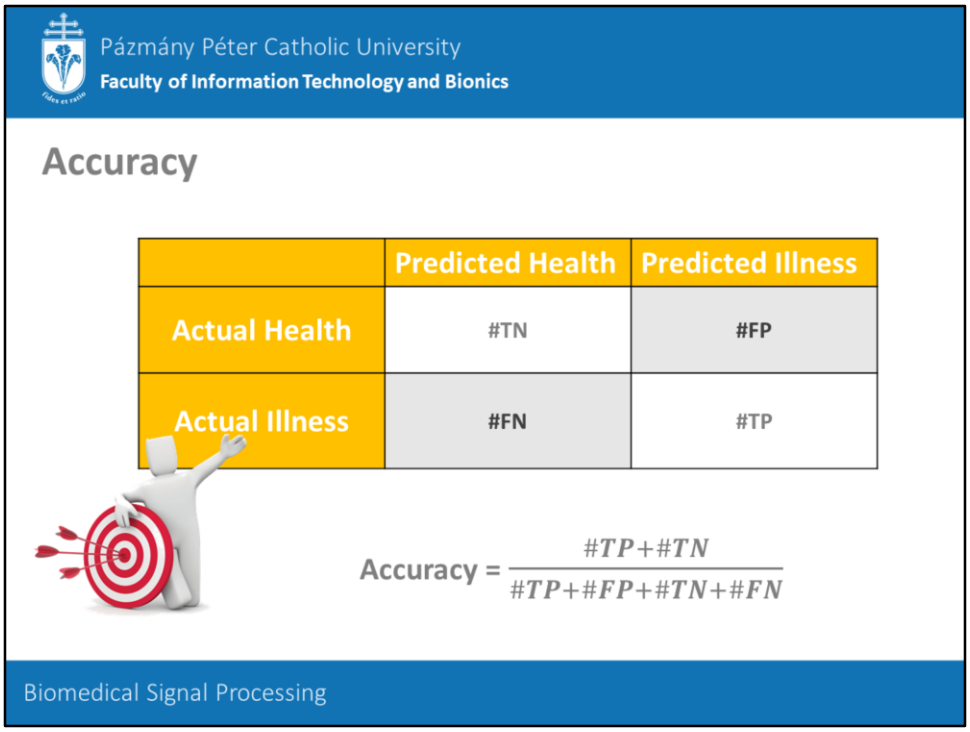
http://news.bbcimg.co.uk/media/images/76096000/gif/_76096068_breast_cancer_304.gif



References:
Rangayyan p. 469
Ling and Sheng: Cost-Sensitive Learning and the Class Imbalance Problem

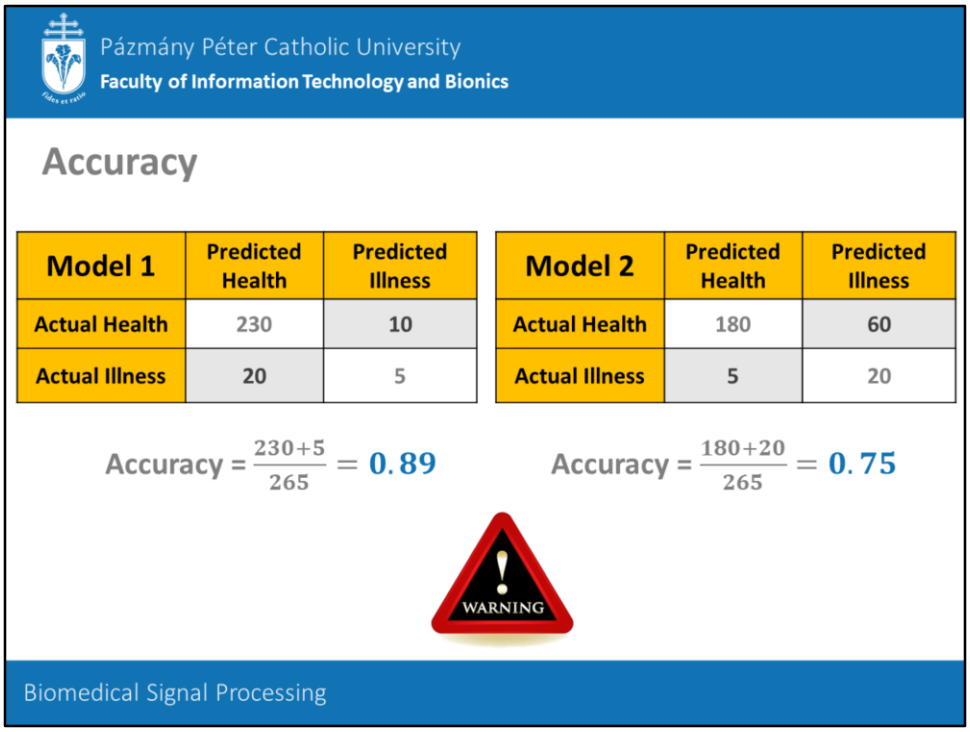


References:
Rangayyan p. 469
Ling and Sheng: Cost-Sensitive Learning and the Class Imbalance Problem



An overall performance of the classifier can be defined as the accuracy.

References:
Rangayyan p. 469
Ling and Sheng: Cost-Sensitive Learning and the Class Imbalance Problem



Model 1 classifies almost all test as negative, still has a higher accuracy.
Model 2 classifies the abn

In the breast cancer example the positive set is 1% compared to the 99% N test. If the positive cases would be classified as negatives, the accuracy would be still high.

References:
Rangayyan p. 469
Ling and Sheng: Cost-Sensitive Learning and the Class Imbalance Problem

Pázmány Péter Catholic University
Faculty of Information Technology and Bionics

Cost matrix

	Predicted Health	Predicted Illness
Actual Health	C(TN): cost of screening (including complications)	C(FP): cost of further tests/therapy therapy on healthy individual
Actual Illness	C(FN): cost of delayed therapy	C(TP): cost of further tests/therapy

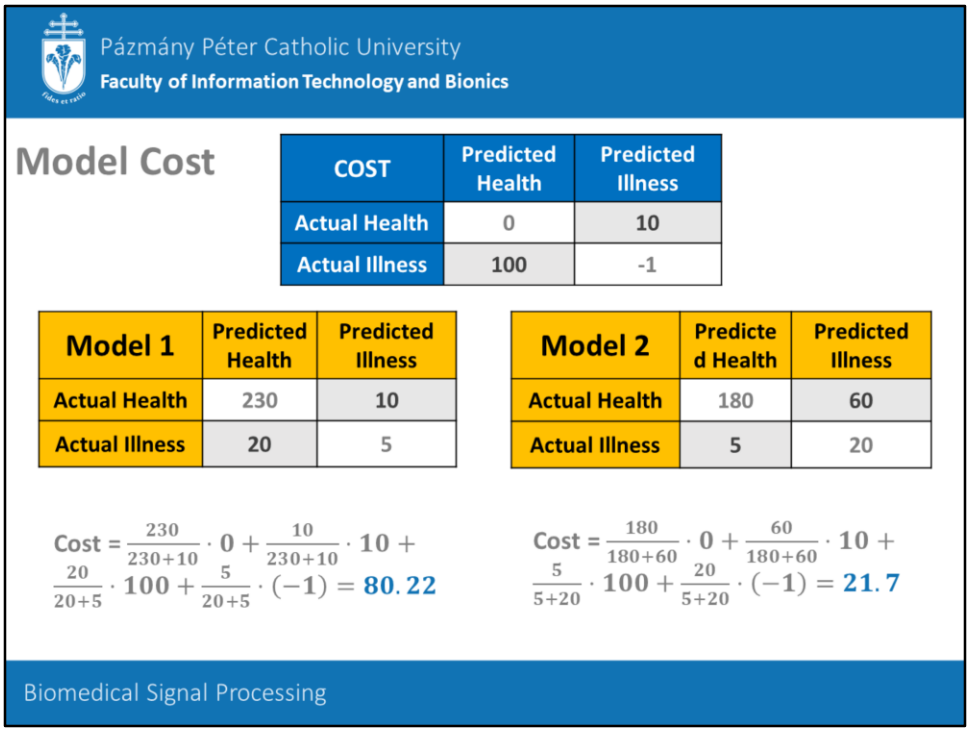
$$Cost = TNF \cdot C(TN) + FPF \cdot C(FP) + FNF \cdot C(FN) + TPF \cdot C(TP)$$

Biomedical Signal Processing

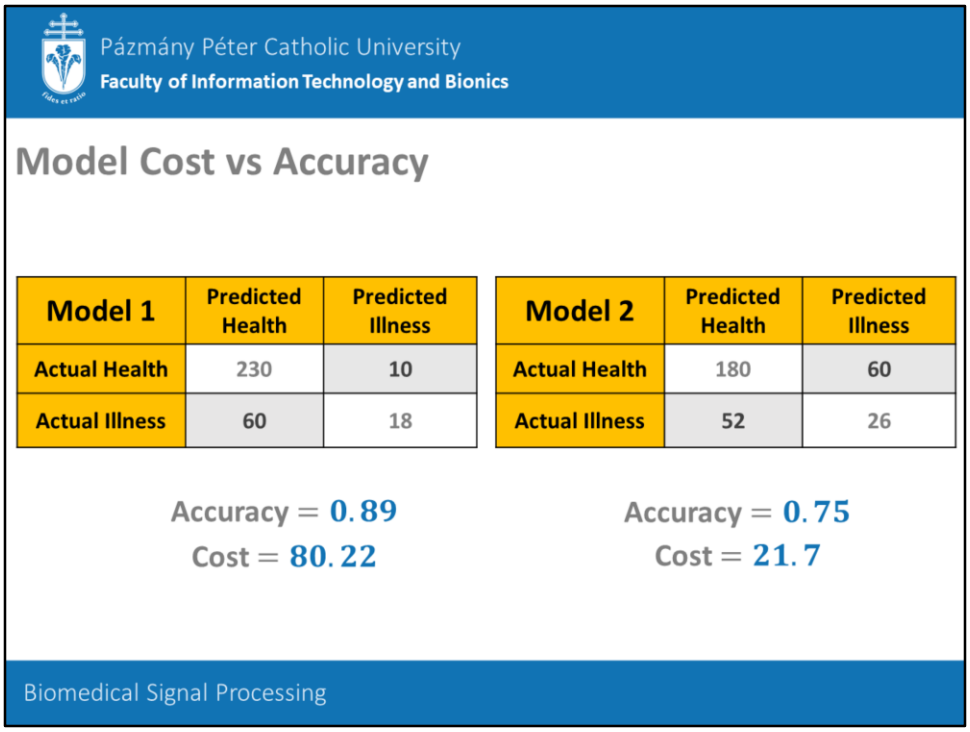
So far we have made diagnostic decisions based on the implicit assumption that the cost of a false positive and false negative are the same.
This leads to us choosing a positive diagnosis if the probability of a positive diagnosis is greater than the probability of a negative diagnosis.
However, what if there are difference costs associated with false positives, false negatives? Or maybe there are even costs to a true positive or true negative?

These misclassification cost values can be given by domain experts, or learned via other approaches. In cost-sensitive learning, it is usually assume that such a cost matrix is given and known.

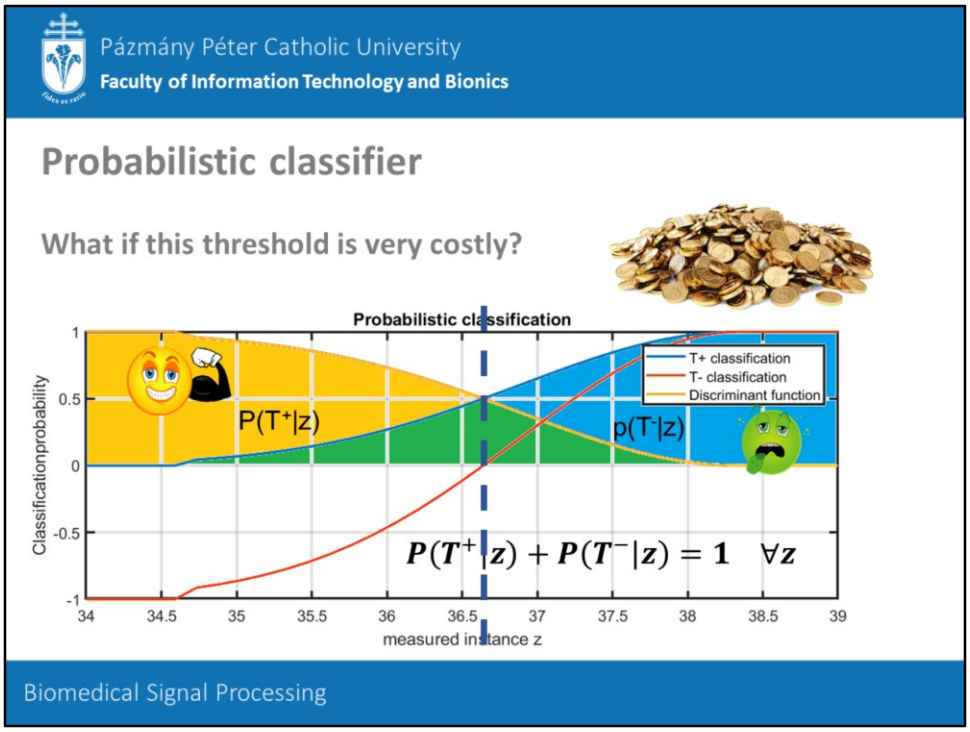
References:
Rangayyan p. 469
Ling and Sheng: Cost-Sensitive Learning and the Class Imbalance Problem



References:
Rangayyan p. 469
Ling and Sheng: Cost-Sensitive Learning and the Class Imbalance Problem



„If we compare both the models and if we check their accuracy. Accuracy for Model 1 is higher compared to Model 2, however the cost for Model 1 is higher compared to Model 2.
So it depends on what kind of problem statement we are facing.
If we are focusing on accuracy then we will go with the Model 1 (In this case we need to compromise on cost) , however if we are focusing on cost then we will go with the Model 2 (In this case we need to compromise on accuracy).”



probabilistic classifier is a classifier that is able to predict, given an observation of an input, a probability distribution over a set of classes, rather than only outputting the most likely class that the observation should belong to. (Wikipedia)

$$P(T^+|z) + P(T^-|z) = 1 \quad \forall z$$

Cost minimisation

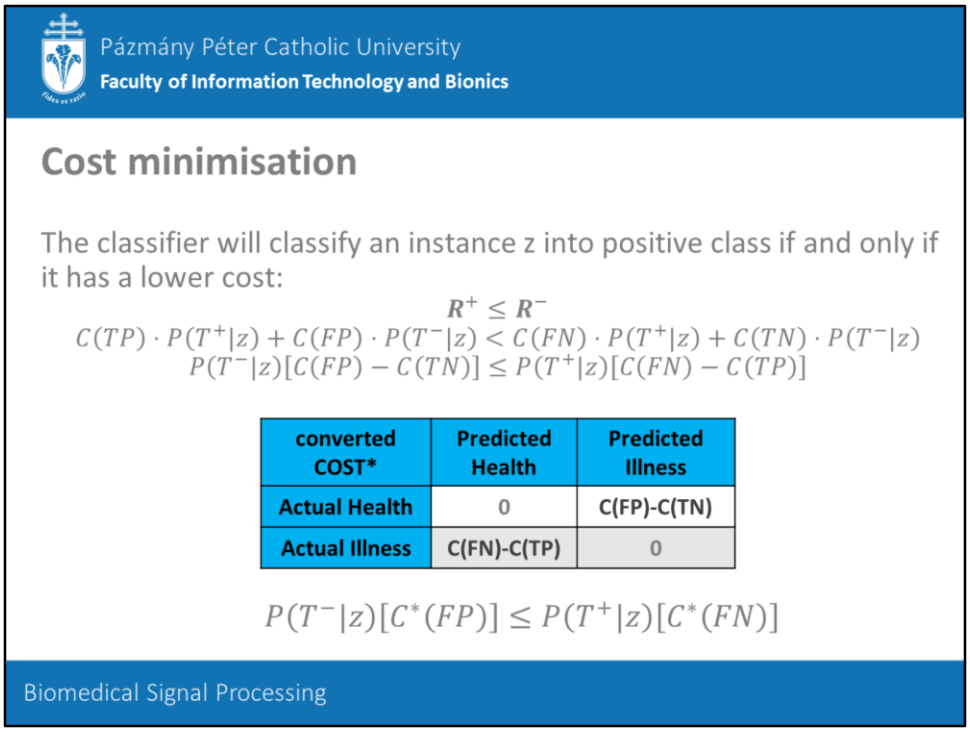
How to minimize the cost of a classification problem?

The expected cost of classifying a case z into positive/negative class:

$$R(T^+|z) = C(TP) \cdot P(T^+|z) + C(FP) \cdot P(T^-|z)$$
$$R(T^-|z) = C(FN) \cdot P(T^-|z) + C(TN) \cdot P(T^-|z)$$

COST	Predicted Health	Predicted Illness
Actual Health	C(TN)	C(FP)
Actual Illness	C(FN)	C(TP)

$P(T^-\|z)$	Probability estimation of classifying a case z as healthy
$P(T^+\|z)$	Probability estimation of classifying a case z as healthy

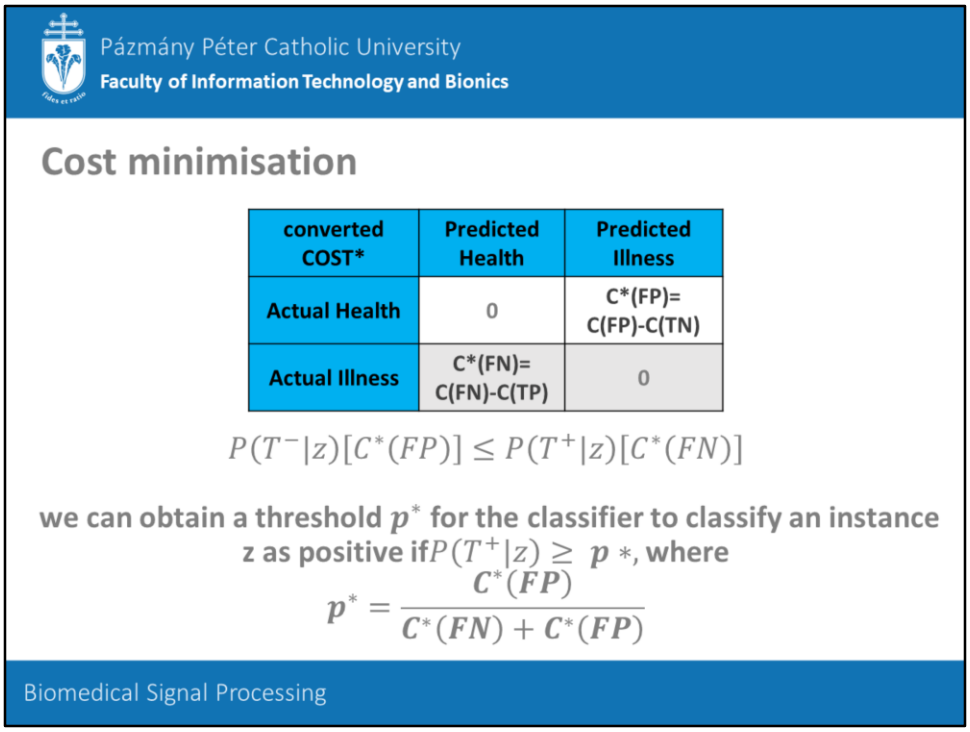


Given the cost matrix, an example should be classified into the class that has the minimum expected cost. This is the minimum expected cost principle.

The line, $P(T^-|z)[C(FP) - C(TN)] \leq P(T^+|z)[C(FN) - C(TP)]$ indicates that the decision (of classifying an example into positive) will not be changed if a constant is added into a row of the original cost matrix. Thus, the original cost matrix can always be converted to a simpler one by subtracting C(TN) from the first row, and C(TP) from the second row

Under this assumption, the classifier will classify a case into positive class if and only if:

$$(T^-|z)[C^*(FP)] \leq P(T^+|z)[C^*(FN)]$$



Thus, if a cost-insensitive classifier can produce a posterior probability estimation p(I|z) for test examples z, we can make it cost-sensitive by setting the classification threshold p* to the limit, where

$$P(T^-|z)[C^*(FP)] = p^*[C^*(FN)]$$

However, here $P(T^+|z)$ should be (1-p*):

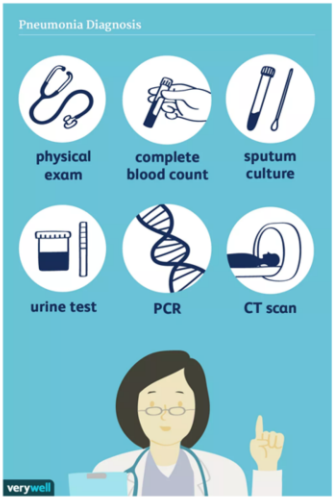
$$(1-p^*)[C^*(FP)] = p^*[C^*(FN)]$$
$$C^*(FP) = p^*\left(C^*(FN) + C^*(FP)\right)$$
$$p^* = \frac{C^*(FP)}{\left(C^*(FN)+C^*(FP)\right)}$$

$$p^* = \frac{C^*(FN)}{C^*(FN) + C^*(FP)}$$

and classify any example to be positive whenever $P(T^+|z)$≥ p* .

Multi-stage diagnosis

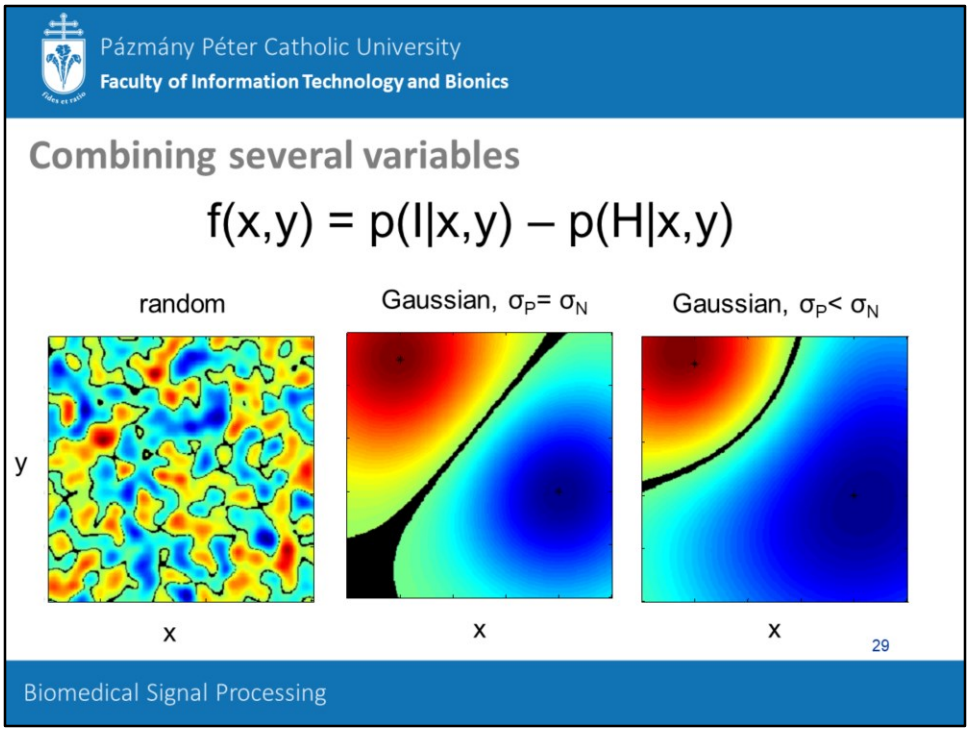
- Use more economic ways of diagnosis first
- Generally (unless it is a round of eliminating diseases), cost (and probability) of diagnosis increases as more and more tests are performed



https://www.verywellhealth.com/diagnosis-of-pneumonia-4160855

Inevitability of having to exercise ethical/moral/practical judgment

- **What cost to put on someone's life**
  - years of „useful life"
- **What cost to put on someone's discomfort**
  - physical
  - psychological
- **Resource allocation – different area**



The first one ca cnot be separated, the second one can be separated linearly.
With the third one a non-linear classification might be applied.

Or again, weighted substraction if we have unequal costs associated.
But more-generally, what approaches are taken?

How to combine? Classification types

- **Logic-based, rule-based**
- **Pseudo-probabilistic**
  - confidence factor (cf MYCIN)
  - score definition, e.g. using simple counts of positive and negative factors
  - fuzzy logic
- **Probabilistic**
  - log-likelihood
  - Bayesian (and Bayesian/Belief networks)
- **ANN-based methods**
  - Feed-forward vs Recurrent Neural networks
  - Support vector machines

Rule-based classification

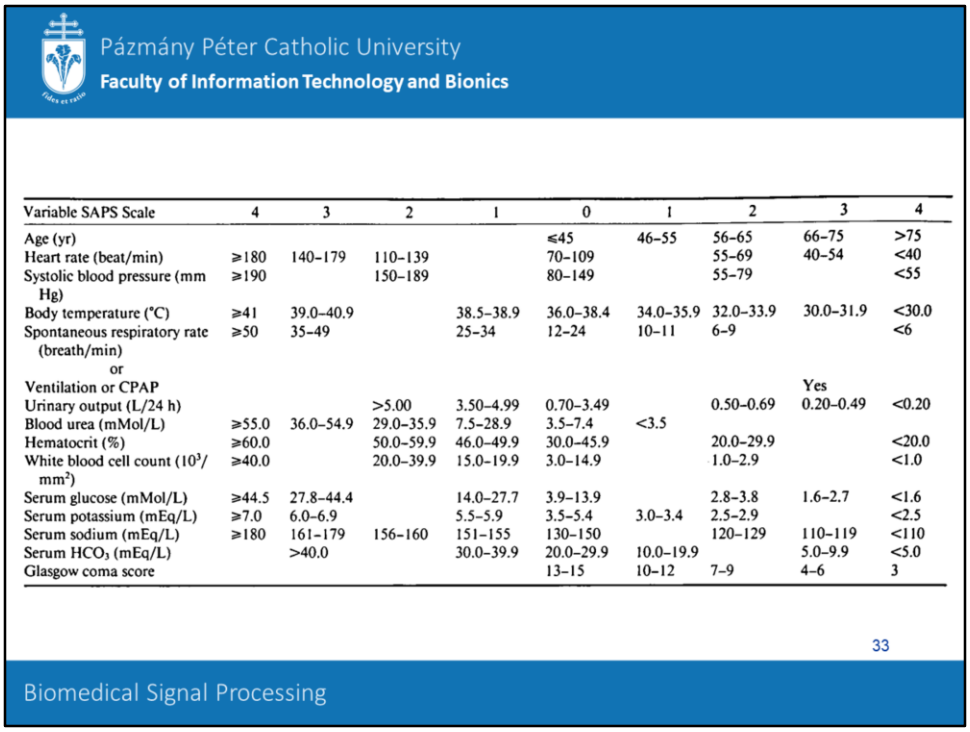
IF (the RR interval of the beat is less than the normal at the current heart rate) AND (the QRS waveshape is markedly different from the normal QRS of the patient) THEN *the beat is a PVC.*

Rangayyan (2002): p. 446-448

Rule-based classification

IF (QRS duration $\geq$ 105 *ms* and $\leq$ 120 *ms*) AND
(QRS amplitude is negative in leads V1 and V2) AND
(Q or S duration $\geq$ 80 *ms* in leads V1 and V2) AND
(no Q wave is present in any two of leads I, V5, and V6) AND
(R duration $>$ 60 *ms* in any two of leads I, aVL, V5, and V6) THEN
the patient has incomplete left bundle-branch block.

Rangayyan (2002): p. 446-448

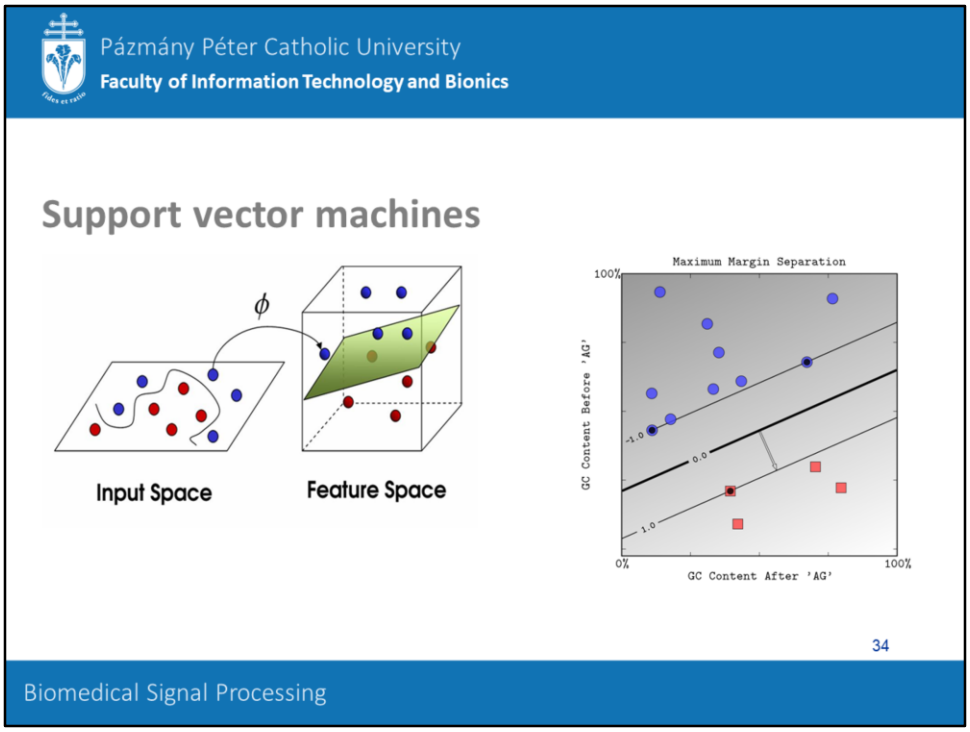


Variable SAPS Scale	4	3	2	1	0	1	2	3	4
Age (yr)					≤45	46–55	56–65	66–75	>75
Heart rate (beat/min)	≥180	140–179	110–139		70–109		55–69	40–54	<40
Systolic blood pressure (mm Hg)	≥190		150–189		80–149		55–79		<55
Body temperature (°C)	≥41	39.0–40.9		38.5–38.9	36.0–38.4	34.0–35.9	32.0–33.9	30.0–31.9	<30.0
Spontaneous respiratory rate (breath/min)	≥50	35–49		25–34	12–24	10–11	6–9		<6
or									
Ventilation or CPAP								Yes	
Urinary output (L/24 h)			>5.00	3.50–4.99	0.70–3.49		0.50–0.69	0.20–0.49	<0.20
Blood urea (mMol/L)	≥55.0	36.0–54.9	29.0–35.9	7.5–28.9	3.5–7.4	<3.5			
Hematocrit (%)	≥60.0		50.0–59.9	46.0–49.9	30.0–45.9		20.0–29.9		<20.0
White blood cell count (10^3/ mm^2)	≥40.0		20.0–39.9	15.0–19.9	3.0–14.9		1.0–2.9		<1.0
Serum glucose (mMol/L)	≥44.5	27.8–44.4		14.0–27.7	3.9–13.9		2.8–3.8	1.6–2.7	<1.6
Serum potassium (mEq/L)	≥7.0	6.0–6.9		5.5–5.9	3.5–5.4	3.0–3.4	2.5–2.9		<2.5
Serum sodium (mEq/L)	≥180	161–179	156–160	151–155	130–150		120–129	110–119	<110
Serum HCO_3 (mEq/L)		>40.0		30.0–39.9	20.0–29.9	10.0–19.9		5.0–9.9	<5.0
Glasgow coma score					13–15	10–12	7–9	4–6	3

How to classify ICU patients? One example is giving them a score based on the table above.

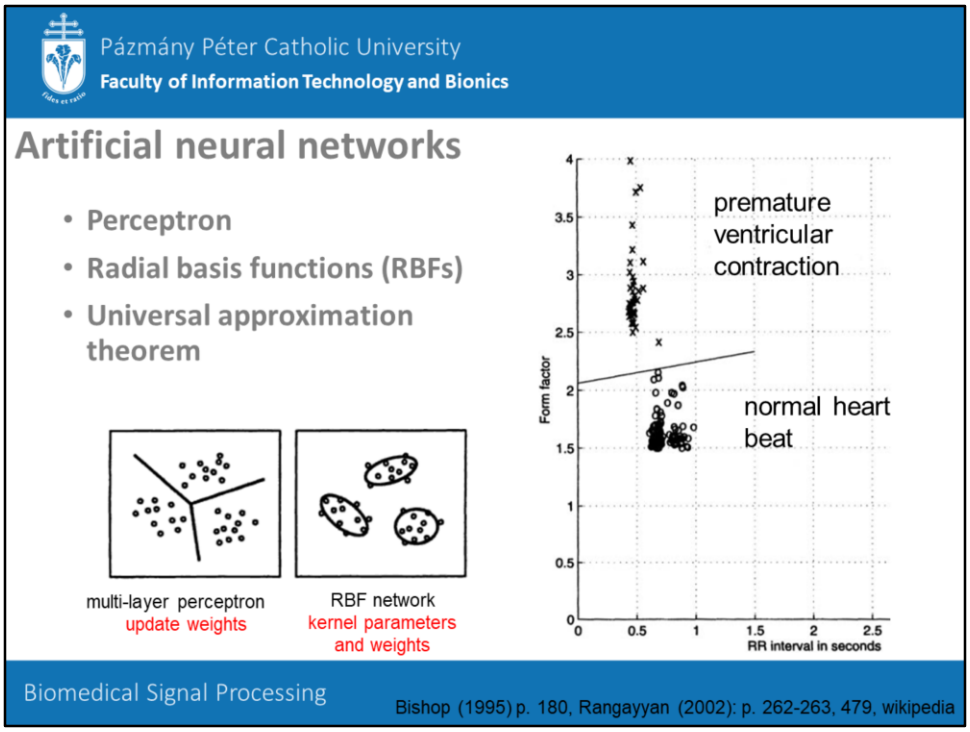
Other scores: APGAR, Manning

References:
Le Gall et al. (1984): A simplified acute physiology score for ICU patients
http://en.wikipedia.org/wiki/ICU_scoring_systems
https://en.wikipedia.org/wiki/Apgar_score
https://en.wikipedia.org/wiki/Biophysical_profile



Outside scope:
Relevance vector machines: SVMs with probabilistic output

Sources:
http://www.imtech.res.in/raghava/rbpred/svm.jpg
http://svmcompbio.tuebingen.mpg.de/examples.html



Perceptron: possibility of non-linear mapping following linear combination of inputs

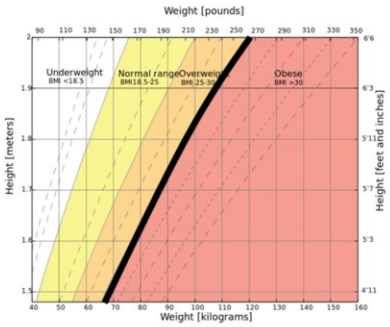
Radial basis functions (RBFs): weighted sum of radial basis function outputs, each of which is a function of the distance of the input vector from some center: $\varphi(\mathbf{x})=\sum 8\_i \llbracket a\_i\, \rho(\|\mathbf{x}-\mathbf{c}\_i\|) \rrbracket$, with e.g. $\rho(\|\mathbf{x}-\mathbf{c}\_i\|)=\exp[-\beta\|\mathbf{x}-\mathbf{c}\_i\|\hat{}2]$

Universal Approximation theorem: feed-forward network with single hidden layer able to approximate continuous functions on compact subsets (https://en.wikipedia.org/wiki/Universal_approximation_theorem)

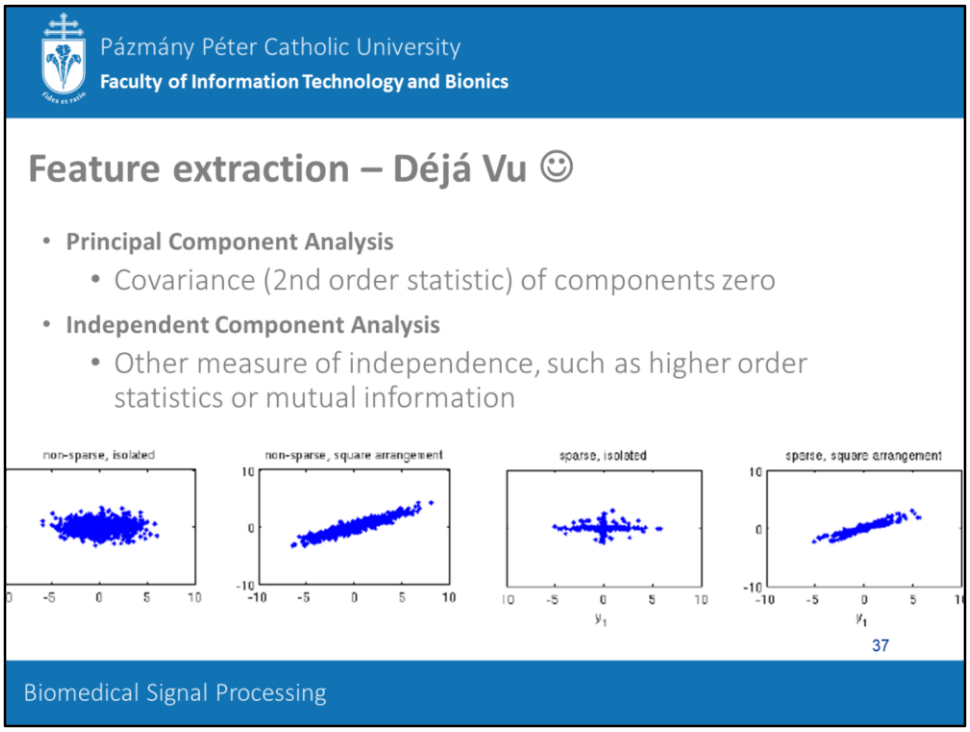
Issues (mainly with ANNs)

- **Curse of dimensionality**
- **Overtraining**
- **Time to run**
- **Main solution:**
  - reduce dimensionality
  - feature extraction

- **How to pick ANN architecture?**
  - self-organising maps (SOM)

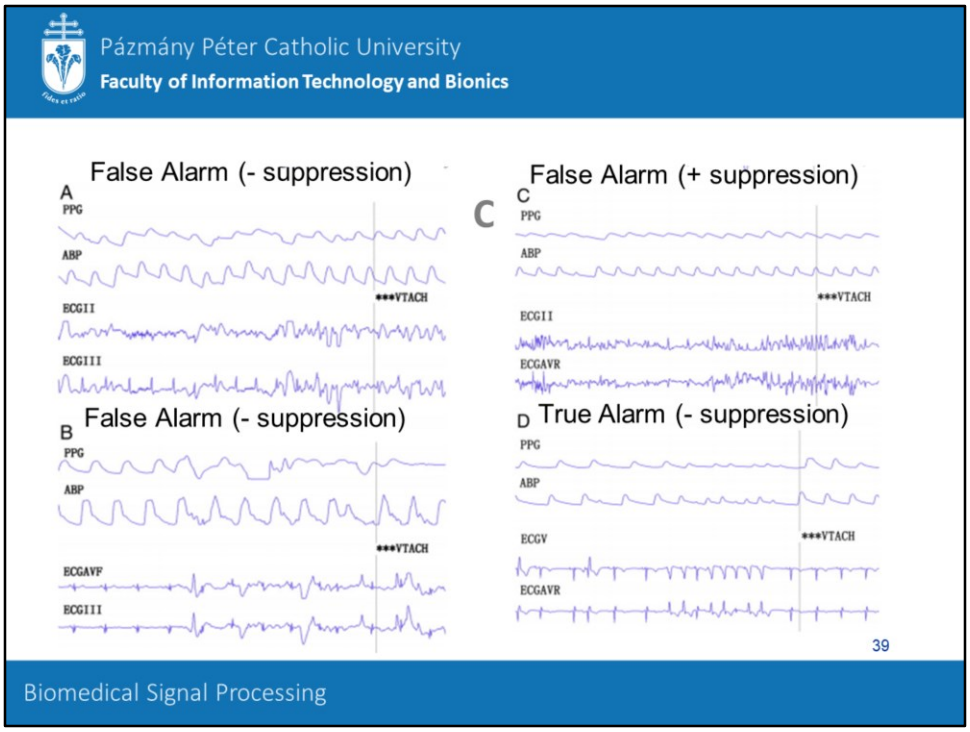


$BMI = mass / height^2$



Comon and Jutten (2010): Handbook of blind source separation: Independent Component Analysis and applications

EXTRA MATERIAL

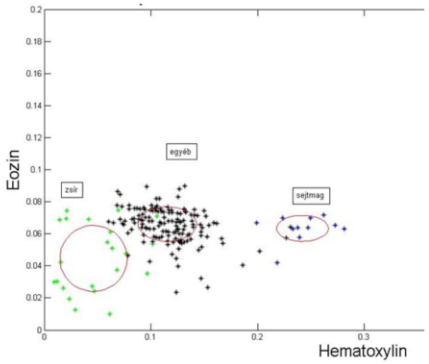


Examples of false and true ventricular tachycardia alarms. Note the vertical line marks the time the alarm sounded. (A and B) False alarms and the algorithm failed to suppress them. (C) A false alarm and is suppressed correctly. (D) A true alarm and is accepted correctly by the algorithm.
When signal quality is high enough for PPG, ABP, then it can suppress alarm

Source:
Li, Qiao, and Gari D. Clifford. "Signal quality and data fusion for false alarm reduction in the intensive care unit." *Journal of electrocardiology*45.6 (2012): 596-603.

Estimation of PDF – Unimodal Gaussian

- Supervised learning
  (vs unsupervised or reinforcement)
- Gaussian est. of p(x | class)
- Bayesian est. of p(class | x)



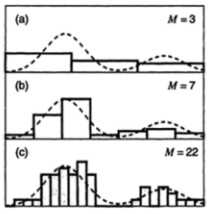
Estimation of PDF – Other

Non-parametric:

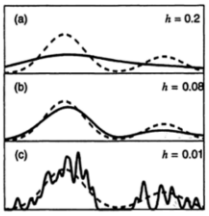
- histogram-based
- kernel-based
- K-nearest neighbour

Mixture model

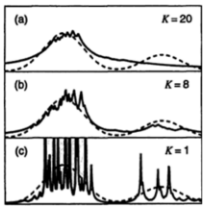
- mixture of Gaussians



histogram-based pdf est.



kernel-based pdf est.



KNN-based pdf est.

Bishop (1995): Neural Networks for Pattern Recognition, p. 49-60