



Biomedical Signal Processing

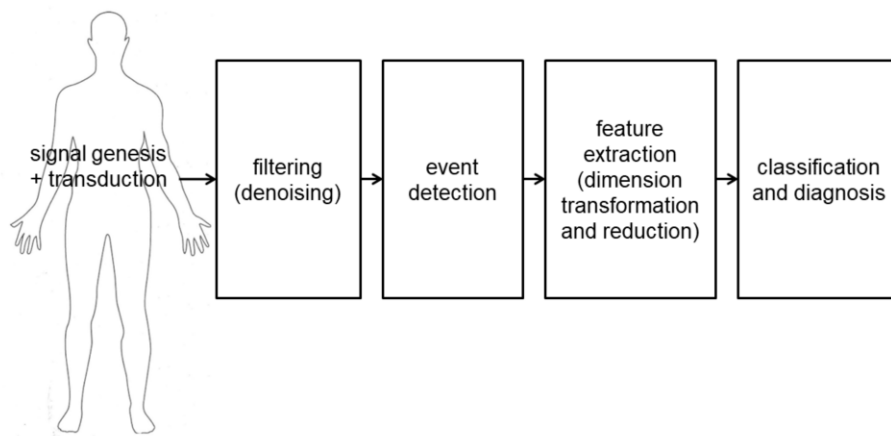
2018-2019 Autumn

Classification

Lecturer: Janka Hatvani
Responsible lecturer: dr. Miklós Gyöngy

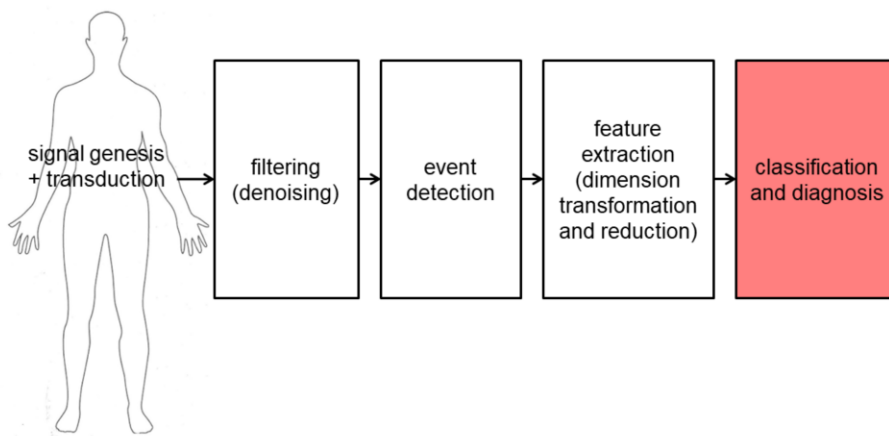


The BSP Flow Chart





The BSP Flow Chart



3



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

4



Overview

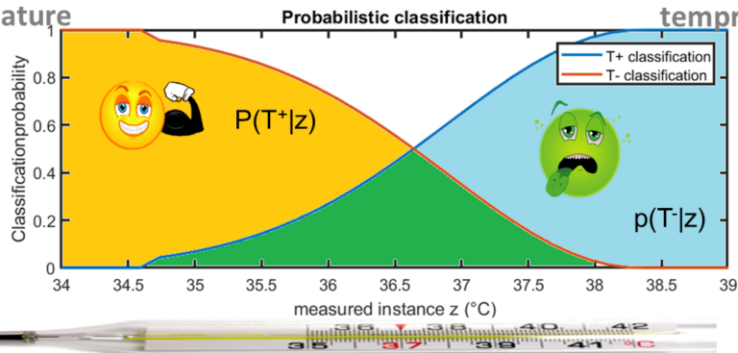
- The classification problem
- Examples of classification
- Types of classification
- Bayesian classification
- Cost matrix (implicit, explicit)
 - the inability to escape from having to exercise moral judgment



Probabilistic classification

Probability of classifying
healthy with this measured
temperature

Probability of classifying
sick with this measured
temperature



Biomedical Signal Processing

z : measured temperature

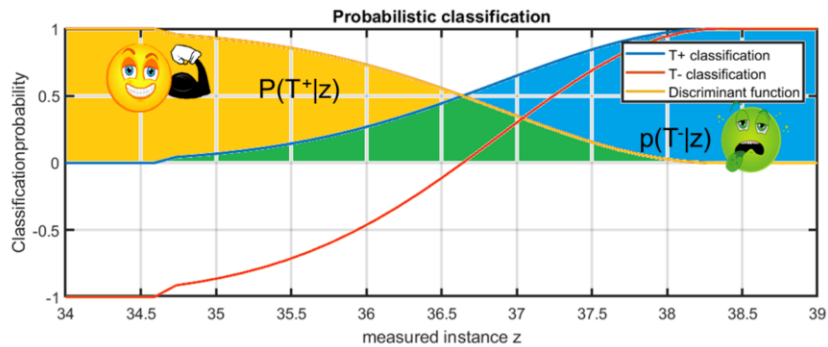
probabilistic classifier is a [classifier](#) that is able to predict, given an observation of an input, a [probability distribution](#) over a [set](#) of classes, rather than only outputting the most likely class that the observation should belong to. (Wikipedia)

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

What determines threshold?

discriminant function:

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



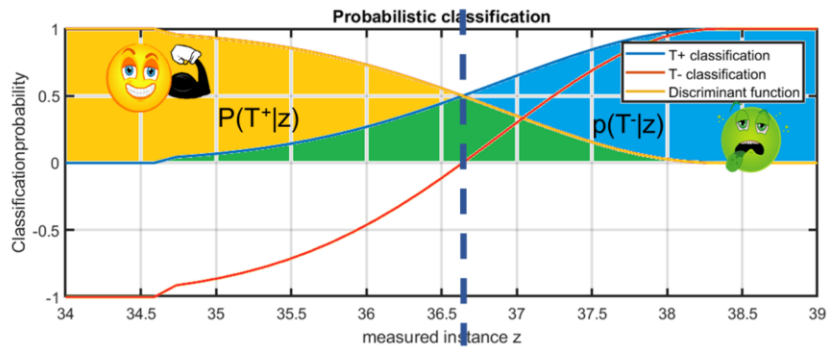
Biomedical Signal Processing

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

What determines threshold?

$$f(z) = p(I|z) - p(H|z)$$

$$z_{thres} = \{z | f(z) = 0\}$$



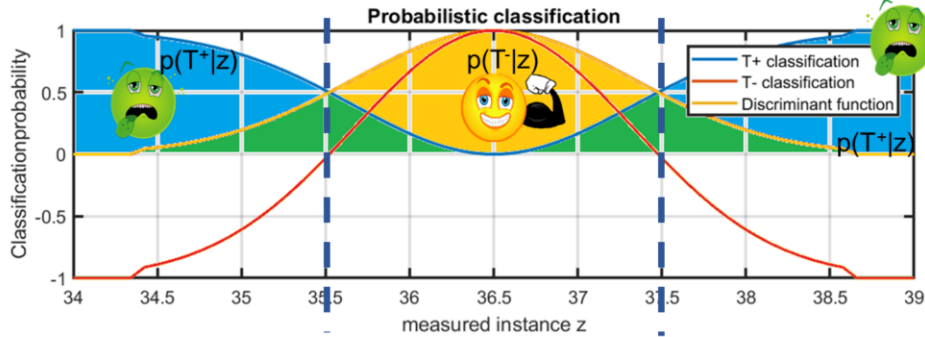
Biomedical Signal Processing

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 determines threshold?

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

$$z_{thres} = \{z \mid f(z) = 0\}$$



Biomedical Signal Processing

What if A low temperature indicates sickness too?

As the threshold is defined as a set of points, I might have multiple thresholds!



Statistical classification at court



10

Biomedical Signal Processing

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” (8543×8543) 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



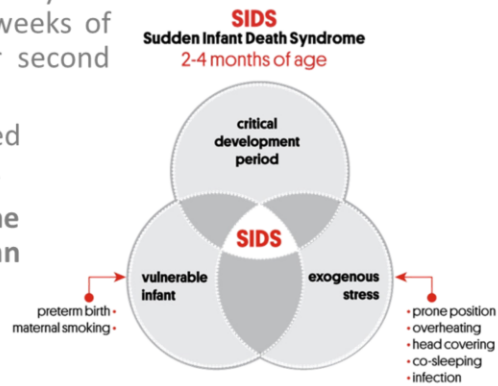
Statistical classification at court

Sally Clark

- Clark's first son died suddenly in December 1996 within a few weeks of his birth. In January 1998 her second died in a similar manner
- A month later, she was arrested and tried for both of the deaths.

- Prosecution evidence: **the chance of two children from an affluent family suffering SIDS was 1 in 73 million**

$$\text{million} \left(\frac{1}{8543} \cdot \frac{1}{8543} \right).$$



Biomedical Signal Processing

~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



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, therefore 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.



Importance of Bayesian Reasoning

R v Adams

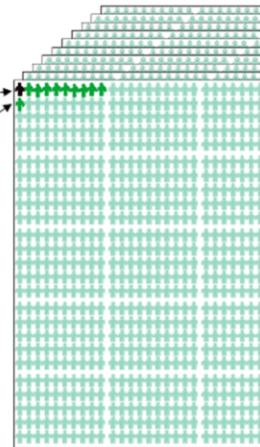
- A rape victim described her attacker as in his twenties.
- A suspect, Denis Adams, was arrested. The woman failed to pick him out. Adams was 37 and he had an alibi for the night in question.
- However, his DNA was a match (1 in 20 million people would be a match)

Imagine 10,000 people who could potentially have committed the crime

One of whom is the actual source

But about 10 out of the other 9,999 people have the matching blood type

- ↑ Actual source
- ↑ Not source but matching type
- ↑ Non matching person



- The SIDSs within one family are not independent events, therefore 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.



Breast cancer question

- A 50-year-old woman, no symptoms, participates in routine mammography screening. **She tests positive**, is alarmed, and wants to know from you whether she has breast cancer for certain or **what the chances are**. Apart from the screening results, you know nothing else about this woman. How many women who test positive actually have breast cancer? What is the best answer?
- The probability that a woman has breast cancer is 1% ("prevalence")
- If a woman has breast cancer, the probability that she tests positive is 90% ("sensitivity")
- If a woman does not have breast cancer, the probability that she nevertheless tests positive is 9% ("false alarm rate")
- **Question: what is the probability that she has breast cancer?**

14

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



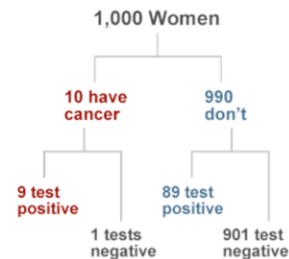
Breast cancer question

- The probability that a woman has breast cancer is 1% ("prevalence") $P(B) = 0.01$;
- If a woman **has breast cancer**, the probability that she **tests positive** is 90% ("sensitivity") $P(\text{test} + | B) = 0.90$;
- If a woman **does not have breast cancer**, the probability that she nevertheless **tests positive** is 9% ("false alarm rate") $P(\text{test} + | \sim B) = 0.09$;

Maths answer:

$$P(B|\text{test} +) =$$

$$\frac{(P(\text{test} + | B) \cdot P(B))}{(P(\text{test} + | B) \cdot P(B) + P(\text{test} + | \sim B) \cdot P(\sim B))} =$$
$$\frac{(0.9 \cdot 0.01)}{0.9 \cdot 0.01 + 0.09 \cdot 0.99} = \mathbf{0.09}$$



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



Classification matrix

| | Predicted Health | Predicted Illness |
|----------------|---|---|
| Actual Health | #TN <i>Specificity:</i> $TNF = \frac{\#TN}{\#TN + \#FP}$ | #FP $FPF = \frac{\#FP}{\#TN + \#FP}$ |
| Actual Illness | #FN $FNF = \frac{\#FN}{\#FN + \#TP}$ | #TP <i>Sensitivity:</i> $TPF = \frac{\#TP}{\#FN + \#TP}$ |

Biomedical Signal Processing

References:

Rangayyan p. 469

Ling and Sheng: Cost-Sensitive Learning and the Class Imbalance Problem



Sensitivity & Specificity

| Model 1 | Predicted Health | Predicted Illness |
|----------------|------------------|-------------------|
| Actual Health | 230 | 10 |
| Actual Illness | 20 | 5 |

| Model 2 | Predicted Health | Predicted Illness |
|----------------|------------------|-------------------|
| Actual Health | 180 | 60 |
| Actual Illness | 5 | 20 |

$$\text{Specificity} = \frac{230}{230+10} = \mathbf{0.96}$$

$$\text{Sensitivity} = \frac{5}{20+5} = \mathbf{0.2}$$

$$\text{Specificity} = \frac{180}{180+60} = \mathbf{0.75}$$

$$\text{Sensitivity} = \frac{20}{20+5} = \mathbf{0.8}$$

References:

Rangayyan p. 469

Ling and Sheng: Cost-Sensitive Learning and the Class Imbalance Problem



Accuracy

| | Predicted Health | Predicted Illness |
|----------------|------------------|-------------------|
| Actual Health | #TN | #FP |
| Actual Illness | #FN | #TP |



$$\text{Accuracy} = \frac{\#TP + \#TN}{\#TP + \#FP + \#TN + \#FN}$$

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



Accuracy

| Model 1 | Predicted Health | Predicted Illness | Model 2 | Predicted Health | Predicted Illness |
|----------------|------------------|-------------------|----------------|------------------|-------------------|
| Actual Health | 230 | 10 | Actual Health | 180 | 60 |
| Actual Illness | 20 | 5 | Actual Illness | 5 | 20 |

$$\text{Accuracy} = \frac{230+5}{265} = 0.89$$

$$\text{Accuracy} = \frac{180+20}{265} = 0.75$$



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



Cost matrix

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

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



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



Model Cost

| COST | Predicted Health | Predicted Illness |
|----------------|------------------|-------------------|
| Actual Health | 0 | 10 |
| Actual Illness | 100 | -1 |

| Model 1 | Predicted Health | Predicted Illness |
|----------------|------------------|-------------------|
| Actual Health | 230 | 10 |
| Actual Illness | 20 | 5 |

$$\text{Cost} = \frac{230}{230+10} \cdot 0 + \frac{10}{230+10} \cdot 10 + \frac{20}{20+5} \cdot 100 + \frac{5}{20+5} \cdot (-1) = 80.22$$

| Model 2 | Predicted Health | Predicted Illness |
|----------------|------------------|-------------------|
| Actual Health | 180 | 60 |
| Actual Illness | 5 | 20 |

$$\text{Cost} = \frac{180}{180+60} \cdot 0 + \frac{60}{180+60} \cdot 10 + \frac{5}{5+20} \cdot 100 + \frac{20}{5+20} \cdot (-1) = 21.7$$

References:

Rangayyan p. 469

Ling and Sheng: Cost-Sensitive Learning and the Class Imbalance Problem



Model Cost vs Accuracy

| Model 1 | Predicted Health | Predicted Illness | Model 2 | Predicted Health | Predicted Illness |
|----------------|------------------|-------------------|----------------|------------------|-------------------|
| Actual Health | 230 | 10 | Actual Health | 180 | 60 |
| Actual Illness | 60 | 18 | Actual Illness | 52 | 26 |

Accuracy = **0.89**

Cost = **80.22**

Accuracy = **0.75**

Cost = **21.7**

Biomedical Signal Processing

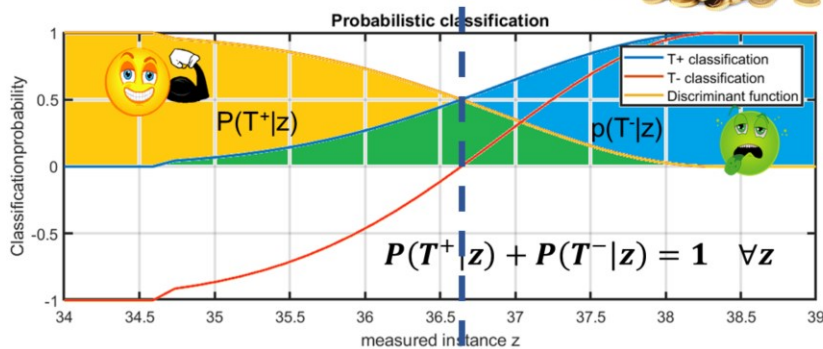
„If we compare both the models and if we check their accuracy. Accuracy for Model 1 is higher compared to Model 2, however the cost for Model 1 is higher compared to Model 2.

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

What if this threshold is very costly?



Biomedical Signal Processing

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 |



Cost minimisation

The classifier will classify an instance z into positive class if and only if it has a lower cost:

$$R^+ \leq R^-$$

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

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

| converted COST* | Predicted Health | Predicted Illness |
|--------------------|---------------------|----------------------|
| Actual Health | 0 | $C(FP)-C(TN)$ |
| Actual Illness | $C(FN)-C(TP)$ | 0 |

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

Biomedical Signal Processing

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:

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



Cost minimisation

| converted COST* | Predicted Health | Predicted Illness |
|--------------------|---------------------------|---------------------------|
| Actual Health | 0 | $C^*(FP) = C(FP) - C(TN)$ |
| Actual Illness | $C^*(FN) = C(FN) - C(TP)$ | 0 |

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

we can obtain a threshold p^* for the classifier to classify an instance z as positive if $P(T^+|z) \geq p^*$, where

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

Biomedical Signal Processing

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 $(1-p^*)$:

$$\begin{aligned} (1-p^*)[C^*(FP)] &= p^*[C^*(FN)] \\ C^*(FP) &= p^*(C^*(FN) + C^*(FP)) \\ p^* &= \frac{C^*(FP)}{(C^*(FN) + C^*(FP))} \end{aligned}$$

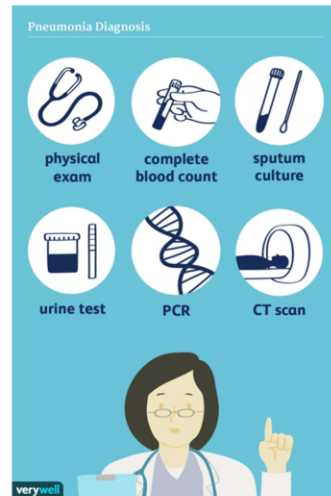
$$p^* = \frac{C^*(FN)}{C^*(FN) + C^*(FP)}$$

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



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





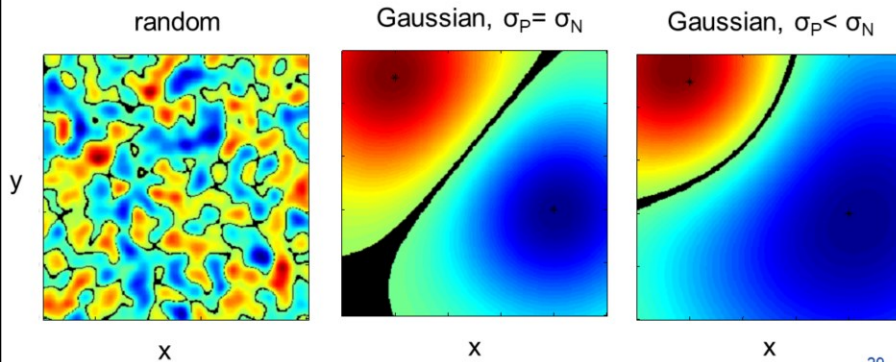
Inevitability of having to exercise ethical/moral/practical judgment

- **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**

28

Combining several variables

$$f(x,y) = p(I|x,y) - p(H|x,y)$$



29

Biomedical Signal Processing

The first one cannot be separated, the second one can be separated linearly. With the third one a non-linear classification might be applied.

Or again, weighted subtraction 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

30



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.*

31

Rangayyan (2002): p. 446-448



Rule-based classification

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

32



| 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 | | | <20.0 |
| 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 ³) | ≥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 |

33

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

Other scores: APGAR, Manning

References:

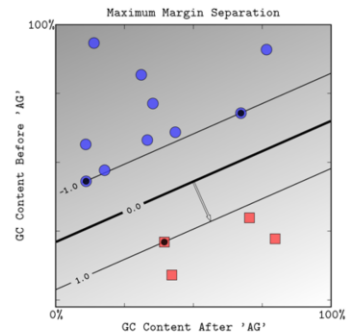
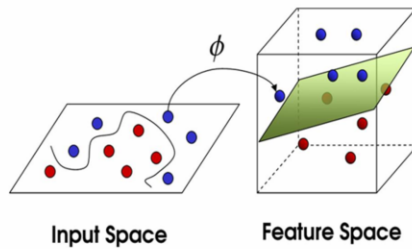
Le Gall et al. (1984): A simplified acute physiology score for ICU patients

http://en.wikipedia.org/wiki/ICU_scoring_systems

https://en.wikipedia.org/wiki/Apgar_score

https://en.wikipedia.org/wiki/Biophysical_profile

Support vector machines



34

Biomedical Signal Processing

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>

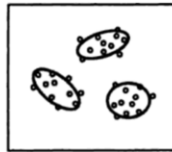


Artificial neural networks

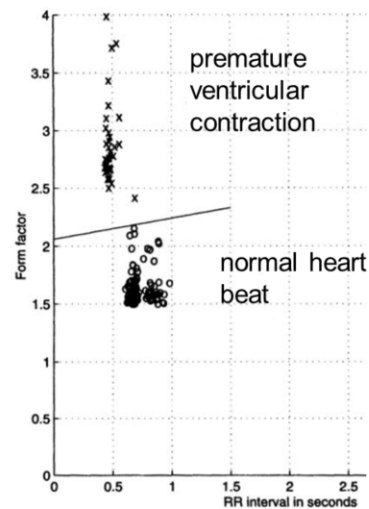
- Perceptron
- Radial basis functions (RBFs)
- Universal approximation theorem



multi-layer perceptron
update weights



RBF network
kernel parameters
and weights



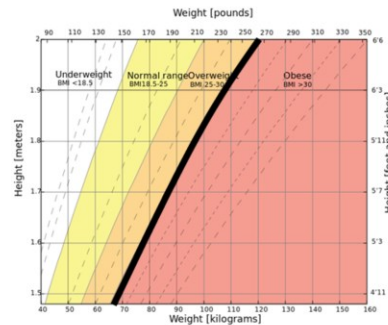
Perceptron: possibility of non-linear mapping following linear combination of inputs

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: $\varphi(\mathbf{x}) = \sum_i a_i \rho(\|\mathbf{x} - \mathbf{c}_i\|)$, with e.g. $\rho(\|\mathbf{x} - \mathbf{c}_i\|) = \exp[-\beta \|\mathbf{x} - \mathbf{c}_i\|^2]$

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)

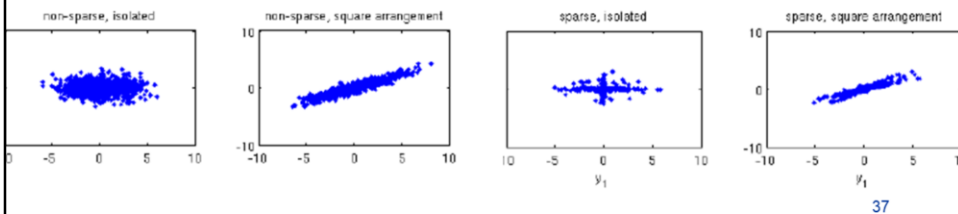


$$\text{BMI} = \text{mass} / \text{height}^2$$



Feature extraction – Déjà Vu 😊

- **Principal Component Analysis**
 - Covariance (2nd order statistic) of components zero
- **Independent Component Analysis**
 - Other measure of independence, such as higher order statistics or mutual information



Biomedical Signal Processing

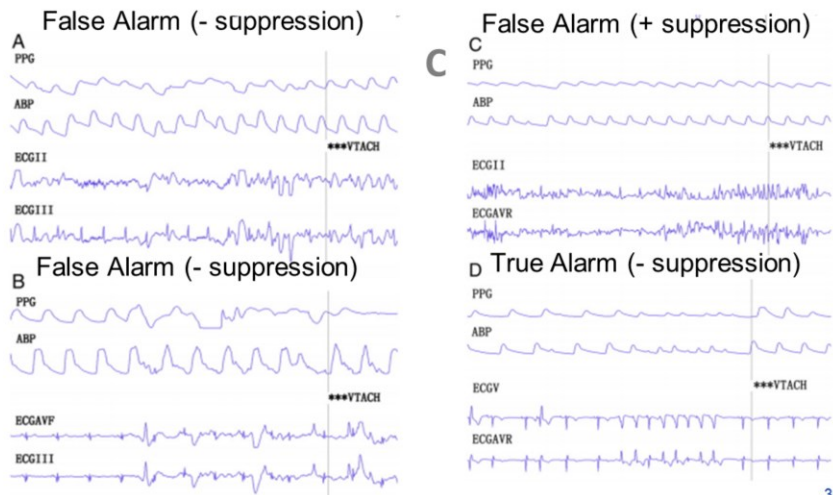
37

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



EXTRA MATERIAL

38



39

Biomedical Signal Processing

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*

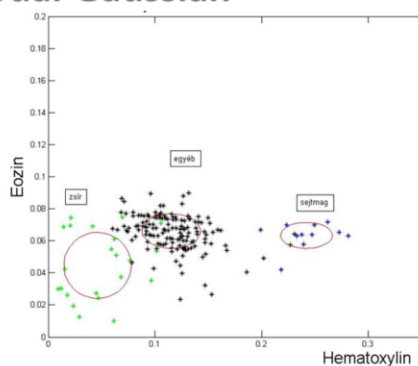
Source:

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 $p(x \mid \text{class})$
- Bayesian est. of $p(\text{class} \mid x)$



40

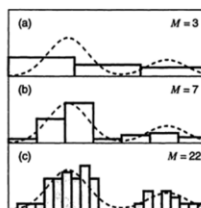


Non-parametric: Estimation of PDF – Other

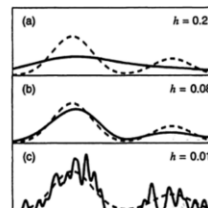
- histogram-based
- kernel-based
- K-nearest neighbour

Mixture model

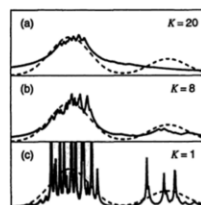
- mixture of Gaussians



histogram-based pdf est.



kernel-based pdf est.



KNN-based pdf est.

41