Introduction to Classification

Antanas Verikas
antanas.verikas@hh.se

IDE, Halmstad University

2013

Task

Assign an object $\mathbf{x}$ into one of several classes $c_1, \ldots, c_Q$.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in X^n \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_Q \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \in C^Q$$

$$X^n \quad \to \quad C^Q$$

`Input space` `Output (class) space`

Classification of $\mathbf{x} \in X^n$ by ANN into Q classes

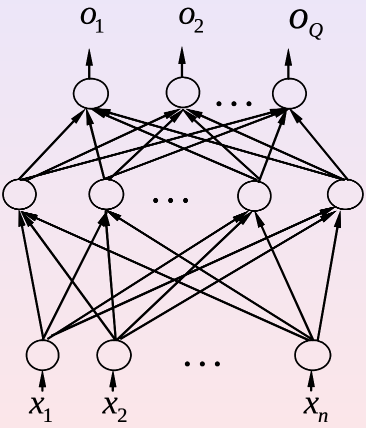


Figure: The multi-layer perceptron

Robot and Lego pieces

Classify the Lego pieces into red, blue, and yellow.

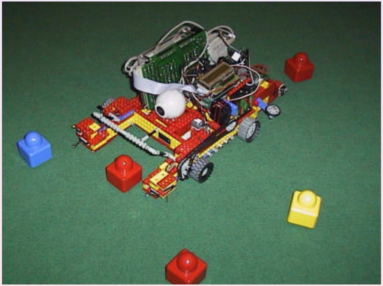


Figure: Robot and Lego pieces.

Camera records images in a 3D space

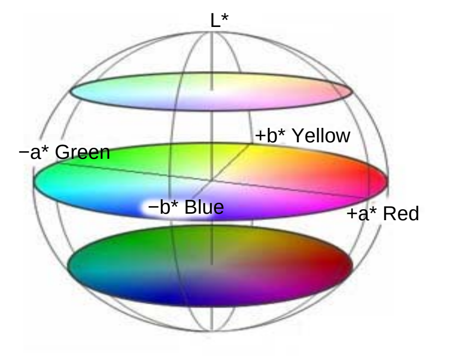


Figure: *Lab* colour space.

Mapping RGB (3D) to rgb (2D)

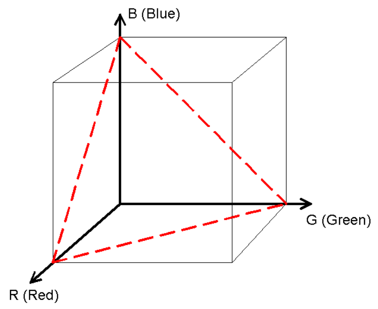
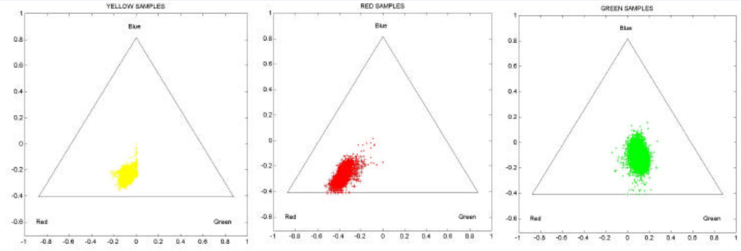


Figure: RGB cube.

$$
\begin{aligned}
r &= \frac{R}{R+G+B} \\
g &= \frac{G}{R+G+B} \\
b &= \frac{B}{R+G+B}
\end{aligned}
$$

Pixels in the normalized rgb colour space



Input is 2D ($\mathbf{x} \in X^2$) and output is 4D ($\mathbf{c} \in C^4$)—red, blue, yellow, green.

All together

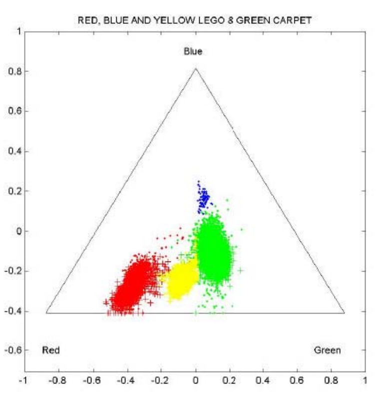
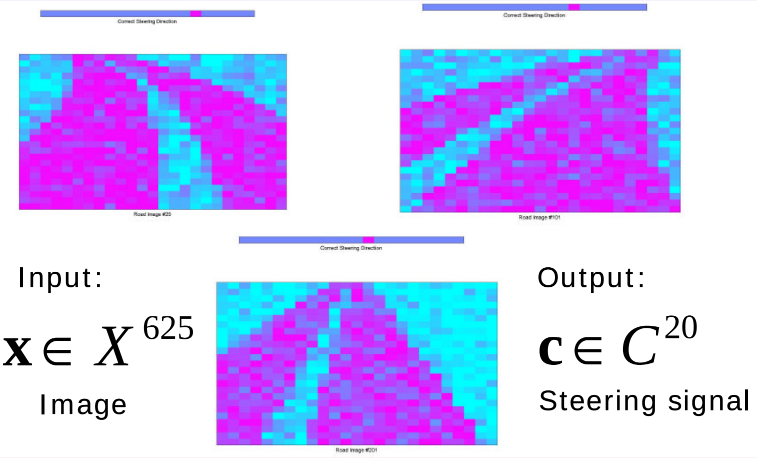


Figure: All pixels.

- The classifier task is to find optimal borders between the different categories.
- Given rgb values, how likely is it that the robot is seeing e.g. a red lego piece?

ANN guided vehicle (1)

ANN guided vehicle (2)



Data types

- Image sequences;
- Voice records;
- Query data

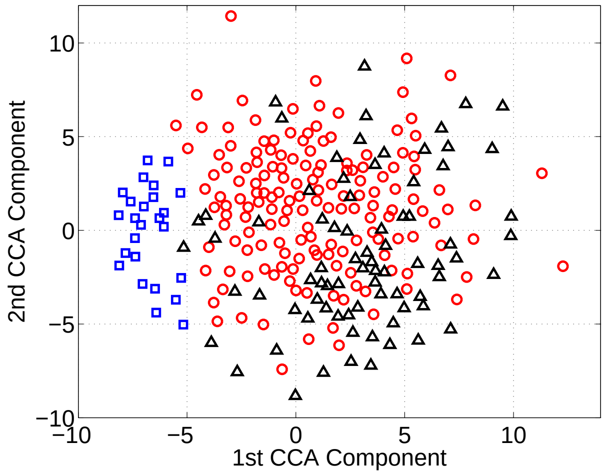
Features (1)

1. Age;
2. Subjectively estimated illness duration (months);
3. Education (five grades);
4. Average duration of intensive speech use (hours/day);
5. Number of days of intensive speech use (days/week);
6. Smoking (Yes/No);
7. Smoked cigarets/day;

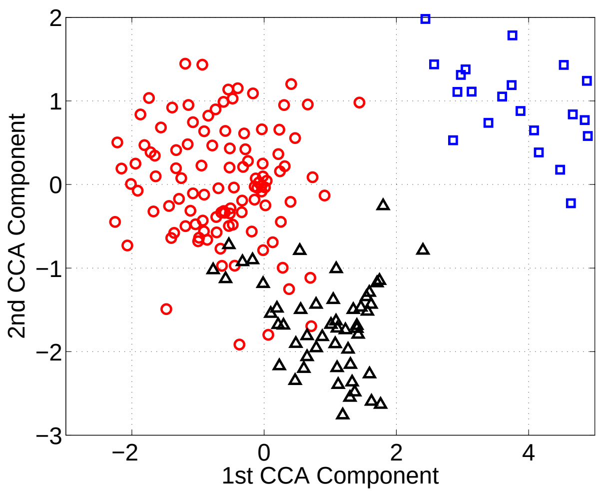
Features (2)

8. Smoking duration (years);
9. Subjective voice function assessment by the patient (0–100);
10. Maximal tonality duration for "aaaaa" (sec);
11. Functional voice index (F);
12. Emotional condition index (E);
13. Physical condition index (P);
14. Voice deficiency index (the maximum value is 120), assessed from answers to questions from a specially designed questionnaire.

Original 14D mapped into 2D



Decisions mapped into 2D



Definitions (1)

Classification means taking a decision

If I believe $\mathbf{x} \in c_k$ then I take an action α_i

Examples

- If I see a yellow Lego brick, then I will lift it up and carry it to my "home".
- If I see a green carpet, then I will keep looking.
- If the road turns left, then I will turn the steering wheel left.

Notation

$p(\mathbf{x})$	Probability density for $\mathbf{x}$.
$p(c_k)$	A priori probability for category c_k.
$p(\mathbf{x}\|c_k)$	Probability density for all $\mathbf{x} \in c_k$.
$p(c_k\|\mathbf{x})$	A posteriori probability for category c_k.
$p(\mathbf{x}, c_k)$	Joint probability for $\mathbf{x}$ and c_k.
α_i	Action i.
$\lambda(\alpha_i\|c_k)$, λ_{ik}	Cost for decision α_i if $\mathbf{x} \in c_k$.

Illustration from health care (1)

- Two categories: $c_1 = \texttt{Healthy}$, $c_2 = \texttt{Ill}$
- $p(c_i) =$ The probability that the person is healthy/ill before the doctor meets him/her. (How many of the people going to see a doctor are actually ill?)
- $\mathbf{x} = \{x_1, x_2, ..., x_n\} =$ The results from the doctor's examination (the doctor may have done many tests).

Illustration from health care (2)

- $p(\mathbf{x})$ = The probability for getting the result $\mathbf{x}$.
- $p(\mathbf{x}, c_i)$ = The probability for observing a person from category c_i with the test results $\mathbf{x}$.

$$p(\mathbf{x}, c_i) = p(\mathbf{x}|c_i)p(c_i) = p(c_i|\mathbf{x})p(\mathbf{x}) \tag{1}$$

- $p(\mathbf{x}|c_i)$ = The probability for getting test results $\mathbf{x}$ when we know the person is from category c_i.

Bayes' rule

$$p(c_k, \mathbf{x}) = p(\mathbf{x}, c_k) \quad \Rightarrow \tag{2}$$

$$p(c_k|\mathbf{x}) = \frac{p(c_k)p(\mathbf{x}|c_k)}{p(\mathbf{x})} \tag{3}$$

$$p(\mathbf{x}) = \sum_{j=1}^{Q} p(c_j)p(\mathbf{x}|c_j) \tag{4}$$

Expected conditional risk

$$R(\alpha_i|\mathbf{x}) = \sum_{k=1}^{Q} \lambda(\alpha_i|c_k)p(c_k|\mathbf{x}) \tag{5}$$

The "Bayes optimal" decision: Choose α_i that minimizes $R(\alpha_i|\mathbf{x})$

Approaches

- Model a posteriori probabilities

$$p(c_k|\mathbf{x}) \tag{6}$$

- Model probability densities & use "Bayes"

$$p(\mathbf{x}|c_k) \tag{7}$$

- Model discrimination functions & discrimination boundaries

Parametric versus Non-parametric

- "Parametric": Assume a parametric form. Few degrees of freedom—usually leads to large model bias.
- "Non-parametric": Assumes no parametric form. Many degrees of freedom—usually leads to large model variance.
- Optimal somewhere in-between.

Linear and 1NN

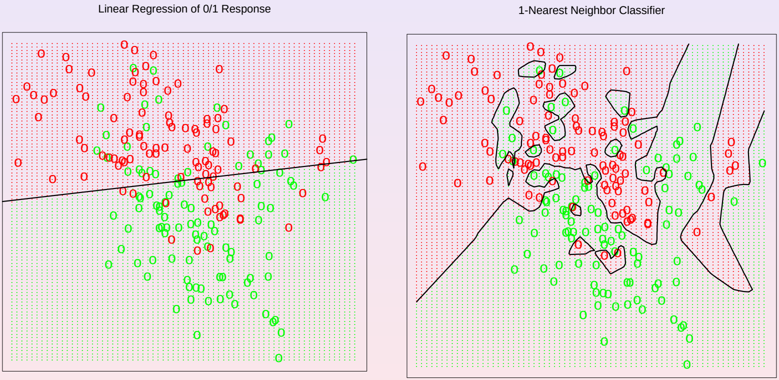


Figure: Decision boundaries of the linear and 1NN classifiers.

15NN and Bayes

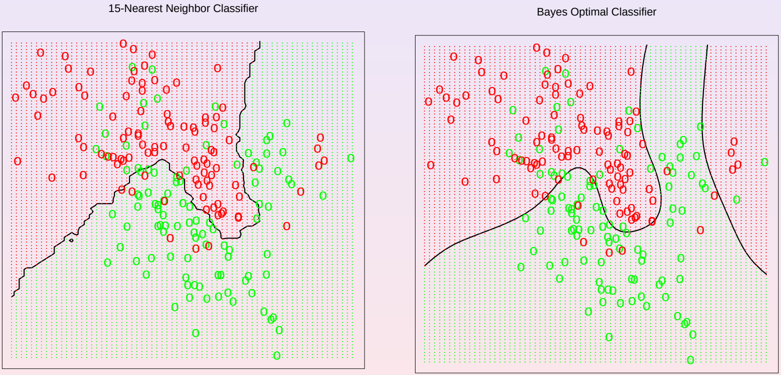


Figure: Decision boundaries of the 15NN and Bayes classifiers.