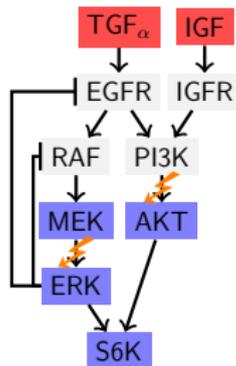
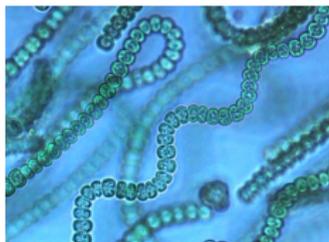


# Logische Analyse von Daten

Softwarepraktikum SoSe 2019

Alexander Bockmayr, Katinka Becker

# Technologischer Fortschritt in Biologie, Medizin und Lebenswissenschaften



# Technologischer Fortschritt in Biologie, Medizin und Lebenswissenschaften



	bak 1	bak 2	bak 3	bak 4
Eigenschaft	nein	nein	ja	ja
Gen 1	0	0	0	1
Gen 2	1	0	1	1
Gen 3	1	1	0	0
Gen 4	0	1	1	1

	MEK	AKT	ERK	S6K
Zustand 1	0	0	1	0
Zustand 2	1	0	0	0
Zustand 3	0	0	0	0
Zustand 4	0	1	0	1
Zustand 5	1	0	1	1
Zustand 6	1	1	1	1

	miRNA 1	miRNA 2	miRNA 3
Gesund 1	0	0	0
Gesund 2	1	1	0
Gesund 3	0	0	0
Mutiert 1	0	1	1
Mutiert 2	1	0	1
Mutiert 3	0	1	1

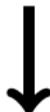
# Technologischer Fortschritt in Biologie, Medizin und Lebenswissenschaften



	bak 1	bak 2	bak 3	bak 4
Eigenschaft	nein	nein	ja	ja
Gen 1	0	0	0	1
Gen 2	1	0	1	1
Gen 3	1	1	0	0
Gen 4	0	1	1	1

	MEK	AKT	ERK	S6K
Zustand 1	0	0	1	0
Zustand 2	1	0	0	0
Zustand 3	0	0	0	0
Zustand 4	0	1	0	1
Zustand 5	1	0	1	1
Zustand 6	1	1	1	1

	miRNA 1	miRNA 2	miRNA 3
Gesund 1	0	0	0
Gesund 2	1	1	0
Gesund 3	0	0	0
Mutiert 1	0	1	1
Mutiert 2	1	0	1
Mutiert 3	0	1	1



**Ziel:** Wichtige Informationen erkennen und heraus filtern!

## Binäre Klassifizierungsprobleme

Gen 1	Gen 2	Gen 3	Gen 4	Eigenschaft
0	1	0	1	ja
1	1	0	1	ja
0	1	1	0	nein
0	0	1	1	nein

}  $\Omega^+$

}  $\Omega^-$

MEK	AKT	ERK	S6K
0	0	1	0
1	0	0	0
0	0	0	0
0	1	0	1
1	0	1	1
1	1	1	1

}  $\Omega^+$

}  $\Omega^-$

miRNA 1	miRNA 2	miRNA 3	Gesund
0	0	0	ja
1	1	0	ja
0	0	0	ja
0	1	1	nein
1	0	1	nein
0	1	1	nein

}  $\Omega^+$

}  $\Omega^-$

# Logische Analyse von Daten (LAD)

Eine Methode die in den späten 80ern von Peter L. Hammer entwickelt wurde

- Diskretisierung
- **Muster Bestimmung**
- Theorie Bildung

[1] Crama, Y., Hammer, P. L., Ibaraki, T. *Cause-effect relationships and partially defined boolean functions*. Annals of Operations Research, 16: 299-325, 1988.

# Muster Bestimmung

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
1	0	1	1	1
0	0	0	1	1
1	1	1	0	1
1	1	1	0	0
1	0	0	1	0
0	0	1	0	1
1	1	1	1	1
1	0	1	0	0
1	0	0	0	1
0	0	1	1	0

}  $\Omega^+$

}  $\Omega^-$

**Definition:** Ein **Positives (Negatives) Muster**

- deckt **mindestens eine positive (negative)** Beobachtung ab
- deckt **keine negative (positive)** Beobachtung ab

# Muster Bestimmung

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
1	0	1	1	1
0	0	0	1	1
1	1	1	0	1
1	1	1	0	0
1	0	0	1	0
0	0	1	0	1
1	1	1	1	1
1	0	1	0	0
1	0	0	0	1
0	0	1	1	0

}  $\Omega^+$

}  $\Omega^-$

**Definition:** Ein **Positives (Negatives) Muster**

- deckt **mindestens eine positive (negative)** Beobachtung ab
- deckt **keine negative (positive)** Beobachtung ab

# Muster Bestimmung

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
1	0	1	1	1
0	0	0	1	1
1	1	1	0	1
1	1	1	0	0
1	0	0	1	0
0	0	1	0	1
1	1	1	1	1
1	0	1	0	0
1	0	0	0	1
0	0	1	1	0

}  $\Omega^+$

}  $\Omega^-$

## Definition: Prim-Muster

- kein kürzerer enthaltener Term ist selbst Muster

## Input: Binäre Datensätze

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
1	0	1	1	1
0	0	0	1	1
1	1	1	0	0
1	1	1	0	0
1	0	0	1	0
0	0	1	0	1
1	1	1	1	1
1	0	1	0	0
1	0	0	0	1
0	0	1	0	1



## Wir nutzen einen Optimierungsansatz.

[1] Hong Seo Ryou and In-Yong Jang. *Milp approach to pattern generation in logical analysis of data*. *Discrete Applied Mathematics*, 157(4):749–761, 2009.



## Wir nutzen die Software ZIMPL und SCIP.



- Modellierungssprache und Optimierungspaket entwickelt am ZIB



## Ergebnis: Wir finden Muster!

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
1	0	1	1	1
0	0	0	1	1
1	1	1	0	0
1	1	1	0	0
1	0	0	1	0
0	0	1	0	1
1	1	1	1	1
1	0	1	0	0
1	0	0	0	0
0	0	1	0	1

# Daten

## **Einführung**

Dienstag, 5.3.19, 10.00-13.00

Donnerstag, 7.3.19, 10.00-13.00



## **Wöchentliches Treffen**

Dienstags, 12.3.19 - 30.4.19,  
10.00-12.00

## **Präsentationen**

Ende April



1. März

1. April

1. Mai