

Data Mining Tutorial

Kernel, Perceptron, Regression

Erich Schubert, Arthur Zimek

Ludwig-Maximilians-Universität München

2015-07-08 — KDD Übung

Wiederholung:

- ▶ Frequent itemsets sind *typische* Teilmengen
- ▶ Assoziationsregeln *erweitern* ein kleines FI in ein größeres FI
- ▶ Apriori ist der bekannteste Algorithmus hierfür, und ein Verfahren, das *kennen muss* (im data mining)
- ▶ Das Prinzip wird auch außerhalb der Warenkorbanalyse verwendet! (Das ist nur das Paradebeispiel)
- ▶ Mehr auf den Vorlesungsfolien!

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	
B	
C	
D	
E	
F	
G	
H	
I	
J	
K	
L	

Database

→ 1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	
B	1
C	
D	
E	1
F	
G	1
H	1
I	
J	
K	
L	

Database

1	B E G H
→ 2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	1
B	2
C	1
D	
E	2
F	
G	2
H	2
I	
J	
K	
L	

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C1

A	2
B	3
C	2
D	
E	3
F	1
G	2
H	3
I	
J	
K	
L	

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C1

A	2
B	4
C	3
D	1
E	4
F	2
G	3
H	4
I	
J	
K	
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
→ 5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	3
B	5
C	3
D	1
E	5
F	2
G	3
H	5
I	
J	
K	1
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
→ 6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	3
B	6
C	3
D	1
E	6
F	3
G	4
H	6
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	4
B	7
C	3
D	2
E	6
F	3
G	5
H	7
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	5
B	8
C	3
D	3
E	6
F	3
G	6
H	7
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C1

A	5
B	9
C	3
D	4
E	6
F	4
G	7
H	7
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	5
B	9
C	4
D	4
E	7
F	5
G	7
H	7
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C1

A	6
B	9
C	5
D	4
E	8
F	6
G	7
H	8
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C1

A	7
B	10
C	5
D	4
E	9
F	6
G	8
H	8
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A	7
B	10
C	5
D	4
E	9
F	6
G	8
H	8
I	4
J	0
K	2
L	4

Minsupport 30% \Leftrightarrow Support $\geq 4/12$

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB		CE	
AC		CF	
AD		CG	
AE		CH	
AF		DE	
AG		DF	
AH		DG	
BC		DH	
BD		EF	
BE		EG	
BF		EH	
BG		FG	
BH		FH	
CD		GH	

Database

→ 1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB		CE	
AC		CF	
AD		CG	
AE		CH	
AF		DE	
AG		DF	
AH		DG	
BC		DH	
BD		EF	
BE	1	EG	1
BF		EH	1
BG	1	FG	
BH	1	FH	
CD		GH	1

Database

1	B E G H
→ 2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB	1	CE	1
AC	1	CF	
AD		CG	1
AE	1	CH	1
AF		DE	
AG	1	DF	
AH	1	DG	
BC	1	DH	
BD		EF	
BE	2	EG	2
BF		EH	2
BG	2	FG	
BH	2	FH	
CD		GH	2

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	2	CE	2
AC	2	CF	1
AD		CG	1
AE	2	CH	2
AF	1	DE	
AG	1	DF	
AH	2	DG	
BC	2	DH	
BD		EF	1
BE	3	EG	2
BF	1	EH	3
BG	2	FG	
BH	3	FH	1
CD		GH	2

Database

1	B E G H
2	A B C E G H
3	A B C E F H
→	4 B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB	2	CE	3
AC	2	CF	2
AD		CG	2
AE	2	CH	3
AF	1	DE	1
AG	1	DF	1
AH	2	DG	1
BC	3	DH	1
BD	1	EF	2
BE	4	EG	3
BF	2	EH	4
BG	3	FG	1
BH	4	FH	2
CD	1	GH	3

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB	3	CE	3
AC	2	CF	2
AD		CG	2
AE	3	CH	3
AF	1	DE	1
AG	1	DF	1
AH	3	DG	1
BC	3	DH	1
BD	1	EF	2
BE	5	EG	3
BF	2	EH	5
BG	3	FG	1
BH	5	FH	2
CD	1	GH	3

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	3	CE	3
AC	2	CF	2
AD		CG	2
AE	3	CH	3
AF	1	DE	1
AG	1	DF	1
AH	3	DG	1
BC	3	DH	1
BD	1	EF	3
BE	6	EG	4
BF	3	EH	6
BG	4	FG	2
BH	6	FH	3
CD	1	GH	4

Database

1	B E G H	
2	A B C E G H	
3	A B C E F H	
4	B C D E F G H L	
5	A B E K H	
6	B E F G H I K	
→	7	A B D G H
8	A B D G	
9	B D F G	
10	C E F	
11	A C E F H	
12	A B E G	

L1

A
B
C
D
E
F
G
H

C2

AB	4	CE	3
AC	2	CF	2
AD	1	CG	2
AE	3	CH	3
AF	1	DE	1
AG	2	DF	1
AH	4	DG	2
BC	3	DH	2
BD	2	EF	3
BE	6	EG	4
BF	3	EH	6
BG	5	FG	2
BH	7	FH	3
CD	1	GH	5

Database

1	B E G H	
2	A B C E G H	
3	A B C E F H	
4	B C D E F G H L	
5	A B E K H	
6	B E F G H I K	
7	A B D G H	
→	8	A B D G
9	B D F G	
10	C E F	
11	A C E F H	
12	A B E G	

L1

A
B
C
D
E
F
G
H

C2

AB	5	CE	3
AC	2	CF	2
AD	2	CG	2
AE	3	CH	3
AF	1	DE	1
AG	3	DF	1
AH	4	DG	3
BC	3	DH	2
BD	3	EF	3
BE	6	EG	4
BF	3	EH	6
BG	6	FG	2
BH	7	FH	3
CD	1	GH	5

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	5	CE	3
AC	2	CF	2
AD	2	CG	2
AE	3	CH	3
AF	1	DE	1
AG	3	DF	2
AH	4	DG	4
BC	3	DH	2
BD	4	EF	3
BE	6	EG	4
BF	4	EH	6
BG	7	FG	3
BH	7	FH	3
CD	1	GH	5

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
→ 10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB	5	CE	4
AC	2	CF	3
AD	2	CG	2
AE	3	CH	3
AF	1	DE	1
AG	3	DF	2
AH	4	DG	4
BC	3	DH	2
BD	4	EF	4
BE	6	EG	4
BF	4	EH	6
BG	7	FG	3
BH	7	FH	3
CD	1	GH	5

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	5	CE	5
AC	3	CF	4
AD	2	CG	2
AE	4	CH	4
AF	2	DE	1
AG	3	DF	2
AH	5	DG	4
BC	3	DH	2
BD	4	EF	5
BE	6	EG	4
BF	4	EH	7
BG	7	FG	3
BH	7	FH	4
CD	1	GH	5

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	6	CE	5
AC	3	CF	4
AD	2	CG	2
AE	5	CH	4
AF	2	DE	1
AG	4	DF	2
AH	5	DG	4
BC	3	DH	2
BD	4	EF	5
BE	7	EG	5
BF	4	EH	7
BG	8	FG	3
BH	7	FH	4
CD	1	GH	5

Database

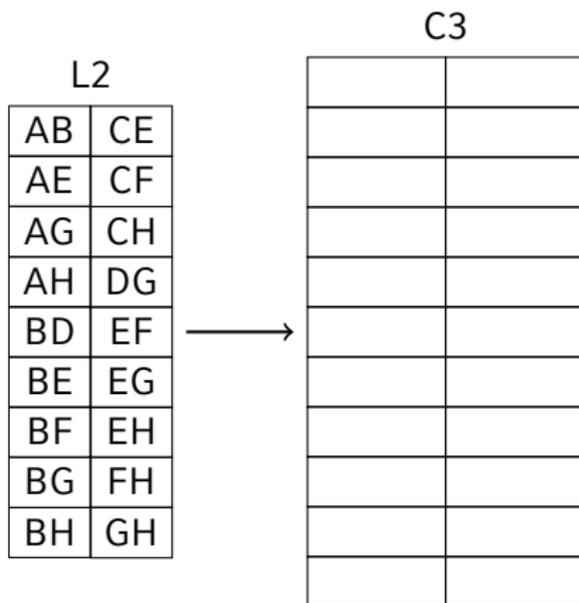
1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

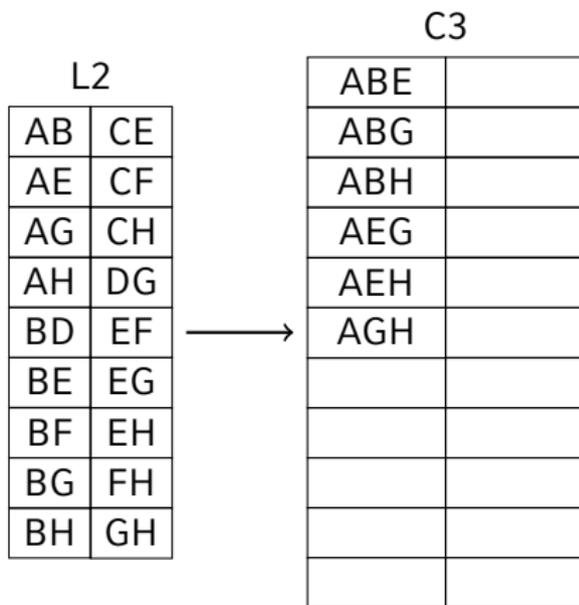
A
B
C
D
E
F
G
H

L2

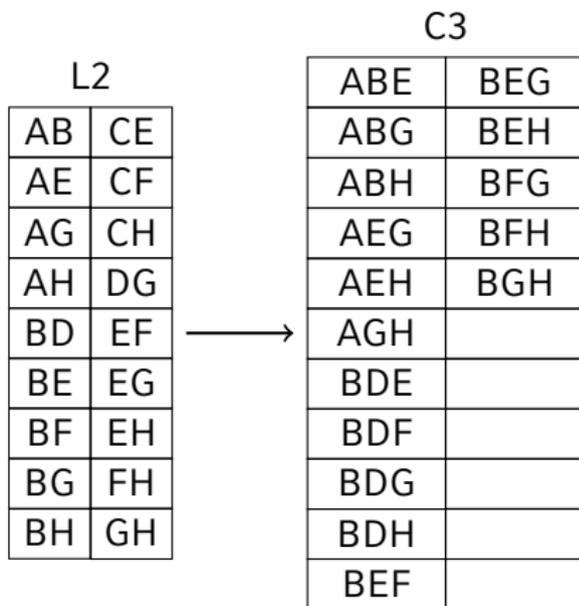
AB	6	CE	5
AC	3	CF	4
AD	2	CG	2
AE	5	CH	4
AF	2	DE	1
AG	4	DF	2
AH	5	DG	4
BC	3	DH	2
BD	4	EF	5
BE	7	EG	5
BF	4	EH	7
BG	8	FG	3
BH	7	FH	4
CD	1	GH	5



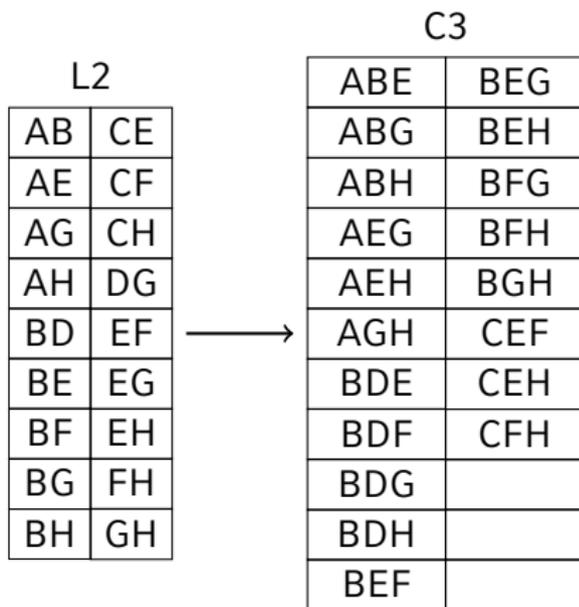
Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



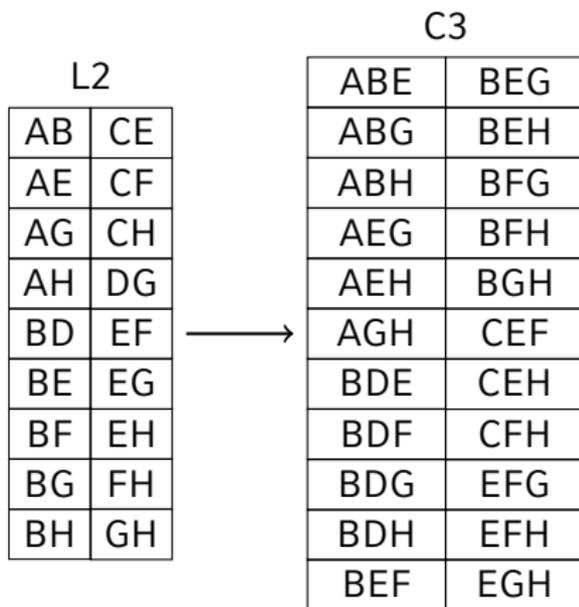
Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



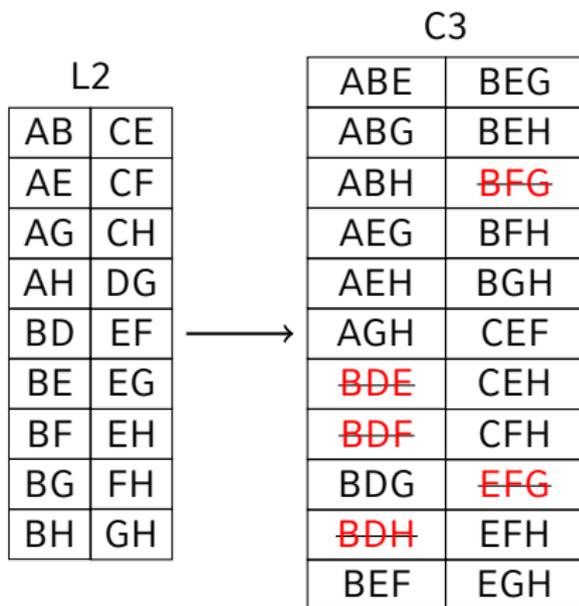
Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE		BEH	
ABG		BFH	
ABH		BGH	
AEG		CEF	
AEH		CEH	
AGH		CFH	
BDG		EFH	
BEF		EGH	
BEG			

Database

→ 1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE		BEH	1
ABG		BFH	
ABH		BGH	1
AEG		CEF	
AEH		CEH	
AGH		CFH	
BDG		EFH	
BEF		EGH	1
BEG	1		

Database

1	B E G H
→ 2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE	1	BEH	2
ABG	1	BFH	
ABH	1	BGH	2
AEG	1	CEF	
AEH	1	CEH	1
AGH	1	CFH	
BDG		EFH	
BEF		EGH	2
BEG	2		

Database

1	B E G H
2	A B C E G H
→ 3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE	2	BEH	3
ABG	1	BFH	1
ABH	2	BGH	2
AEG	1	CEF	1
AEH	2	CEH	2
AGH	1	CFH	1
BDG		EFH	1
BEF	1	EGH	2
BEG	2		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
→ 4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE	2	BEH	4
ABG	1	BFH	2
ABH	2	BGH	3
AEG	1	CEF	2
AEH	2	CEH	3
AGH	1	CFH	2
BDG	1	EFH	2
BEF	2	EGH	3
BEG	3		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
→ 5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE	3	BEH	5
ABG	1	BFH	2
ABH	3	BGH	3
AEG	1	CEF	2
AEH	3	CEH	3
AGH	1	CFH	2
BDG	1	EFH	2
BEF	2	EGH	3
BEG	3		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

ABE	3	BEH	6
ABG	1	BFH	3
ABH	3	BGH	4
AEG	1	CEF	2
AEH	3	CEH	3
AGH	1	CFH	2
BDG	1	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

ABE	3	BEH	6
ABG	2	BFH	3
ABH	4	BGH	5
AEG	1	CEF	2
AEH	3	CEH	3
AGH	2	CFH	2
BDG	2	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

ABE	3	BEH	6
ABG	3	BFH	3
ABH	4	BGH	5
AEG	1	CEF	2
AEH	3	CEH	3
AGH	2	CFH	2
BDG	3	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

ABE	3	BEH	6
ABG	3	BFH	3
ABH	4	BGH	5
AEG	1	CEF	2
AEH	3	CEH	3
AGH	2	CFH	2
BDG	4	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
→ 10	C E F
11	A C E F H
12	A B E G

C3

ABE	3	BEH	6
ABG	3	BFH	3
ABH	4	BGH	5
AEG	1	CEF	3
AEH	3	CEH	3
AGH	2	CFH	2
BDG	4	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

ABE	3	BEH	6
ABG	3	BFH	3
ABH	4	BGH	5
AEG	1	CEF	4
AEH	4	CEH	4
AGH	2	CFH	3
BDG	4	EFH	4
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
→ 12	A B E G

C3

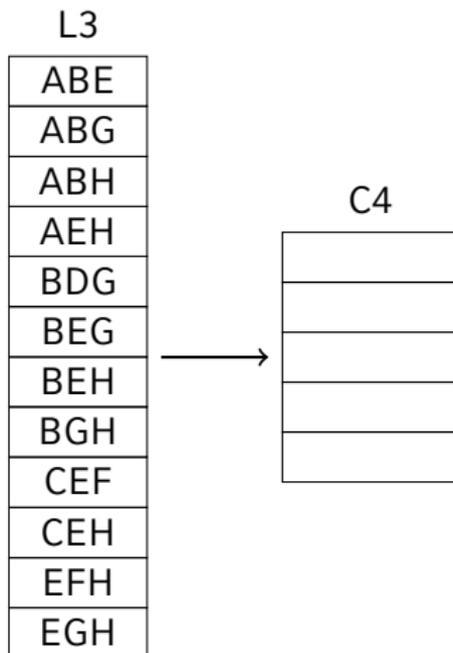
ABE	4	BEH	6
ABG	4	BFH	3
ABH	4	BGH	5
AEG	2	CEF	4
AEH	4	CEH	4
AGH	2	CFH	3
BDG	4	EFH	4
BEF	3	EGH	4
BEG	5		

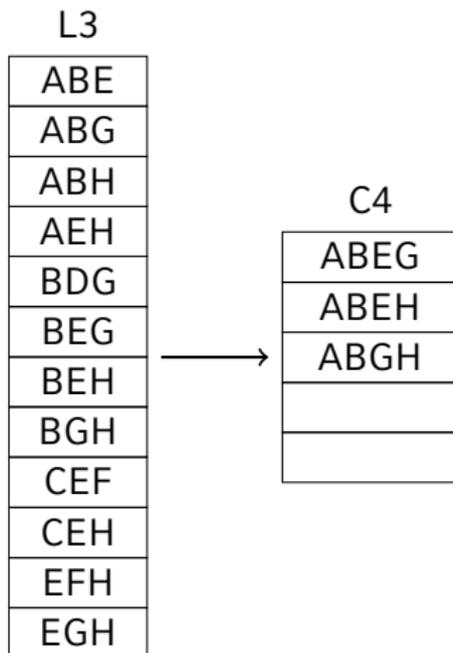
Database

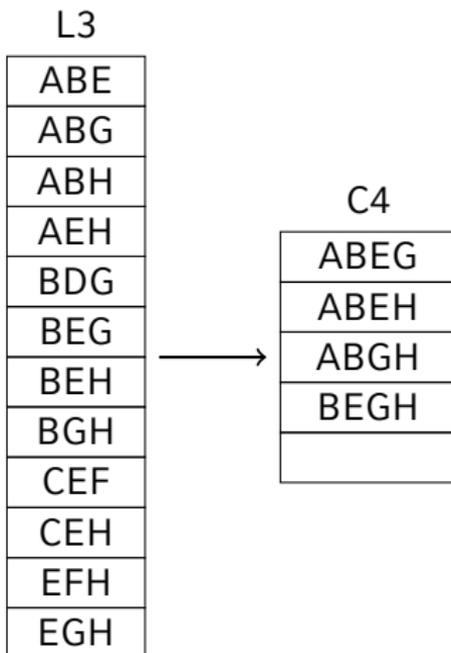
1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

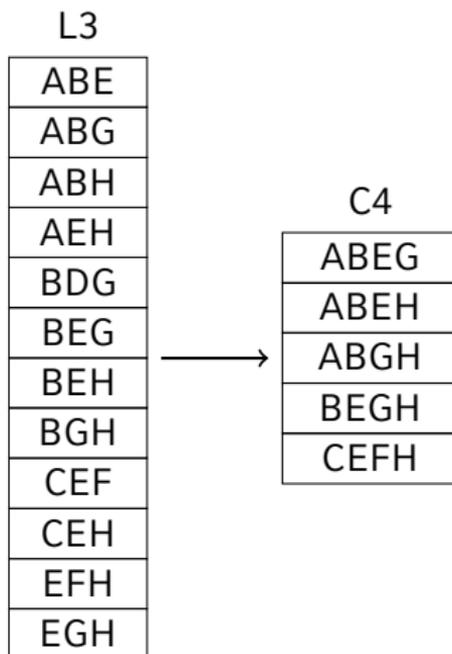
L3

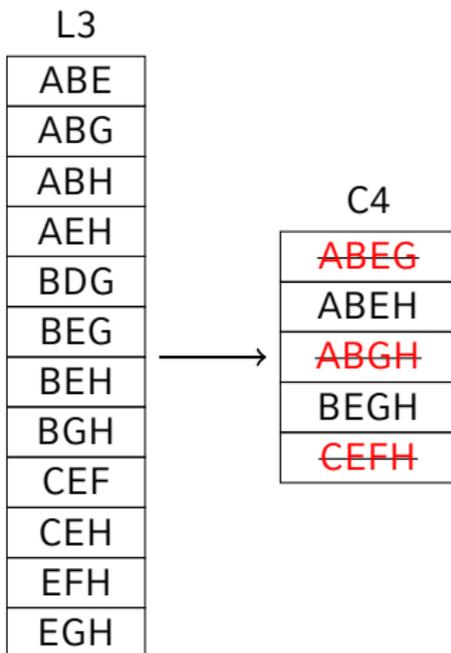
ABE	4	BEH	6
ABG	4	BFH	3
ABH	4	BGH	5
AEG	2	CEF	4
AEH	4	CEH	4
AGH	2	CFH	3
BDG	4	EFH	4
BEF	3	EGH	4
BEG	5		











AEG, AGH und CFH sind nicht *frequent*!

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	
BEGH	

Database

→ 1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	
BEGH	1

Database

1	B E G H
→ 2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	1
BEGH	2

Database

1	B E G H
2	A B C E G H
→ 3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	2
BEGH	2

Database

1	B E G H
2	A B C E G H
3	A B C E F H
→ 4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	2
BEGH	3

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
→ 5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	3
BEGH	3

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C4

ABEH	3
BEGH	4

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
→ 7	A B D G H
→ 8	A B D G
→ 9	B D F G
→ 10	C E F
→ 11	A C E F H
→ 12	A B E G

C4

ABEH	3
BEGH	4

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L4

ABEH	3
BEGH	4

Nur ein frequent 4-Itemset übrig!

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-
EH	BG	7	$4/7 \approx 0.571$	-

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-
EH	BG	7	$4/7 \approx 0.571$	-
BGH	E	5	$4/5 = \mathbf{0.800}$	BGH \Rightarrow E

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-
EH	BG	7	$4/7 \approx 0.571$	-
BGH	E	5	$4/5 = \mathbf{0.800}$	BGH \Rightarrow E
GH	BE	5	$4/5 = \mathbf{0.800}$	GH \Rightarrow BE

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-
EH	BG	7	$4/7 \approx 0.571$	-
BGH	E	5	$4/5 = \mathbf{0.800}$	BGH \Rightarrow E
GH	BE	5	$4/5 = \mathbf{0.800}$	GH \Rightarrow BE
EGH	B	4	$4/4 = \mathbf{1.000}$	EGH \Rightarrow B

“Kernel” kann mehrdeutig sein! Unterscheidet zwischen:

- ▶ Kernel function (diese Aufgabe)
- ▶ Kernel density function (in der Statistik)
- ▶ Kernel matrix (oftmals: eine vorberechnete Distanz-Matrix – z.B. mit Kernelfunktion)
- ▶ Positiv (semi-) definite Matrix A in $d(x, x) := x^T A x \geq 0$

“Kernel” kann mehrdeutig sein! Unterscheidet zwischen:

- ▶ Kernel function (diese Aufgabe)
- ▶ Kernel density function (in der Statistik)
- ▶ Kernel matrix (oftmals: eine vorberechnete Distanz-Matrix – z.B. mit Kernelfunktion)
- ▶ Positiv (semi-) definite Matrix A in $d(x, x) := x^T A x \geq 0$

Positiv definite Matrix $A \Rightarrow x^T A y$ ist eine Kernel-Funktion.

“Kernel” kann mehrdeutig sein! Unterscheidet zwischen:

- ▶ Kernel function (diese Aufgabe)
- ▶ Kernel density function (in der Statistik)
- ▶ Kernel matrix (oftmals: eine vorberechnete Distanz-Matrix – z.B. mit Kernelfunktion)
- ▶ Positiv (semi-) definite Matrix A in $d(x, x) := x^T A x \geq 0$

Positiv definite Matrix $A \Rightarrow x^T A y$ ist eine Kernel-Funktion.

Aber nicht jede Kernelfunktion ist als positiv definite *Matrix* repräsentierbar!

Positiv semi-definit \leftrightarrow Generalisierte Skalarprodukte

Standardskalarprodukt: $\langle x, y \rangle = \sum_i x_i y_i$

Generalisiertes Skalarprodukt: $\langle x, y \rangle_A = x^T \cdot A \cdot y$

Positiv semi-definit \leftrightarrow Generalisierte Skalarprodukte

Standardskalarprodukt: $\langle x, y \rangle = \sum_i x_i y_i$

Generalisiertes Skalarprodukt: $\langle x, y \rangle_A = x^T \cdot A \cdot y$

Matrix E so dass $x^T \cdot E \cdot y = \langle x, y \rangle$?

$$\langle x, y \rangle = \sum_i \sum_j e_{ij} \cdot x_i \cdot y_j$$

Positiv semi-definit \leftrightarrow Generalisierte Skalarprodukte

Standardskalarprodukt: $\langle x, y \rangle = \sum_i x_i y_i$

Generalisiertes Skalarprodukt: $\langle x, y \rangle_A = x^T \cdot A \cdot y$

Matrix E so dass $x^T \cdot E \cdot y = \langle x, y \rangle$?

$$\langle x, y \rangle = \sum_i \sum_j e_{ij} \cdot x_i \cdot y_j$$

$$e_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Positiv semi-definit \leftrightarrow Generalisierte Skalarprodukte

Standardskalarprodukt: $\langle x, y \rangle = \sum_i x_i y_i$

Generalisiertes Skalarprodukt: $\langle x, y \rangle_A = x^T \cdot A \cdot y$

Matrix E so dass $x^T \cdot E \cdot y = \langle x, y \rangle$?

$$\langle x, y \rangle = \sum_i \sum_j e_{ij} \cdot x_i \cdot y_j$$

$$e_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Das ist die Einheitsmatrix!

Beweise für ein paar Kernel-Funktionen:

$$0) k_0(x, y) = \langle x, y \rangle = x^T \cdot y$$

Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

Beweise für ein paar Kernel-Funktionen:

$$0) k_0(x, y) = \langle x, y \rangle = x^T \cdot y$$

$$k_0(x, x) = \langle x, x \rangle = \sum_i x_i x_i$$

Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

Beweise für ein paar Kernel-Funktionen:

$$0) k_0(x, y) = \langle x, y \rangle = x^T \cdot y$$

$$k_0(x, x) = \langle x, x \rangle = \sum_i x_i x_i = \sum_i x_i^2$$

Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

Beweise für ein paar Kernel-Funktionen:

$$0) k_0(x, y) = \langle x, y \rangle = x^T \cdot y$$

$$k_0(x, x) = \langle x, x \rangle = \sum_i x_i x_i = \sum_i x_i^2 \geq 0 \text{ offensichtlich}$$

Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

Beweise für ein paar Kernel-Funktionen:

$$0) k_0(x, y) = \langle x, y \rangle = x^T \cdot y$$

$$k_0(x, x) = \langle x, x \rangle = \sum_i x_i x_i = \sum_i x_i^2 \geq 0 \text{ offensichtlich}$$

$$A) k_1(x, y) = 1$$

Beweise für ein paar Kernel-Funktionen:

$$0) k_0(x, y) = \langle x, y \rangle = x^T \cdot y$$

$$k_0(x, x) = \langle x, x \rangle = \sum_i x_i x_i = \sum_i x_i^2 \geq 0 \text{ offensichtlich}$$

$$A) k_1(x, y) = 1 = c^+ \text{ für nicht-negative Konstante } c^+ \in \mathbb{R}^{0+}$$

Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

Beweise für ein paar Kernel-Funktionen:

$$\begin{aligned} 0) \quad k_0(x, y) &= \langle x, y \rangle = x^T \cdot y \\ k_0(x, x) &= \langle x, x \rangle = \sum_i x_i x_i = \sum_i x_i^2 \geq 0 \text{ offensichtlich} \end{aligned}$$

$$\begin{aligned} A) \quad k_1(x, y) &= 1 = c^+ \text{ für nicht-negative Konstante } c^+ \in \mathbb{R}^{0+} \\ k_1(x, x) &= c^+ \geq 0 \text{ trivial.} \end{aligned}$$

Beweise für ein paar Kernel-Funktionen:

$$\text{B) } k_2(x, y) = 3 \cdot x^T \cdot y$$

Beweise für ein paar Kernel-Funktionen:

$$\text{B) } k_2(x, y) = 3 \cdot x^T \cdot y = c^+ \cdot k_0(x, y)$$

Beweise für ein paar Kernel-Funktionen:

$$\text{B) } k_2(x, y) = 3 \cdot x^T \cdot y = c^+ \cdot k_0(x, y)$$

$$k_2(x, x) = \underbrace{c^+}_{3} \cdot \underbrace{k_0(x, x)}_{x^T \cdot x}$$

Beweise für ein paar Kernel-Funktionen:

$$\text{B) } k_2(x, y) = 3 \cdot x^T \cdot y = c^+ \cdot k_0(x, y)$$

$$k_2(x, x) = \underbrace{c^+}_{\geq 0} \cdot \underbrace{k_0(x, x)}_{\geq 0} \geq 0$$

Beweise für ein paar Kernel-Funktionen:

$$\text{B) } k_2(x, y) = 3 \cdot x^T \cdot y = c^+ \cdot k_0(x, y)$$

$$k_2(x, x) = \underbrace{c^+}_{\geq 0} \cdot \underbrace{k(x, x)}_{\geq 0} \geq 0$$

Beweise für ein paar Kernel-Funktionen:

$$\text{B) } k_2(x, y) = 3 \cdot x^T \cdot y = c^+ \cdot k_0(x, y)$$

$$k_2(x, x) = \underbrace{c^+}_{\geq 0} \cdot \underbrace{k(x, x)}_{\geq 0} \geq 0$$

$$\text{C) } k_3(x, y) = 3 \cdot x^T \cdot y + 5$$

Beweise für ein paar Kernel-Funktionen:

$$\text{B) } k_2(x, y) = 3 \cdot x^T \cdot y = c^+ \cdot k_0(x, y)$$

$$k_2(x, x) = \underbrace{c^+}_{\geq 0} \cdot \underbrace{k(x, x)}_{\geq 0} \geq 0$$

$$\text{C) } k_3(x, y) = 3 \cdot x^T \cdot y + 5 = c^+ \cdot k_0(x, y) + d^+$$

Beweise für ein paar Kernel-Funktionen:

$$B) k_2(x, y) = 3 \cdot x^T \cdot y = c^+ \cdot k_0(x, y)$$

$$k_2(x, x) = \underbrace{c^+}_{\geq 0} \cdot \underbrace{k(x, x)}_{\geq 0} \geq 0$$

$$C) k_3(x, y) = 3 \cdot x^T \cdot y + 5 = c^+ \cdot k_0(x, y) + d^+$$

Genauso! Im allgemeinen: ein beliebiges Polynom aus nicht-negativen Faktoren und positiv semi-definiten Kernel-Funktionen ist wiederum positiv semi-definit!

$$\text{Beispiel: } 2k_0(x, y) \cdot k_1(x, y) + k_0(x, y)^2 + k_1(x, y)^2 + 7$$

Aufgabe 12-1

- C1 und L1
- C2 und L2
- C3 und L3
- C4 und L4
- Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

$$A \wedge B \wedge C$$

$(A, B, C) \mapsto 0:$ $(A, B, C) \mapsto 1:$

0 0 0

1 1 1

0 0 1

0 1 0

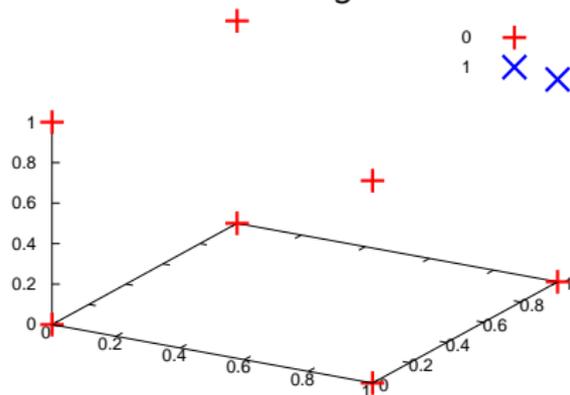
0 1 1

1 0 0

1 0 1

1 1 0

Veranschaulichung:



Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

 $A \vee B$ $(A, B) \mapsto 0:$

0 0

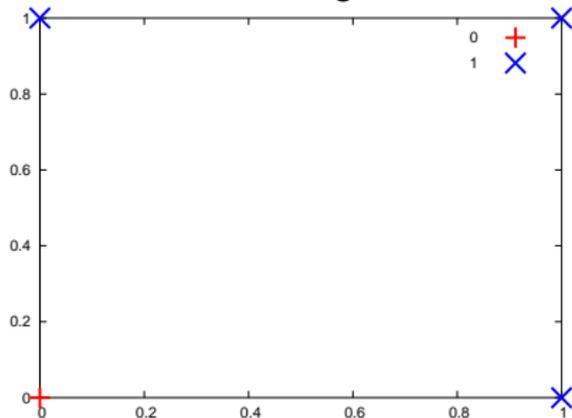
 $(A, B) \mapsto 1:$

0 1

1 0

1 1

Veranschaulichung:



$$(A \vee B) \wedge (A \vee C)$$

Aufgabe 12-1

- C1 und L1
- C2 und L2
- C3 und L3
- C4 und L4
- Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

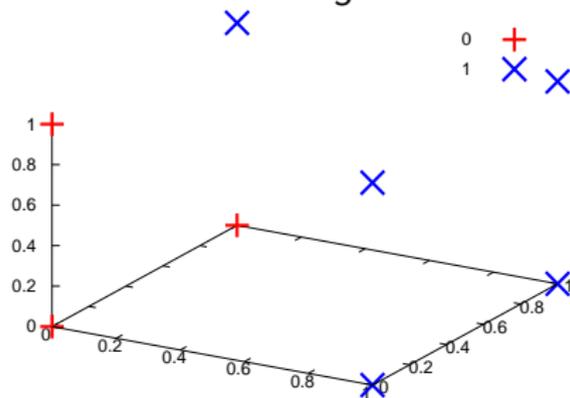
$(A, B, C) \mapsto 0:$

0 0 0
0 0 1
0 1 0

$(A, B, C) \mapsto 1:$

0 1 1
1 0 0
1 0 1
1 1 0
1 1 1

Veranschaulichung:



Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

 $(A, B) \mapsto 0:$

0 0

1 0

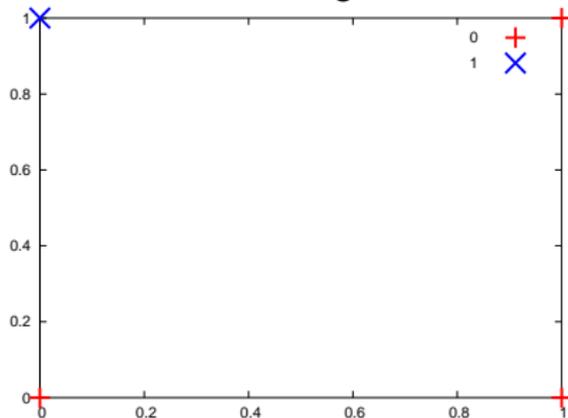
1 1

 $(A, B) \mapsto 1:$

0 1

 $\neg A \wedge B$

Veranschaulichung:



Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

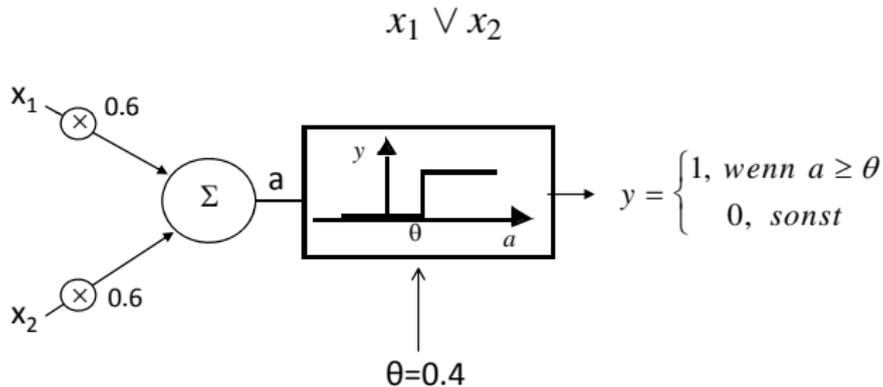
Assoziationsregeln

Aufgabe 12-2

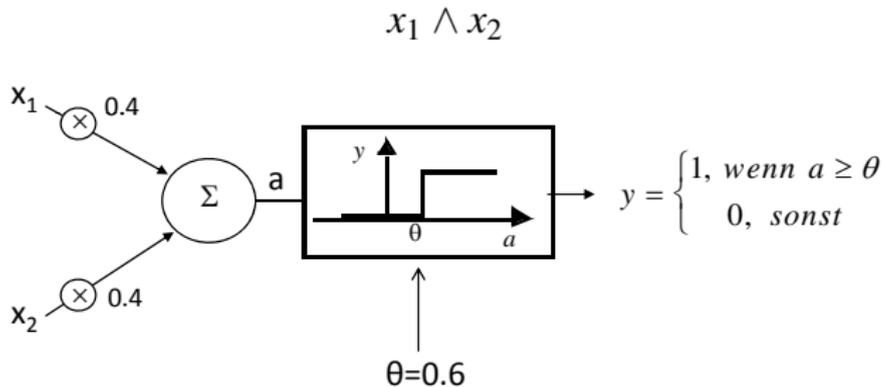
Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5



Bemerkung: θ und Gewichte sind nicht eindeutig!



Bemerkung: θ und Gewichte sind nicht eindeutig!

Eingabedaten:

x	3	8	9	13	3	6	11	21	1	16
y	30	57	64	72	36	43	59	90	20	83

Lineare Regression \Leftrightarrow beste Ausgleichsgerade!

Eingabedaten:

x	3	8	9	13	3	6	11	21	1	16
y	30	57	64	72	36	43	59	90	20	83

Allgemeine Gerade: $y = \alpha \cdot x + \beta$.

Benötigt also: α (Steigung), β (Achsenabschnitt).

Eingabedaten:

x	3	8	9	13	3	6	11	21	1	16
y	30	57	64	72	36	43	59	90	20	83

Allgemeine Gerade: $y = \alpha \cdot x + \beta$.

Benötigt also: α (Steigung), β (Achsenabschnitt).

Optimale Steigung: $\text{Cov}(X, Y) / \text{Var}(X)$

(Kleinste-Quadrate-Schätzer, siehe math. Statistik)

Eingabedaten:

x	3	8	9	13	3	6	11	21	1	16
y	30	57	64	72	36	43	59	90	20	83

Allgemeine Gerade: $y = \alpha \cdot x + \beta$.

Benötigt also: α (Steigung), β (Achsenabschnitt).

Optimale Steigung: $\text{Cov}(X, Y) / \text{Var}(X)$

(Kleinste-Quadrate-Schätzer, siehe math. Statistik)

Optimaler Achsenabschnitt: $\bar{y} = \alpha \bar{x} + \beta$

Eingabedaten:

x	3	8	9	13	3	6	11	21	1	16
y	30	57	64	72	36	43	59	90	20	83

Allgemeine Gerade: $y = \alpha \cdot x + \beta$.

Benötigt also: α (Steigung), β (Achsenabschnitt).

Optimale Steigung: $\text{Cov}(X, Y) / \text{Var}(X)$

(Kleinste-Quadrate-Schätzer, siehe math. Statistik)

Optimaler Achsenabschnitt: $\beta = \bar{y} - \alpha \bar{x}$

Für Mittelwerte \bar{x} , \bar{y} und α .

Eingabedaten:

x	3	8	9	13	3	6	11	21	1	16
y	30	57	64	72	36	43	59	90	20	83

Allgemeine Gerade: $y = \alpha \cdot x + \beta$.

Benötigt also: α (Steigung), β (Achsenabschnitt).

Optimale Steigung: $\text{Cov}(X, Y) / \text{Var}(X)$

(Kleinste-Quadrate-Schätzer, siehe math. Statistik)

Optimaler Achsenabschnitt: $\beta = \bar{y} - \alpha \bar{x}$

Für Mittelwerte \bar{x} , \bar{y} und α .

Also als Erstes: Mittelwerte berechnen, Daten zentrieren.

Eingabedaten:

x	3	8	9	13	3	6	11	21	1	16
y	30	57	64	72	36	43	59	90	20	83

Allgemeine Gerade: $y = \alpha \cdot x + \beta$.

Benötigt also: α (Steigung), β (Achsenabschnitt).

Optimale Steigung: $\text{Cov}(X, Y) / \text{Var}(X)$

(Kleinste-Quadrate-Schätzer, siehe math. Statistik)

Optimaler Achsenabschnitt: $\beta = \bar{y} - \alpha \bar{x}$

Für Mittelwerte \bar{x} , \bar{y} und α .

Also als Erstes: Mittelwerte berechnen, Daten zentrieren.

Mittelwerte: $\bar{x} = 9.1$, $\bar{y} = 55.4$

Zentrierte Daten:

$\bar{x} = 9.1$	$x - \bar{x}$	-6.1	-1.1	-0.1	3.9	-6.1	-3.1	1.9	11.9	-8.1	6.9
$\bar{y} = 55.4$	$y - \bar{y}$	-25.4	1.6	8.6	16.6	-19.4	-12.4	3.6	34.6	-35.4	27.6

Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

Zentrierte Daten:

$\bar{x} = 9.1$	$x - \bar{x}$	-6.1	-1.1	-0.1	3.9	-6.1	-3.1	1.9	11.9	-8.1	6.9
$\bar{y} = 55.4$	$y - \bar{y}$	-25.4	1.6	8.6	16.6	-19.4	-12.4	3.6	34.6	-35.4	27.6
	$(x - \bar{x}) \cdot (y - \bar{y})$	154.94	-1.76	-0.86	64.74	118.34	38.44	6.84	411.74	286.74	190.44

Zentrierte Daten:

$\bar{x} = 9.1$	$x - \bar{x}$	-6.1	-1.1	-0.1	3.9	-6.1	-3.1	1.9	11.9	-8.1	6.9
$\bar{y} = 55.4$	$y - \bar{y}$	-25.4	1.6	8.6	16.6	-19.4	-12.4	3.6	34.6	-35.4	27.6
$(x - \bar{x}) \cdot (y - \bar{y})$		154.94	-1.76	-0.86	64.74	118.34	38.44	6.84	411.74	286.74	190.44
$(x - \bar{x})^2$		37.21	1.21	0.01	15.21	37.21	9.61	3.61	141.61	65.61	47.61

Aufgabe 12-1

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Aufgabe 12-2

Aufgabe 12-3

Aufgabe 12-4

Aufgabe 12-5

Zentrierte Daten:

$\bar{x} = 9.1$	$x - \bar{x}$	-6.1	-1.1	-0.1	3.9	-6.1	-3.1	1.9	11.9	-8.1	6.9
$\bar{y} = 55.4$	$y - \bar{y}$	-25.4	1.6	8.6	16.6	-19.4	-12.4	3.6	34.6	-35.4	27.6
	$(x - \bar{x}) \cdot (y - \bar{y})$	154.94	-1.76	-0.86	64.74	118.34	38.44	6.84	411.74	286.74	190.44
	$(x - \bar{x})^2$	37.21	1.21	0.01	15.21	37.21	9.61	3.61	141.61	65.61	47.61

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\frac{1}{n-1} \sum (x - \bar{x})(y - \bar{y})}{\frac{1}{n-1} \sum (x - \bar{x})(x - \bar{x})}$$

Zentrierte Daten:

$\bar{x} = 9.1$	$x - \bar{x}$	-6.1	-1.1	-0.1	3.9	-6.1	-3.1	1.9	11.9	-8.1	6.9
$\bar{y} = 55.4$	$y - \bar{y}$	-25.4	1.6	8.6	16.6	-19.4	-12.4	3.6	34.6	-35.4	27.6
$(x - \bar{x}) \cdot (y - \bar{y})$		154.94	-1.76	-0.86	64.74	118.34	38.44	6.84	411.74	286.74	190.44
$(x - \bar{x})^2$		37.21	1.21	0.01	15.21	37.21	9.61	3.61	141.61	65.61	47.61

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})(x - \bar{x})} = \frac{1269.6}{358.9} \approx 3.53747562 = \alpha$$

Zentrierte Daten:

$\bar{x} = 9.1$	$x - \bar{x}$	-6.1	-1.1	-0.1	3.9	-6.1	-3.1	1.9	11.9	-8.1	6.9
$\bar{y} = 55.4$	$y - \bar{y}$	-25.4	1.6	8.6	16.6	-19.4	-12.4	3.6	34.6	-35.4	27.6
	$(x - \bar{x}) \cdot (y - \bar{y})$	154.94	-1.76	-0.86	64.74	118.34	38.44	6.84	411.74	286.74	190.44
	$(x - \bar{x})^2$	37.21	1.21	0.01	15.21	37.21	9.61	3.61	141.61	65.61	47.61

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})(x - \bar{x})} = \frac{1269.6}{358.9} \approx 3.53747562 = \alpha$$

$$\beta = \bar{y} - \alpha\bar{x} \approx 23.2089719$$

Zentrierte Daten:

$\bar{x} = 9.1$	$x - \bar{x}$	-6.1	-1.1	-0.1	3.9	-6.1	-3.1	1.9	11.9	-8.1	6.9
$\bar{y} = 55.4$	$y - \bar{y}$	-25.4	1.6	8.6	16.6	-19.4	-12.4	3.6	34.6	-35.4	27.6
$(x - \bar{x}) \cdot (y - \bar{y})$		154.94	-1.76	-0.86	64.74	118.34	38.44	6.84	411.74	286.74	190.44
$(x - \bar{x})^2$		37.21	1.21	0.01	15.21	37.21	9.61	3.61	141.61	65.61	47.61
$\alpha x + \beta \approx$		33.82	51.51	55.05	69.20	33.82	44.43	62.12	97.50	26.74	79.81

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})(x - \bar{x})} = \frac{1269.6}{358.9} \approx 3.53747562 = \alpha$$

$$\beta = \bar{y} - \alpha\bar{x} \approx 23.2089719$$

Zentrierte Daten:

$\bar{x} = 9.1$	$x - \bar{x}$	-6.1	-1.1	-0.1	3.9	-6.1	-3.1	1.9	11.9	-8.1	6.9
$\bar{y} = 55.4$	$y - \bar{y}$	-25.4	1.6	8.6	16.6	-19.4	-12.4	3.6	34.6	-35.4	27.6
$(x - \bar{x}) \cdot (y - \bar{y})$		154.94	-1.76	-0.86	64.74	118.34	38.44	6.84	411.74	286.74	190.44
$(x - \bar{x})^2$		37.21	1.21	0.01	15.21	37.21	9.61	3.61	141.61	65.61	47.61
$\alpha x + \beta \approx$		33.82	51.51	55.05	69.20	33.82	44.43	62.12	97.50	26.74	79.81
reales y		30.	57.	64.	72.	36.	43.	59.	90.	20.	83.
$(\alpha x + \beta - y)^2 \approx$		14.60	30.15	80.17	7.86	4.75	2.06	9.74	56.19	45.51	10.19

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})(x - \bar{x})} = \frac{1269.6}{358.9} \approx 3.53747562 = \alpha$$

$$\beta = \bar{y} - \alpha \bar{x} \approx 23.2089719$$

Zentrierte Daten:

$\bar{x} = 9.1$	$x - \bar{x}$	-6.1	-1.1	-0.1	3.9	-6.1	-3.1	1.9	11.9	-8.1	6.9
$\bar{y} = 55.4$	$y - \bar{y}$	-25.4	1.6	8.6	16.6	-19.4	-12.4	3.6	34.6	-35.4	27.6
$(x - \bar{x}) \cdot (y - \bar{y})$		154.94	-1.76	-0.86	64.74	118.34	38.44	6.84	411.74	286.74	190.44
$(x - \bar{x})^2$		37.21	1.21	0.01	15.21	37.21	9.61	3.61	141.61	65.61	47.61
$\alpha x + \beta \approx$		33.82	51.51	55.05	69.20	33.82	44.43	62.12	97.50	26.74	79.81
reales y		30.	57.	64.	72.	36.	43.	59.	90.	20.	83.
$(\alpha x + \beta - y)^2 \approx$		14.60	30.15	80.17	7.86	4.75	2.06	9.74	56.19	45.51	10.19

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})(x - \bar{x})} = \frac{1269.6}{358.9} \approx 3.53747562 = \alpha$$

$$\beta = \bar{y} - \alpha \bar{x} \approx 23.2089719$$

Summe quadratischer Fehler: ≈ 261.220

Wurzel aus mittlerem quadratischen Fehler: ≈ 5.11097

Prognose:

$$\text{Formel: } \hat{y} = 3.53747562 \cdot x + 23.2089719$$

Prognose:

$$\text{Formel: } \hat{y} = 3.53747562 \cdot x + 23.2089719$$

A) $x = 20$

Prognose:

$$\text{Formel: } \hat{y} = 3.53747562 \cdot x + 23.2089719$$

A) $x = 20 \Rightarrow \hat{y} \approx 93.9585$

B) $x = 8$

Prognose:

$$\text{Formel: } \hat{y} = 3.53747562 \cdot x + 23.2089719$$

A) $x = 20 \Rightarrow \hat{y} \approx 93.9585$

B) $x = 8 \Rightarrow \hat{y} \approx 51.5088$

C) $x = 11$

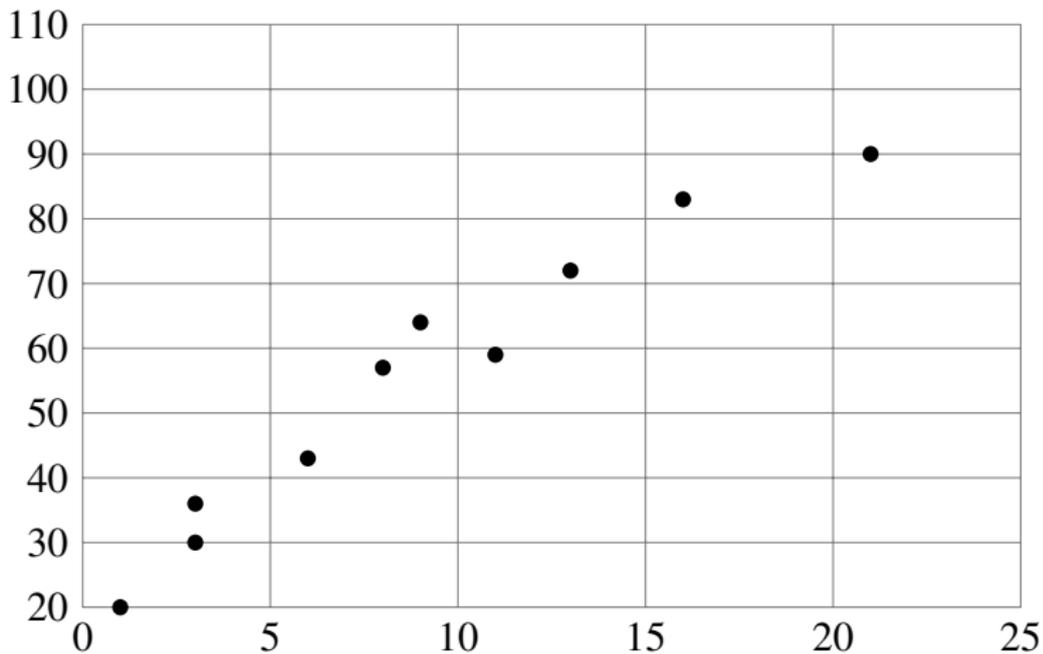
Prognose:

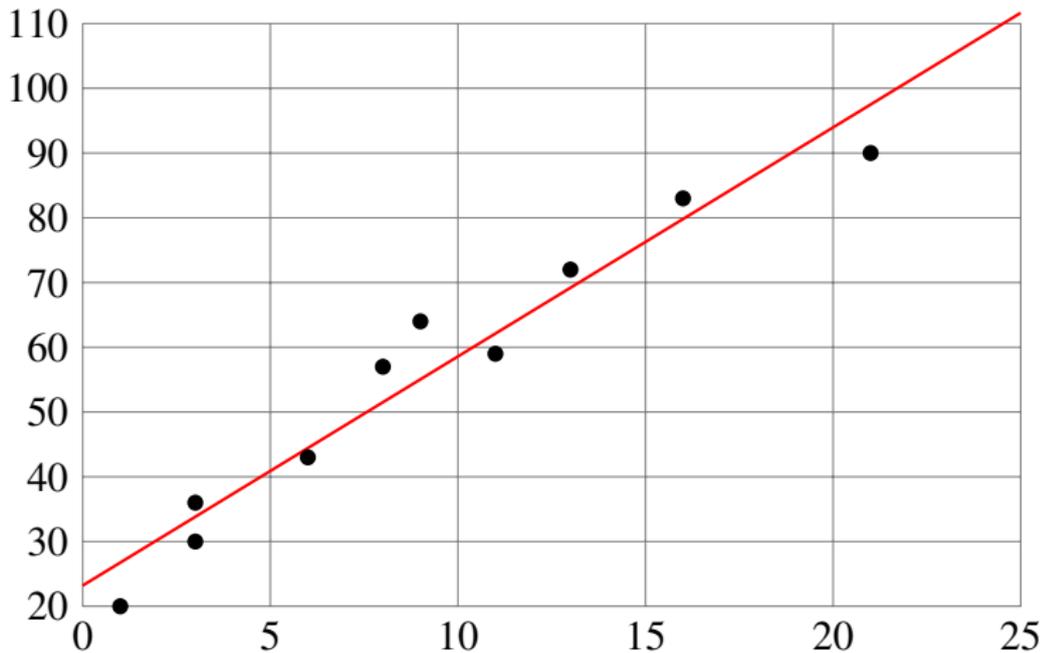
$$\text{Formel: } \hat{y} = 3.53747562 \cdot x + 23.2089719$$

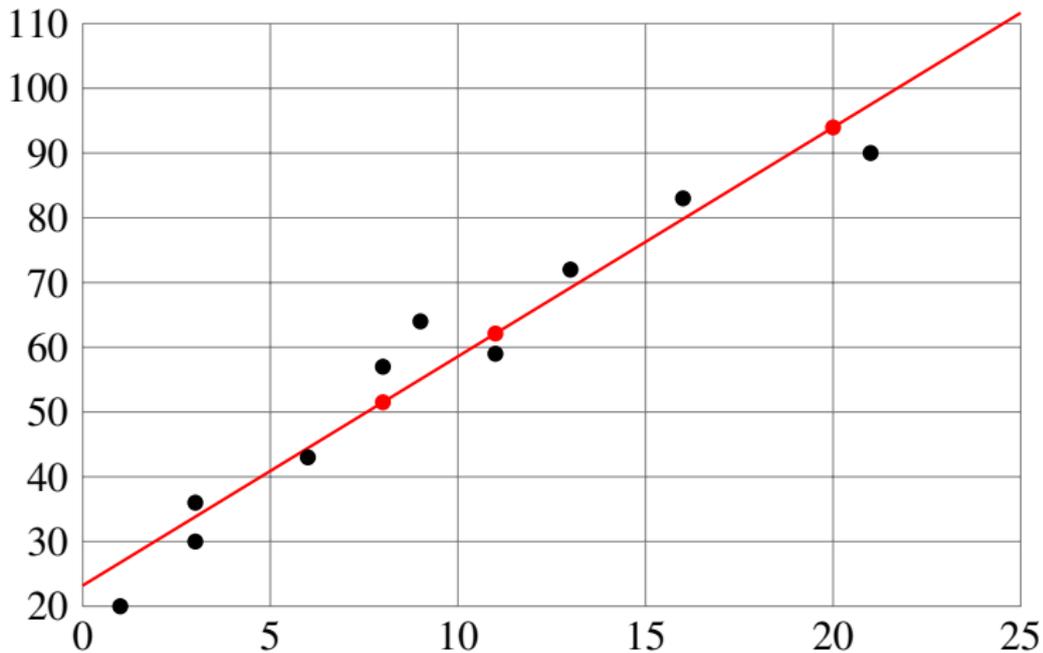
$$\text{A) } x = 20 \Rightarrow \hat{y} \approx 93.9585$$

$$\text{B) } x = 8 \Rightarrow \hat{y} \approx 51.5088$$

$$\text{C) } x = 11 \Rightarrow \hat{y} \approx 62.1212$$







```
import numpy as np
x = np.array([ 3,  8,  9, 13,  3,  6, 11, 21,  1, 16])
y = np.array([30, 57, 64, 72, 36, 43, 59, 90, 20, 83])

x.mean(), y.mean()
# (9.0999999999999996, 55.399999999999999)

x-x.mean()
# array([-6.1, -1.1, -0.1,  3.9, -6.1, ...

(x-x.mean())*(y-y.mean())
# array([ 154.94, -1.76, -0.86,  64.74, ...

(x-x.mean())**2
# array([ 3.72100000e+01,  1.21000000e+00, ...

sum((x-x.mean())*(y-y.mean())), sum((x-x.mean())**2)
# (1269.6000000000001, 358.90000000000003)
```

```
alpha=sum((x-x.mean())*(y-y.mean()))/sum((x-x.mean())**2)
beta =y.mean() - alpha * x.mean()
alpha, beta
# (3.5374756199498467, 23.208971858456394)
```

```
alpha * x + beta
# array([ 33.82139872,  51.50877682,  55.04625244, ...
```

```
(alpha * x + beta - y) ** 2
# array([ 14.60308816,  30.15353203,  80.1695954 , ...
```

```
sum(((alpha * x + beta - y) ** 2))
# 261.22095291167477
```

```
np.sqrt(sum(((alpha * x + beta - y) ** 2))/len(x))
# 5.1109779192604101
```

```
np.array([20, 8, 11]) * alpha + beta
# array([ 93.95848426,  51.50877682,  62.12120368])
```

(The first two lines are the actual regression, the last is prediction!)

Es gibt zahlreiche Varianten von Regression!
Hier: ein optimaler KQ-Schätzer bekannt.
Oftmals: numerische Suche nach lokalem Maximum!

- ▶ quadratischer Fehler vs linearer Fehler
- ▶ Polynome statt linearen Funktionen
- ▶ Regularisierung, um Overfitting zu vermeiden
- ▶ RANSAC – RANdom SAMple Consensus, robuster gegen Ausreißer
- ▶ Gradient descent für differenzierbare Funktionen
- ▶ u.v.m.

Mehr in der Vorlesung “Maschinelles Lernen”, denn Regression braucht Trainingsdaten, und “lernt”!