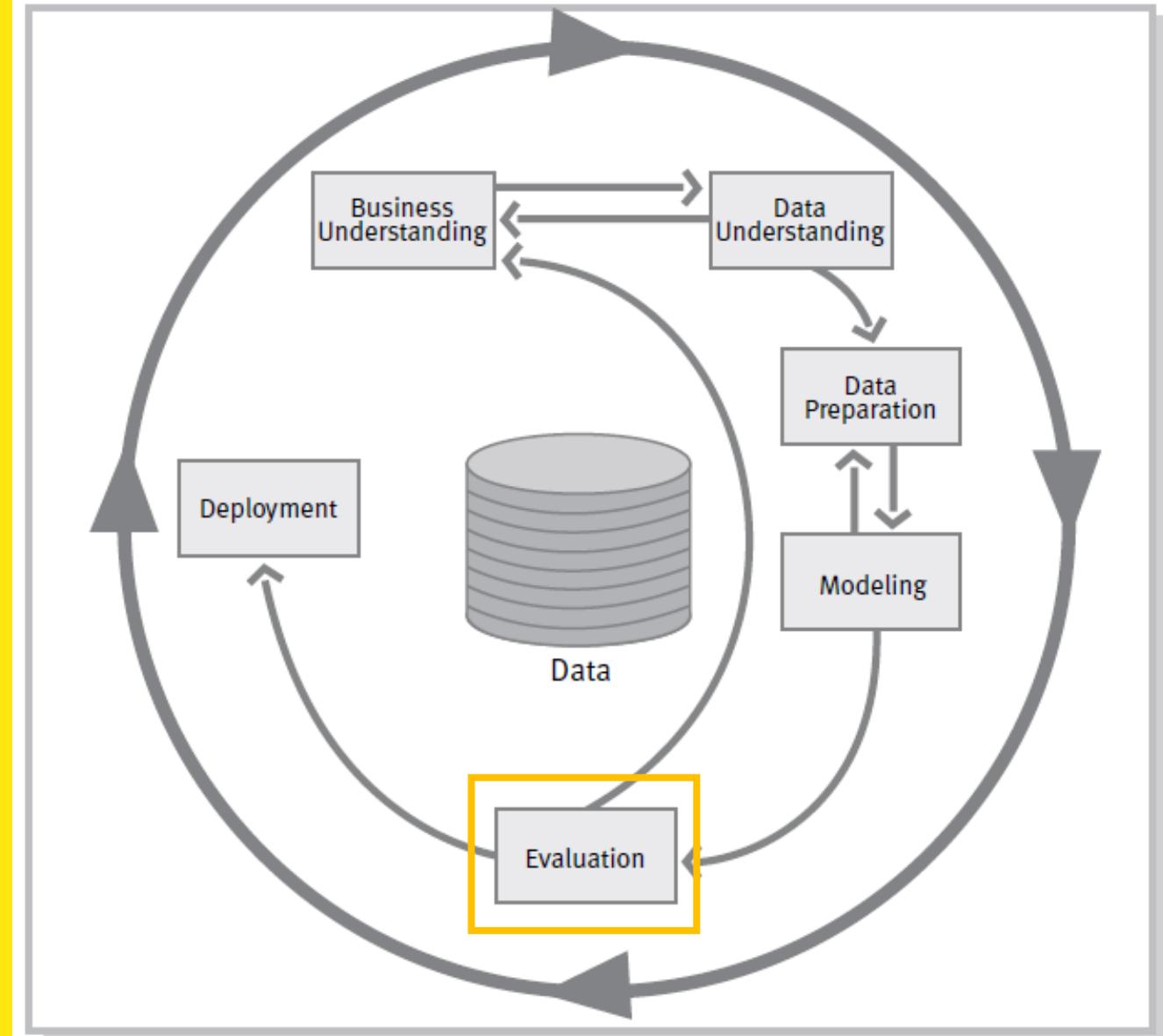


Maschinelles Lernen – Evaluierung

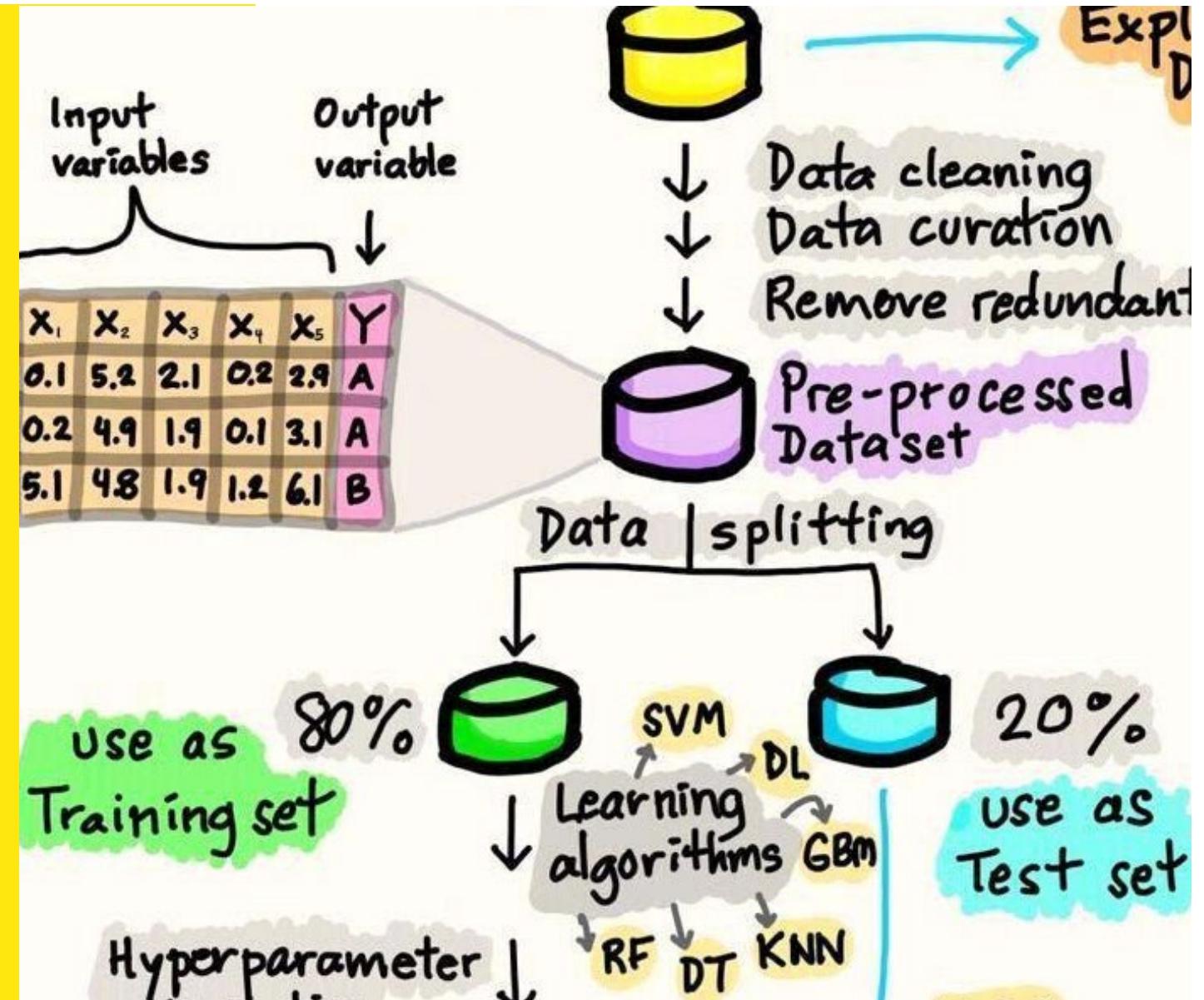
– Hans Friedrich Witschel, Andreas Martin



Testverfahren und -metriken

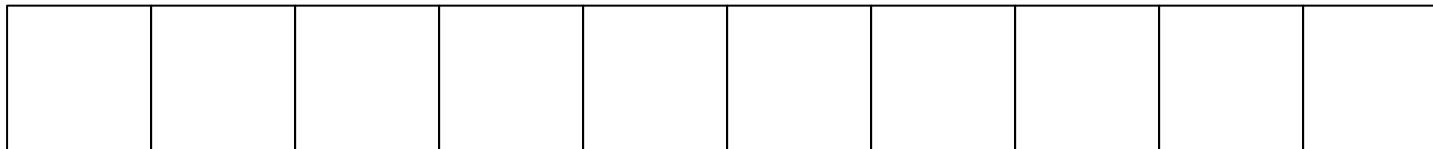
- Verfahren
 - a. Holdout, Kreuzvalidierung
 - b. Lernkurven
- Metriken
 - a. Genauigkeit / Fehlerrate
 - b. Precision, Recall, F-measure
 - c. Area Under (ROC) Curve
 - d. Kosten

Testverfahren



Holdout und Kreuzvalidierung

- **Grundidee:** Qualität der Vorhersagen auf **unbekannten** Daten evaluieren!
- Typische **Holdout-Evaluierung:** nutze 80% der Daten zum Trainieren, den Rest zum Testen
- **Kreuzvalidierung**, z.B. 10-fach:

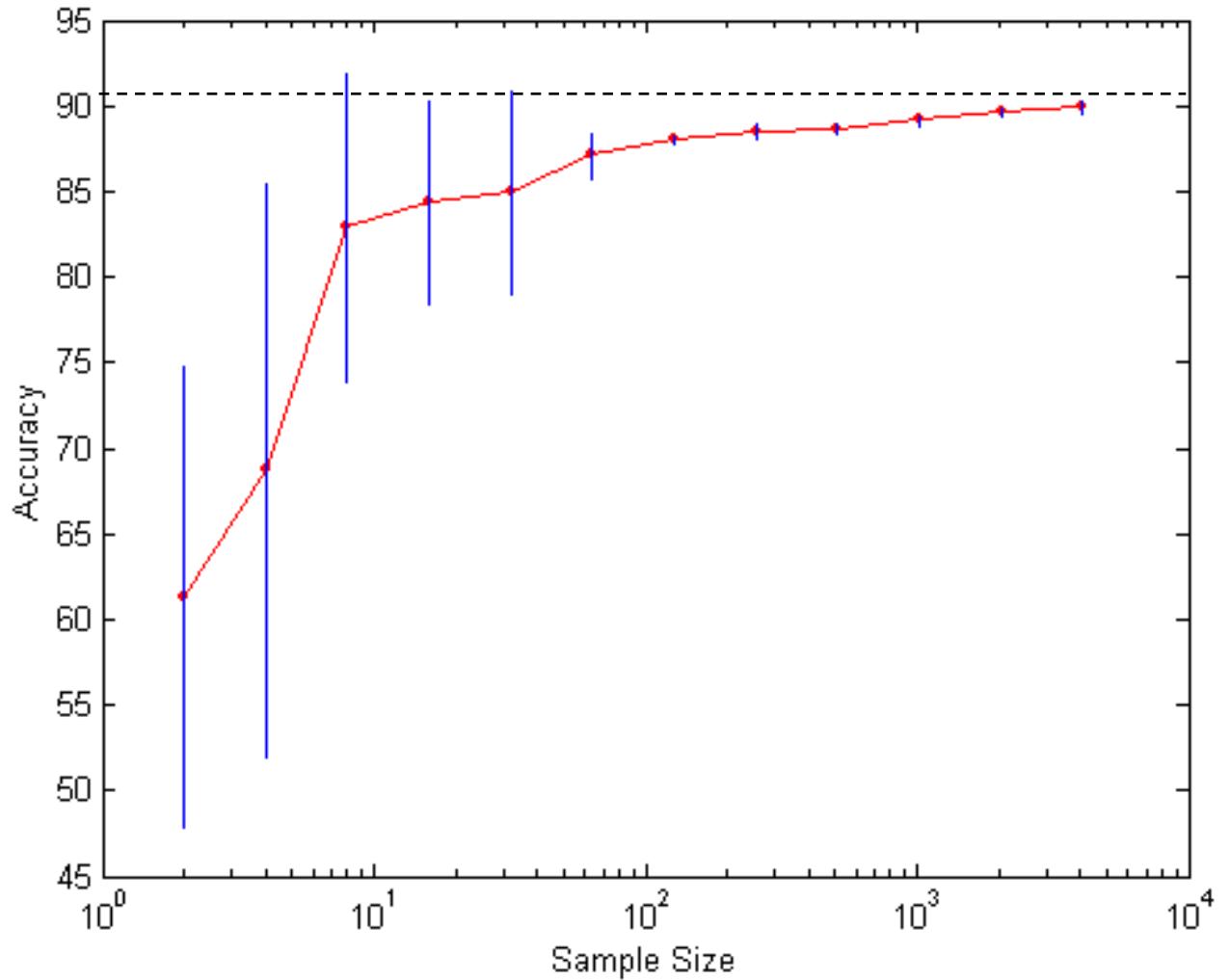


Training

Test

Lernkurve

- Zeigt, wie die Vorhersagegenauigkeit von der Grösse der Trainingsmenge abhängt
- Wichtig, wenn Trainings- und Testdaten manuell vorklassifiziert werden müssen



Testmetriken

CONFUSION MATRIX

		Actual
		Class1
Predicted	Class1	183
	Class2	141
Predicted	Class2	13
	Class1	663

DETAILS

Sensitivity	Specificity	Precision	Recall	F1
0.934	0.825	0.565	0.934	0.704
Accuracy			Kappa	
0.846			0.608	

Genauigkeit bzw. Fehlerrate

- $\text{Genauigkeit} = \frac{\text{Anzahl korrekt klassifizierte Instanzen}}{N}$
- $\text{Fehlerrate} = \frac{\text{Anzahl inkorrekt klassifizierte Instanzen}}{N} = 1 - \text{Genauigkeit}$

Dabei ist N jeweils die Grösse der Testmenge

Genauigkeit: immer eine gute Metrik?

- Wir betrachten eine Spendenkampagne (Daten des KDD-Cup 1998). Personen werden durch folgende Attribute beschrieben:
 - a. Hobbys
 - b. Anzahl der Fälle, in denen sie in der Vergangenheit auf bestimmte Arten von Angeboten geantwortet haben
 - c. Geschlecht

	<p>The following variables indicate the number of known times the donor has responded to other types of mail order offers.</p> <p>MBCRAFT MBGARDEN MBBOOKS MBCOLECT MAGFAML MAGFEM MAGMALE <small>MBGARDEN</small></p> <p>Buy Craft Hobby Buy Gardening Buy Books Buy Collectables Buy General Family Mags Buy Female Mags Buy Sports Mags <small>Buy Sports Mags</small></p> <p><small>Condoning Books</small></p>	<p>PHOTO</p>	<p>The following variables reflect donor interests, as collected from third-party data sources</p> <p>COLLECTABLE (Y/N) VETERANS (Y/N) BIBLE READING (Y/N) SHOP BY CATALOG (Y/N) WORK FROM HOME (Y/N) HOUSEHOLD PETS (Y/N) CD PLAYER OWNERS (Y/N) STEREO/RECORDS/TAPES/CD (Y/N) HOME PC OWNERS/USERS PHOTOGRAPHY (Y/N)</p>
--	---	--------------	--

- Wir sagen erst das Geschlecht voraus, dann das Interesse an Male Magazines
 - a. Warum ist die Genauigkeit so unterschiedlich? Wie gut sind die Klassifikatoren wirklich?

Confusion Matrix, Precision, Recall und F-Measure

$$\text{Precision (p)} = \frac{a}{a + c}$$

$$\text{Recall (r)} = \frac{a}{a + b}$$

$$\text{F - measure (F)} = \frac{2rp}{r + p} = \frac{2a}{2a + b + c}$$

		VORHERGESAGTE KLASSE	
ECHTE KLASSE		Ja	Nein
	Ja	a	b
	Nein	c	d

a: TP (true positive)

b: FN (false negative)

c: FP (false positive)

d: TN (true negative)

Precision + Recall: in Worten

- **Precision:** wieviele der als «Ja» vorhergesagten Instanzen sind wirklich welche?
z.B. Fraud Detection, Precision = 0.84 → 84% der vorhergesagten Betrugsfälle («Warnungen») sind wirklich betrügerisch
- **Recall:** wieviele der echten «Ja»-Fälle wurden gefunden?
z.B. Fraud Detection, Recall = 0.28 → 28% der echten Betrugsfälle wurden gefunden

Area Under (ROC) Curve

- **Grundidee:**
 - a. Klassifikatoren geben (fast) immer einen Score (Wahrscheinlichkeit oder Konfidenz) aus, nach dem man die Instanzen sortieren kann
 - b. AUC misst, inwieweit die positiven Beispiele nach oben sortiert werden
 - c. Formal: $AUC = \text{Wahrscheinlichkeit, dass ein zufällig gewähltes positives Beispiel höher als ein zufällig gewähltes negatives Beispiel gerankt wird}$
- **Wertebereich:**
 - a. Zufällige Sortierung: $AUC = 0.5$
 - b. Optimale Sortierung (erst alle positiven, dann alle negativen): $AUC = 1$

Area Under (ROC) Curve: Berechnung

- Lasse den Klassifikator für jede Instanz A die Wahrscheinlichkeit ausgeben, dass A positiv ist (z.B. Interesse am Wintercheck hat)
- Gehe die sortierte Liste durch, nimm jeweils an, dass die aktuelle Position der Schwellenwert für die Klassifizierung ist und berechne
 - a. True Positive Rate / TPR = Recall = $TP/(TP+FN)$
 - b. False Positive Rate / FPR = $FP/(FP + TN)$

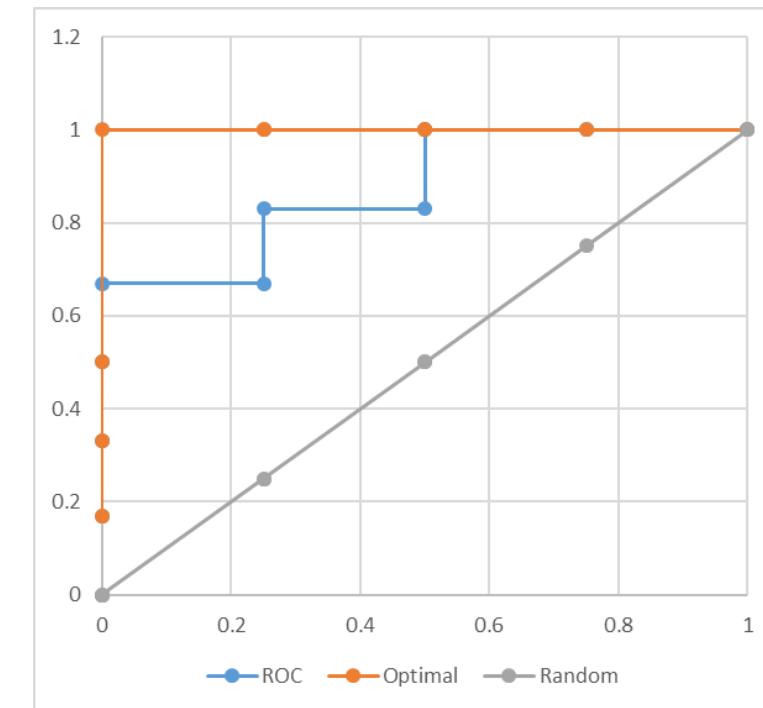
Instanz A	$P(+ A)$	Echte Klasse
1	0.95	+
2	0.93	+
3	0.87	-
4	0.85	-
5	0.85	-
6	0.85	+
7	0.76	-
8	0.53	+
9	0.43	-
10	0.25	+

Area Under (ROC) Curve: Beispiel

Instanz	P(+ A)	Echte Klasse	TPR	FPR
1	0.95	+	$1/6 = 0.17$	0
2	0.93	+	$2/6 = 0.33$	0
3	0.87	+	$3/6 = 0.5$	0
4	0.85	+	$4/6 = 0.67$	0
5	0.84	-	$4/6 = 0.67$	$1/4 = 0.25$
6	0.83	+	$5/6 = 0.83$	$1/4 = 0.25$
7	0.76	-	$5/6 = 0.83$	$2/4 = 0.5$
8	0.53	+	1	$2/4 = 0.5$
9	0.43	-	1	$3/4 = 0.75$
10	0.25	-	1	1

Wieviele Interessenten wurden bis hierher gefunden?

Wieviele «falsche Alarme» wurden bis hierher generiert?



Kostenbasierte Evaluation: Idee

- Oft wird ML eingesetzt, um Geld zu verdienen oder zu sparen. Daher resultieren aus Klassifikator-Vorhersagen (unterschiedliche) Gewinne oder Verluste.
 - Beispiel Swiss Bikes:
 - a. Jeder Wintercheck erzielt einen Umsatz von 100 CHF und einen Gewinn von 40 CHF für Swiss Bikes
 - b. Versand eines Werbebriefs kostet 1 CHF
- Kostenmatrix

The diagram illustrates the transition from a standard confusion matrix to a cost matrix. On the left, a confusion matrix is labeled «Fehlerrate» (error rate). It has two rows for the "True" class and two columns for the "Predicted class". The first row is "True" class and the second row is "Predicted class". The first column is "True" class and the second column is "Predicted class". The matrix values are: True class (Yes, Yes) = 0, True class (Yes, No) = 1, Predicted class (Yes, Yes) = 1, Predicted class (Yes, No) = 0. A large blue arrow points to the right, leading to a second confusion matrix labeled «Kosten» (costs). This matrix has two rows for the "true" class and two columns for the "Predicted class". The first row is "true" class and the second row is "Predicted class". The first column is "true" class and the second column is "Predicted class". The matrix values are: true class (Yes, Yes) = -39, true class (Yes, No) = 0, Predicted class (Yes, Yes) = 1, Predicted class (Yes, No) = 0.

		Predicted class	
		Yes	No
“True” class	Yes	0	1
	No	1	0

		Predicted class	
		Yes	No
“true” class	Yes	-39	0
	No	1	0

Berechnung der Kosten

- Kostenmatrix

		Predicted class	
		Yes	No
“true” class	Yes	-39	0
	No	1	0

- Confusion matrix auf Testmenge:

		Predicted class	
		Yes	No
“true” class	Yes	12	20
	No	67	101

Welche Kosten ergeben sich aus den Vorhersagen des Klassifikators auf der Testmenge?

Übung: Kostenmatrizen formulieren

Wie sieht die Kostenmatrix in den folgenden Fällen aus?

- erste Hälfte der Gruppen: Kreditvergabe in einer Bank
- zweite Hälfte der Gruppen: Aufdeckung von Versicherungsbetrug in einer Unfallversicherung

Macht nötige Annahmen über die jeweils relevanten Kosten!