



max planck institut
informatik

Ideen und Konzepte der Informatik

Künstliche Intelligenz

Maschinelles Lernen

Kurt Mehlhorn

Übersicht

- Was ist Künstliche Intelligenz und Maschinelles Lernen?
- Beispiele für den Stand der Kunst
- Teil I, einfache Methoden
 - Regel von Bayes und Spamerkennung
 - Handschriftenerkennung mit und ohne Trainingsdaten
 - Gesichts- und Personenerkennung
- Teil II, Neuronale Netzwerke
 - Maschinelles Sehen
- Teil III, verstärkendes Lernen (Reinforcement Learning)
 - Schach

Künstliche Intelligenz und M-Lernen

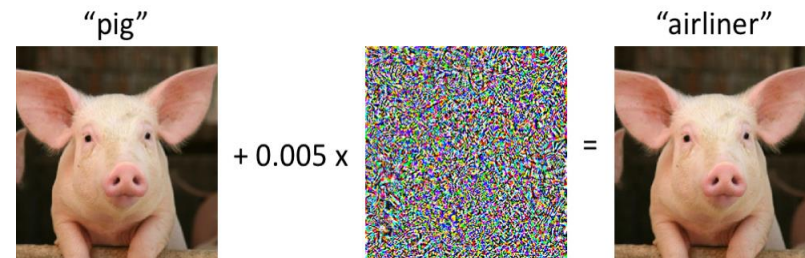
- Künstliche Intelligenz = intelligentes Verhalten von Maschinen
 - Schachspielen, Sprache verstehen, Autofahren, Rätsel lösen, Gefühle verstehen, ..., alles Verhalten, das Menschen ausmacht. Name spielt Alltagswissen herunter.
- Lernen = Fähigkeit, Verhalten zu verbessern aufgrund von Erfahrungen; Verallgemeinern von Erfahrungen; aus Daten wird Verhalten
- Zieht in den Alltag ein: Verkaufsempfehlungen, Sprachschnittstelle, personalisierte Nachrichten,

Riesiges Potential

- Suchmaschinen mit Bild / Sprach-Anfragen
 - Personenerkennung, Überwachung
 - Bessere Benutzerschnittstellen, Sprache, Gesten
 - Selbstfahrende Autos, Roboter
 - Maschinelle Übersetzung
 - Sachbearbeiter
 - Persönliche Assistenten
- “If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future.”
— **Andrew Ng (2016)**
- Aber wie steht es mit den Aufgaben, bei denen Menschen gar nicht denken? Alltagswissen wird unterschätzt.

Aber auch große Gefahren

- Missbrauch (Überwachungsstaat à la China, Verlust der Privatsphäre, Google, Facebook,)
- Lernende Verfahren können in der Regel ihre Entscheidungen nicht erklären.
- Gewaltige Änderungen in der Arbeitswelt, z.B.: Sachbearbeiter, Lastwagenfahrer, ...
- Verfahren werden überschätzt.
- Verantwortung: „Die KI hat ...“



KI-Winter (keine Forschungsgelder für KI)

- KI-Forscher lieben kühne Behauptungen. Haben sich oft nicht erfüllt. Das führte zu KI-Wintern.
- H.A. Simon (Turing Award 75, Nobel Prize 78): stated in 57, that by 67 a computer would be world champion in chess, would discover and prove an important new mathematical theorem, and most theories in psychology would take the form of computer programs.
- Weitere kühne Behauptungen: sehr gute maschinelle Übersetzung ist möglich, Perceptrons (künstliche neuronale Netze) bilden Gehirn nach und liefern ähnliche Leistung. The spirit is willing, but the flesh is weak = the vodka is good but the meat is rotten,
- In the 80ern: Expertensysteme ersetzen menschliche Experten.
- KM glaubt: Der jetzige Hype wird bleiben.

Arten von Lernen

- **Überwachtes Lernen (Supervised Learning):** mit Trainingsdaten oder sogar mit Lehrer.
 - Das ist ein Auto, das ist ein Affe.
- **Unüberwachtes Lernen (Unsupervised Learning):** ohne Trainingsdaten; dann mehr Entdecken als Lernen.
 - Ich entnehme den Daten, dass es in diesem Alphabet 26 verschiedene Buchstaben gibt.
- **Bestärkendes Lernen (Reinforcement Learning):** Verstärkung durch Belohnung/Strafe
 - Gut gemacht!

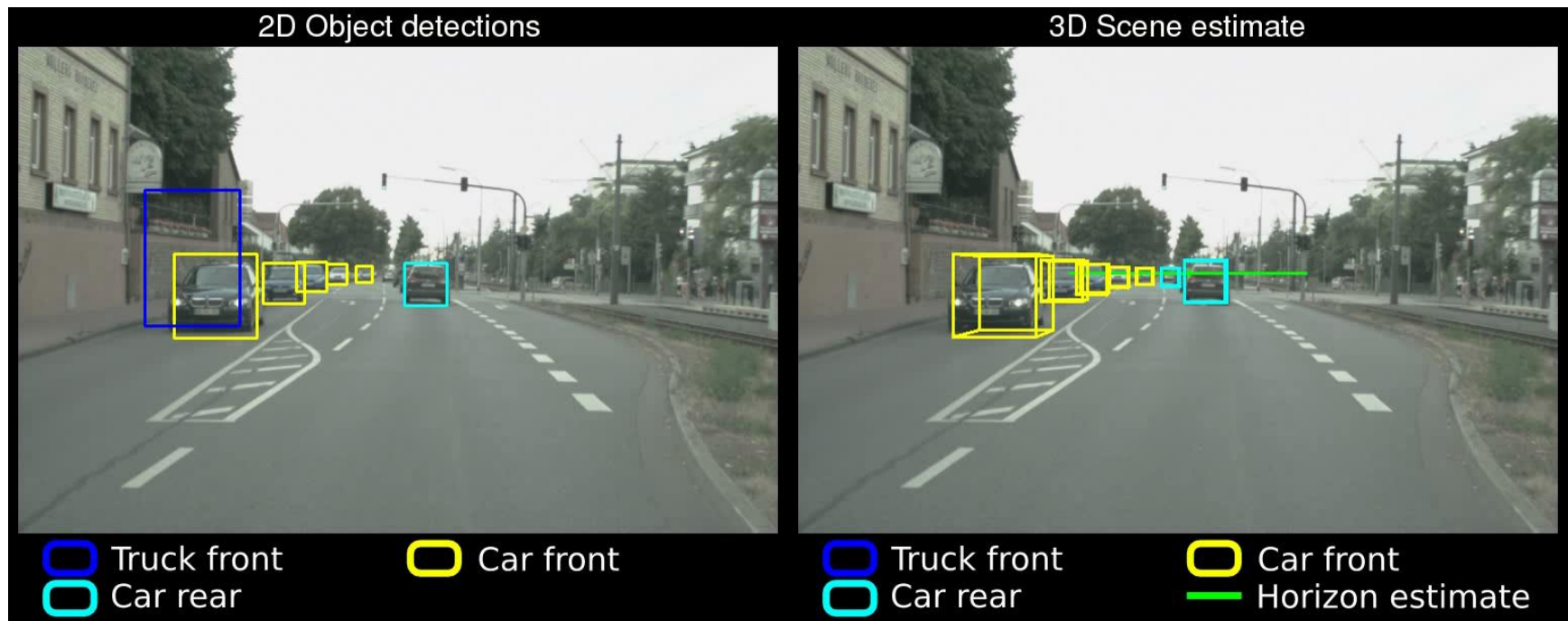


Der Stand der Kunst

- Personenerkennung
- Sprachübersetzung
- Klassifikation von Bildern
- Suchen in Bildersammlungen
- Spiele, Planen und strategisches Verhalten

Objekterkennung (Beispiel von 2014)

Abteilung Schiele: MPI für Informatik



Personenerkennung (2014)

Abteilung Schiele: MPI für Informatik

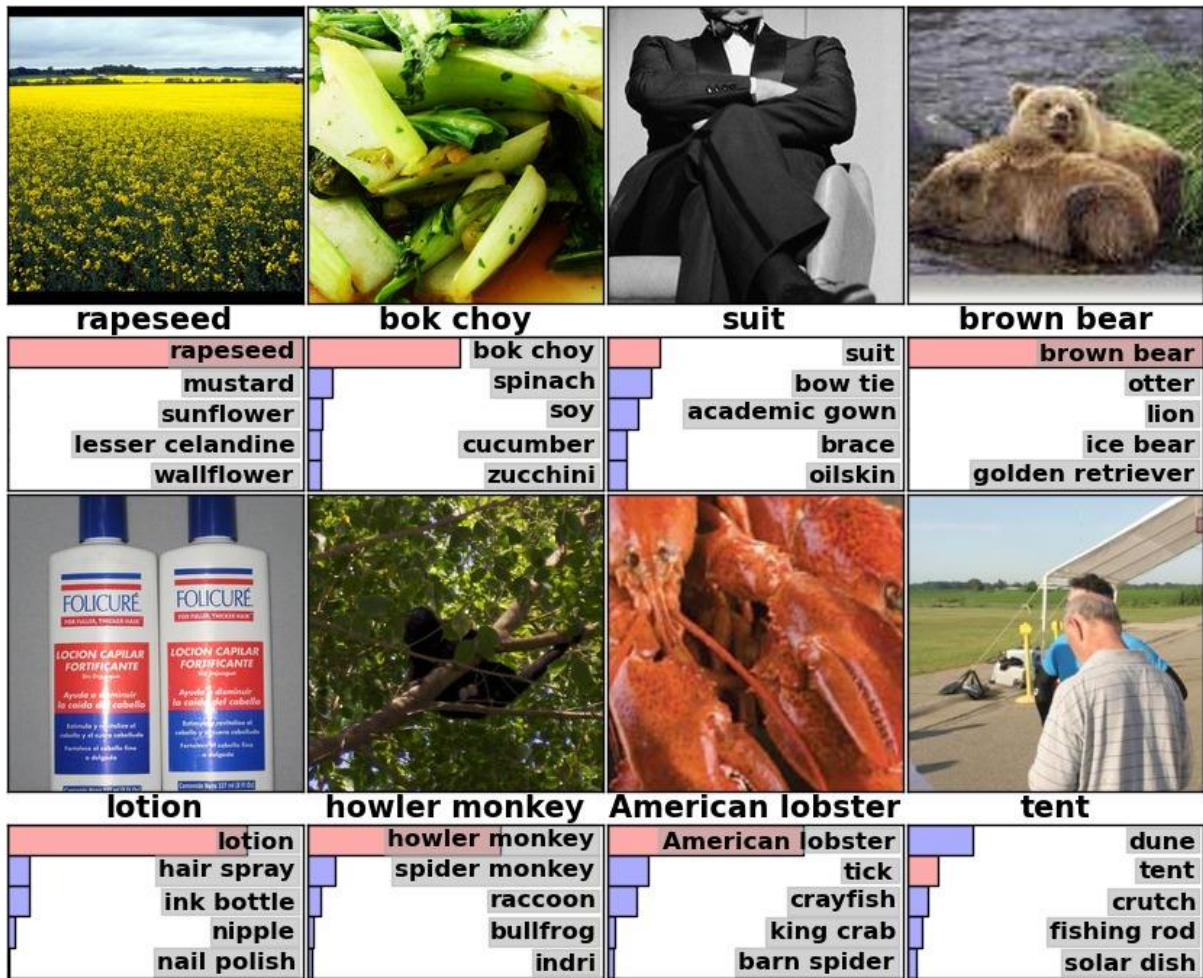


Übersetzung durch DeepL (2020)

Noch immer bedrohen die Buschbrände in Australien zahlreiche Küstenorte. Die Behörden haben nun eine Sperrzone eingerichtet, die Urlauber unverzüglich räumen müssen. Es sei die "größte Massenverlegung" der Region. Die Bilder aus dem Militärhubschrauber zeigen brennende Wälder soweit das Auge reicht. Hochhaushohe Flammenwände, dichter Qualm verdunkelt den Himmel. Die Buschbrände in Australien haben bis zu 1300 Häuser zerstört, aber so genau weiß man es nicht.

The bush fires in Australia are still threatening numerous coastal towns. The authorities have now set up a restricted zone, which holidaymakers must evacuate immediately. It is said to be the "largest mass relocation" in the region. The pictures from the military helicopter show burning forests as far as the eye can see. Skyscraper-high walls of flames, dense smoke darkens the sky. The bush fires in Australia have destroyed up to 1300 houses, but it is not known for sure.

Klassifikation (Krizhevsky et al., 2012)



Trainingsdaten: 1.2 Millionen Bilder aus 1000 Klassen.

Aufgabe: ordne neue Bilder der korrekten Klasse zu.

Hinweis: Bilder zeigen im Wesentlichen ein Objekt.

Suche (Krizhevsky et al., 2012)

Linke Spalte: Suchanfrage, andere Spalten: ähnliche Bilder im Fundus



Spiele

- AlphaGo schlägt besten menschlichen Spieler 3:0.
- Hat sich das Spielen selbst beigebracht.
 - Anfangswissen = Regeln von Schach + gelernt werden soll eine Funktion Stellung → Zug, die Belohnung maximiert.
 - Danach Spiel gegen sich selbst für einige Tage.
 - Methode: Reinforcement Learning, siehe Teil 3.
- Hat Muster und Spielzüge gefunden, die Menschen bisher verborgen blieben.
- Schachmeister nutzen Computer für das Training.
- Schach hat nichts von seiner Faszination verloren.

Regel von Bayes + Spamerkennung

Spam, Junk = unerwünschte Nachrichten

Ham = erwünschte Nachrichten

Bayes-Filter zur Unterscheidung von Ham und Spam. Auch im Alltagsleben sehr nützlich, z. B. zum Bewerten von Testergebnissen.

Spam (ursprünglich): Spiced Ham, Markenname für sehr verbreitetes Dosenfleisch, Monty Python Video

Regel von Bayes

(Thomas Bayes: englischer Pfarrer und Mathematiker, 1701 – 1761)

In einem Sack sind 900 Äpfel und 100 Paprika. Von den Äpfeln sind 10% rot und 90% grün. Bei den Paprika sind es jeweils 50%.

Ich entnehme eine Frucht zufällig. Sie ist rot. Was für eine Frucht ist es?

- Bayes: Entscheide dich für den wahrscheinlicheren Fall.

Regel von Bayes im Alltag

- 5% der Bevölkerung erkranken an der Krankheit X. Es gibt einen Test:
 - Gesunde: 90% negativ, 10% positiv.
 - Kranke: 0% negativ, 100% positiv.

Sie bekommen einen positiven Befund. Was bedeutet das?

Regel von Bayes

In einem Sack sind 900 Äpfel und 100 Paprika. Von den Äpfeln sind 10% rot und 90% grün. Bei den Paprika sind es jeweils 50%.

Bruchteil der Äpfel unter den roten Früchten = $\frac{\text{\# rote Äpfel}}{\text{\# rote Früchte}}$

$P(\text{Apfel} \mid \text{rot}) = \text{Bruchteil der Äpfel unter den roten Früchten}$

Regel von Bayes, math. Formel

In einem Sack sind 900 Äpfel und 100 Paprika. Von den Äpfeln sind 10% rot und 90% grün. Bei den Paprika sind es jeweils 50%.

$$\begin{aligned} P(\text{Apfel} \mid \text{rot}) &= \frac{P(\text{Apfel und rot})}{P(\text{rot})} \\ &= \frac{P(\text{rot} \mid \text{Apfel}) P(\text{Apfel})}{P(\text{Apfel und rot}) + P(\text{Paprika und rot})} \\ &= \frac{P(\text{rot} \mid \text{Apfel}) P(\text{Apfel})}{P(\text{rot} \mid \text{Apfel}) P(\text{Apfel}) + P(\text{rot} \mid \text{Paprika}) P(\text{Paprika})} \end{aligned}$$

Spam versus Ham (Junk-Mail)

- Absenderbasiert
 - E-Mail von Bekannten ist kein Spam.
 - Schwarze Listen
- Inhaltsbasiert (Beispiel für Supervised Learning)
 - Nutzer klassifiziert E-Mails als gut und schlecht; System lernt daraus; Nutzer muss immer weniger eingreifen.

Inhaltsbasierte Filter

- In der Trainingsphase lernen wir
 - Wahrscheinlichkeit von Ham und Spam.
 - Jeweils Wahrscheinlichkeiten für Worte.
- 70% ist Ham, 30% ist Spam.

- Ham

	Freund	Vorle- sung	Algorith- mus	Geld	Viagra	schnell
	0.1	0.3	0.3	0.1	0.1	0.1

- Spam

	Freund	Vorle- sung	Algorith- mus	Geld	Viagra	schnell
	0.2	0.1	0.1	0.2	0.3	0.1

Trainingsphase

- Nutzer klassifiziert E-Mails als Spam und Ham (damit beide Wahrscheinlichkeiten).
- Sei n die Gesamtlänge meiner guten E-Mails (in Worten), sei v die Anzahl der Vorkommen eines bestimmten Wortes. Dann
- Wahrscheinlichkeit des Wortes in Ham $= \frac{v}{n}$.

Inhaltsbasierte Filter (Bayes-Modell)

- Wahrscheinlichkeitsverteilung auf Worten

- Ham

	Freund	Vorlesung	Algorithmus	Geld	Viagra	schnell
	0.1	0.3	0.3	0.1	0.1	0.1

- Spam

	Freund	Vorlesung	Algorithmus	Geld	Viagra	schnell
	0.2	0.1	0.1	0.2	0.3	0.1

- $P(\text{Text} \mid \text{Ham}) = \text{Produkt der Ham-Wahrscheinlichkeiten der Worte im Text.}$

Inhaltsbasierte Filter

- Ham

Freund	Vorlesung	Algorithmus	Geld	Viagra	schnell
0.1	0.3	0.3	0.1	0.1	0.1

- Spam

Freund	Vorlesung	Algorithmus	Geld	Viagra	schnell
0.2	0.1	0.1	0.2	0.3	0.1

- Text = Viagra Geld Freund
 - $P(\text{Text} \mid \text{Ham}) = 0.1 \cdot 0.1 \cdot 0.1 =$
 - $P(\text{Text} \mid \text{Spam}) =$
- Bei 70% Ham und 30% Spam

Inhaltsbasierte Filter

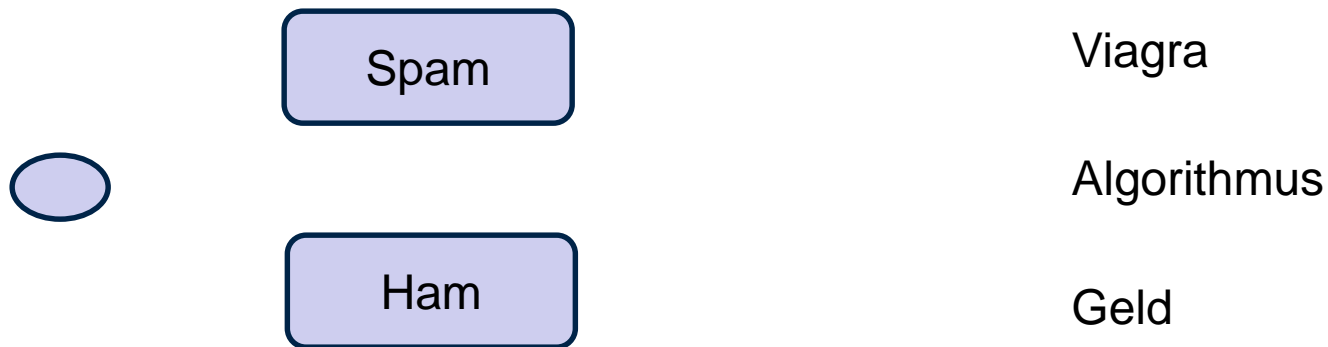
- Viagra Geld Freund
 - $P(\text{Text} | \text{Ham}) = 1/1000$
 - $P(\text{Text} | \text{Spam}) = 12/1000$
- Bei 70% Ham und 30% Spam erhalten wir
- $$P(\text{Spam} | \text{Text}) = \frac{P(\text{Text} | \text{Spam})P(\text{Spam})}{(P(\text{Text} | \text{Ham})P(\text{Ham}) + P(\text{Text} | \text{Spam})P(\text{Spam}))} =$$
$$12 \cdot 3 / (1 \cdot 7 + 12 \cdot 3) = 36/43 \approx 0.84.$$

Nutzungsphase

- Nutzungsphase: System klassifiziert.
- Verteilung wird weiter trainiert (seltene Worte).
- Nutzer kann widersprechen.
- Spammer lernen auch dazu: V!agra statt Viagra.

Zusammenfassung

- Es gibt ein Modell, wie Ereignisse (E-Mails) erzeugt werden:



- Lernen Modell in der Trainingsphase und geben dann für jedes Ereignis die wahrscheinlichste Erklärung (Bayes).
- Klassifizierung in Geschäftspost, Privatpost, Spam.

Jedua Pearl: Turing Award 2012, Bayesian Networks and Causal Inference.



Ziffernerkennung

- Darstellung von Bildern in Rechnern
- Trainingsdaten: handgeschriebene Ziffern
- Überwachtes Lernen (Supervised Learning): mit Label (die Ziffer)
- Unüberwachtes Lernen (Unsupervised Learning): ohne Label

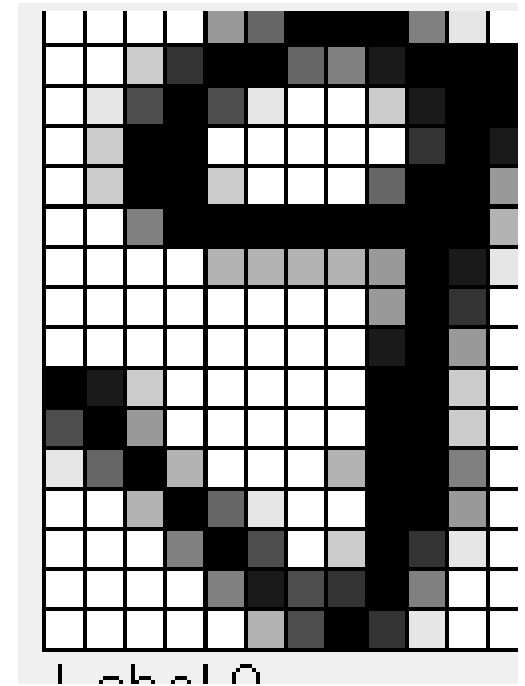
Experimente und erste Version der Folien: Adrian Neumann

Bilder = Matrizen von Zahlen

Ziffer = 12 x 16 Matrix von Grauwerten
in [0,1]

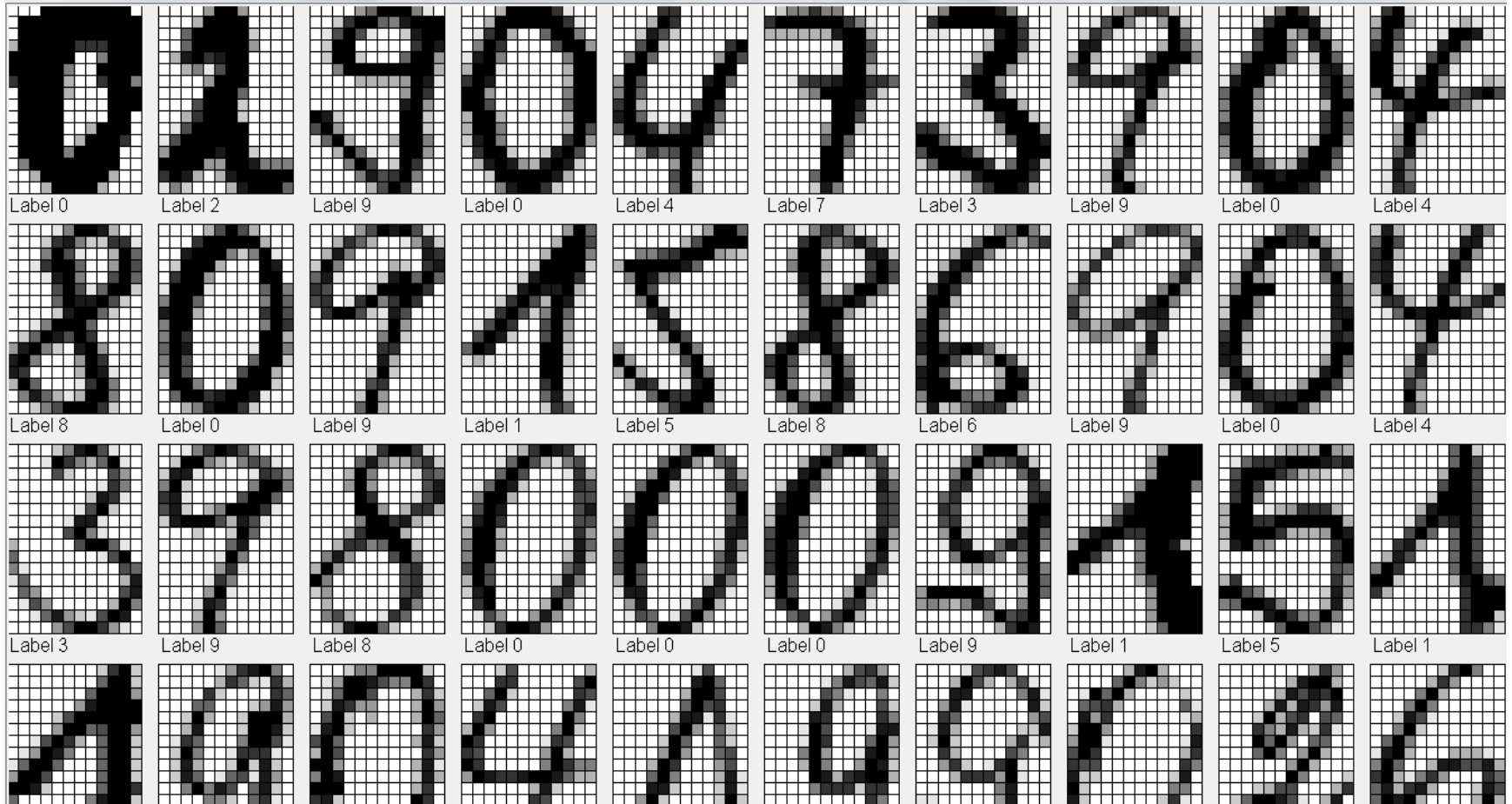
Vektor von Grauwerten der Länge 192

0.0	0.0	0.0	0.2	0.3	0.4	0.8	1.0	1.0	0.7	0.3	0.1
0.0	0.5	0.8	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5
0.3	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9
0.3	1.0	1.0	1.0	1.0	1.0	0.8	1.0	1.0	1.0	1.0	0.8
0.3	1.0	1.0	1.0	1.0	0.8	0.1	0.9	1.0	1.0	0.8	0.2
0.0	0.7	1.0	1.0	1.0	0.8	0.8	1.0	1.0	1.0	0.4	0.0
0.0	0.2	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.3	0.0	0.0
0.0	0.1	0.7	1.0	1.0	1.0	1.0	0.8	0.1	0.0	0.0	0.0
0.0	0.6	1.0	1.0	1.0	1.0	0.9	0.1	0.0	0.0	0.0	0.0
0.6	1.0	1.0	1.0	1.0	1.0	1.0	0.3	0.0	0.0	0.0	0.0
0.8	1.0	1.0	0.5	0.1	0.7	1.0	0.8	0.2	0.0	0.0	0.0
0.5	1.0	1.0	0.3	0.0	0.0	0.9	1.0	0.9	0.1	0.0	0.0
0.4	1.0	1.0	0.3	0.0	0.0	0.5	1.0	1.0	0.5	0.0	0.0
0.0	0.4	1.0	1.0	0.5	0.3	0.5	1.0	1.0	1.0	0.2	0.0
0.0	0.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.1	0.0
0.0	0.0	0.0	0.2	0.5	0.7	1.0	1.0	0.9	0.3	0.0	0.0



Ihr Gehirn sieht
Ziffern, ihr Auge und
Computer sehen nur
eine Matrix von
Grauwerten.

Trainingsdaten



Ziemlich gutmütig, 10000 Beispiele

Grundidee

- Zwei Bilder repräsentieren die gleiche Ziffer, wenn die Bilder sich ähnlich sind.
- Ähnlich = ähnliche Grauwertverteilung.
- Bild = Vektor von Grauwerten.
- Ähnlichkeit der Grauwert-Vektoren:
 - Euklidische Distanz zweier Vektoren
 - Cos-Distanz, Winkel zwischen Vektoren

Ähnlichkeit von Vektoren

- Zwei Vektoren x und y sind ähnlich,
 - wenn $x - y$ kurz ist.
 - wenn der aufgespannte Winkel klein ist.

- Länge eines Vektors $x = (x_1, \dots, x_n)$

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

- Winkel zwischen x und y

$$\cos \alpha = \frac{x \cdot y}{\|x\| \cdot \|y\|},$$

Verfahren: Nearest Neighbor

Um die Bedeutung des Bildes p zu finden, finde das Trainingsbild x mit $\text{dist}(p,x)$ minimal (durch lineare Suche über alle Trainingsdaten) und gib das Label von x aus.

- Erkennungsrate bei Euklidischer Distanz 0.934
- Majority of 3 nearest neighbors 0.945
- Majority of 9 nearest neighbors 0.920

- Mit cos-Distanz 0.940
- Majority of 3 nearest neighbors 0.920

Detaillierte Ergebnisse

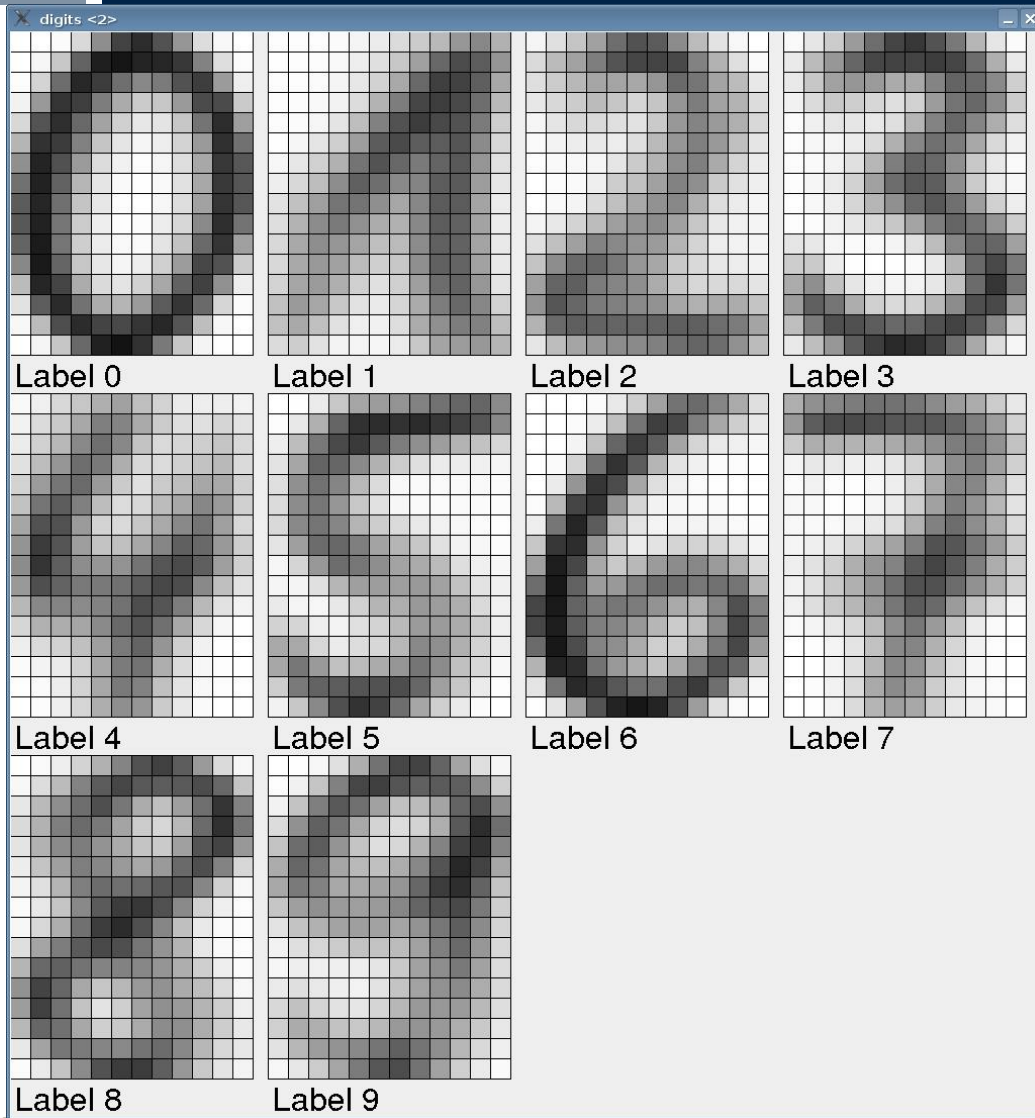
digit = 0 accuracy = 1.0
digit = 1 accuracy = 0.90
digit = 2 accuracy = 0.92
digit = 3 accuracy = 1.0
digit = 4 accuracy = 0.95
digit = 5 accuracy = 0.85
digit = 6 accuracy = 0.84
digit = 7 accuracy = 1.0
digit = 8 accuracy = 0.7
digit = 9 accuracy = 0.94

Klassifizierung ist recht gut,
aber sie dauert sehr **lange**,
da jedes Mal ALLE
Trainingsdaten angeschaut
werden.

Klassen → Klassenzentren

- Vorbereitung: Berechne für jede Klasse (Ziffer) das Klassenzentrum durch Durchschnittsbildung (*siehe nächste Folie*).
- Suche: Finde das nächstgelegene Zentrum (10 Vergleiche).
- Erkennungsrate bei Euklidischer Distanz: 0.683.
- Mit cos-distance: 0.728.
- Sehr effizient, aber deutlich schlechter.

Die Klassenzentren



Durchschnitt (Schwerpunkt einer Klasse)

=

Summe der Elemente der Klasse / Anzahl der Elemente

Verschiedene Schreibweise der gleichen Ziffer: verschwommene Zentren

Mehr als 10 Zentren?

Mehr Klassen ist eine gute Idee.

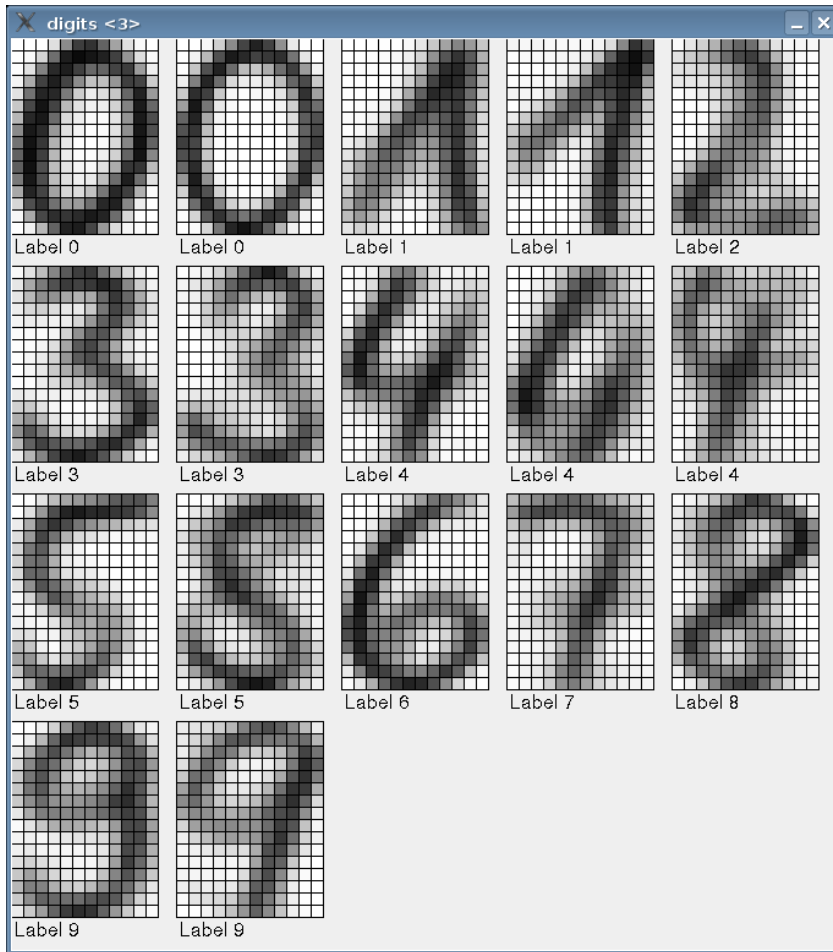
- Danach Vergleich mit den Klassenzentren
 - $k = 10$, Rate 0.683
 - $k = 17$, 0.733

 - Mit cos-distance,
 - $k = 10$, 0.728
 - $k = 17$, 0.783
 - $k = 30$, 0.864
- Aber: Wer macht die Klasseneinteilung?
 - Können wir Klassen entdecken, ohne dass uns Klassenlabels gesagt werden?

 - Automatisches Entdecken von Klassen durch k -Means Algorithmus.

 - **Unsupervised Learning.**

k-Means Algorithmus



- Automatische Klassifizierung in 17 Klassen
- Danach (!!!) Zuweisung eines Labels per Hand und Wegwerfen von schlechten Zentren
- Identifiziert die zwei Schreibweisen der Neun und der Eins

k -Means Algorithmus

Teilt n Punkte in k Cluster (Haufen) ein.

1. Starte mit k beliebigen (zufälligen) Zentren.
2. Weise jeden Punkt dem nächstgelegenen Zentrum zu und bilde so k Cluster.
3. Berechne für jeden Cluster seinen Schwerpunkt; das sind die neuen Zentren.
4. Gehe nach 2.

k -Means Algorithmus

Teilt n Punkte in k Cluster (Haufen) ein.

1. Starte mit k beliebigen (zufälligen) Zentren.
2. Weise jeden Punkt dem nächstgelegenen Zentrum zu und bilde so k Cluster.
3. Berechne für jeden Cluster seinen Schwerpunkt; das sind die neuen Zentren.
4. Gehe nach 2.

Falls ein Cluster leer ist, übernimm den Schwerpunkt unverändert oder wähle einen neues Zentrum zufällig. Falls zwei Zentren identisch, teile Cluster zufällig.

Gesichtserkennung (vor 2010).

- Gesicht \rightarrow Vektor
 - Augenfarbe, Abstand der Augen, Abstand Nase – Augen, Hautfarbe, Abstand Ohren, und so weiter, dann wie bei Ziffernerkennung.
- Aufgaben
 - Zwei Bilder: gleiche Person?
 - Klassifiziere Bilder nach abgebildeter Person.
 - Finde Gesicht in einer Datenbank.

