Maschinelles Lernen

Grundlegende Begriffe

- Überwachtes Lernen
  - Regression
    - Z.B. Schätzen des Benzinverbrauchs eines Autos
  - Klassifikation
    - Z.B. Handschriftenerkennung
- Unüberwachtes Lernen
  - Clustering
    - Z.B. Erkennung eines Kreditkartenbetrugs
- Bestärkendes Lernen („reinforcement learning“)
    - Z.B. Anwendung in der Robotik

Regression

- Gegeben: Features X_i und zugehörige Werte Y_i
- Gesucht: Abbildung $X \rightarrow Y$
- Beispiel: Wie lange brauche ich für meinen Weg zur Arbeit und zum Sportverein?

Klassifikation

- Gegeben: Features X_i und zugehörige Attribute Y_i
- Gesucht: Abbildung $X \rightarrow Y$
- Beispiel: Nach einem Blick auf die Wetterstation, regnet es heute?

Regression: Einfachster Fall

- Keine Features, nur Werte A_i
- Gesucht: eine Schätzung $\hat{A}$ für A
- Beispiele:
  - Anzahl der Regentage im Monat Mai
  - Fahrzeit zur Arbeit in Sekunden
  - Körpergrösse eines erwachsenen Europäers in cm
- Eigentlich gesucht: **Verteilung**

Normalverteilung

- Modell: A ~ **N(μ,σ²)**

$$p(A=a)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(a-\mu)^2}{2\sigma^2}}$$

Warum Normalverteilung?

Zentraler Grenzwertsatz: Gegeben unabhängige und identisch verteilte Zufallsvariablen $X_1, X_2, X_3, \ldots$ Wenn Erwartungswert μ und Varianz σ^2 der Verteilung existieren, dann konvergiert

$$\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}}$$

für $n \rightarrow \infty$ gegen die Standard-Normalverteilung N(0,1)

Rechnen mit Zufallsvariablen

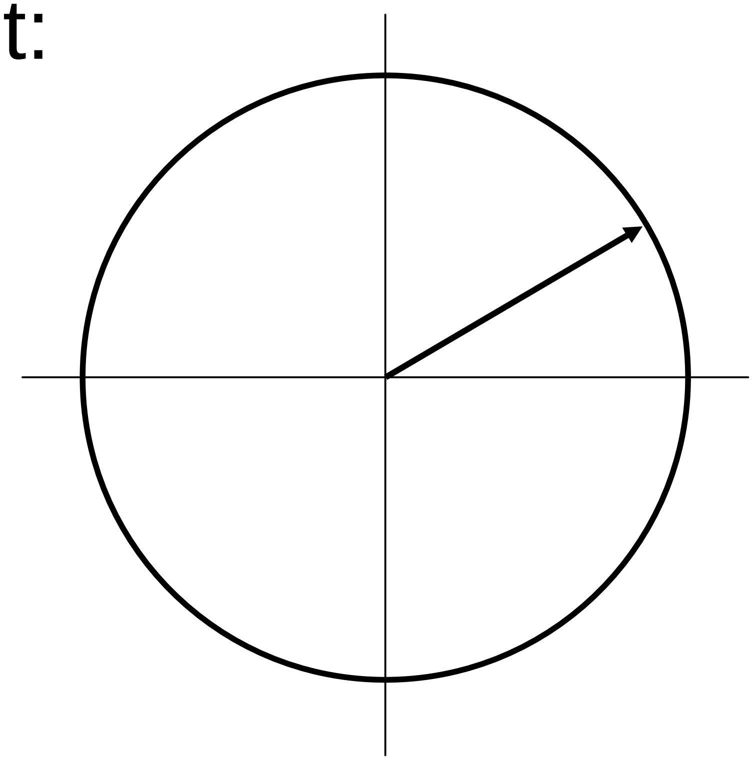
- Summen
- Erwartungswert
- Varianz
- Diskrete und kontinuierliche Zufallsvariablen

- **Übung: ZGWS simulieren**
- Wer kein `randn` hat: Box-Muller Methode:

$$z = \sqrt{-2\ln u_1}\cos(2\pi u_2)$$

Herleitung Box-Muller

- X_1 und X_2 seien N(0,1) verteilt
- $X_1^2+X_2^2$ ist χ^2_2 – verteilt:
  $f(x) = ½\, e^{-½x}$
- $X_1^2+X_2^2$ kann durch
  $-2 \ln u_1$ erzeugt
  werden
- Die zweidimensionale Normalverteilung ist rotationsinvariant



Der ZGWS in der Simulation

```
for i=1:100
  A = rand(i,10000)-.5;
  m(i) = mean((sum(A,1)/sqrt(i)));
  y(i) = mean((sum(A,1)/sqrt(i)).^2);
end
plot(m);
plot(y);
hist(sum(A,1)/sqrt(i),100);
```

Starkes Gesetz der Grossen Zahlen

Gegeben unabhängige und identisch verteilte Zufallsvariablen X_1, X_2, X_3, … Wenn Erwartungswert μ und Varianz σ^2 der Verteilung existieren, dann konvergiert

$$\frac{X_1 + X_2 + \ldots + X_n}{n}$$

für $n \rightarrow \infty$ *fast sicher* gegen den Mittelwert μ

Übung

- Starkes Gesetz der grossen Zahlen simulieren
- Was sind die Unterschiede zum ZGWS?

Schätzung einer Verteilung

- Keine Features, nur Werte A_i
- Gesucht: eine Schätzung $\hat{A}$ für A
- Modell: $A \sim N(\mu,\sigma^2)$
- $p(A_{i:i=1\ldots n}) \sim \exp\{ - [\Sigma(A_i-\mu)^2] / (2\sigma^2) \}$
- Welches Paar (μ,σ^2) ist am besten?
- Für die Daten $A_{i:i=1\ldots n}$, welches Paar (μ,σ^2) ist am wahrscheinlichsten?
- Umgekehrt: für welches Paar (μ,σ^2) sind die Daten $A_{i:i=1\ldots n}$ am wahrscheinlichsten?

Maximum-Likelihood Schätzer

- $\exp\{ - [\Sigma(A_i-\mu)^2] / (2\sigma^2) \} \sim p(A_{i:i=1\ldots n}) = \max!$

$\Leftrightarrow [\Sigma(A_i-\mu)^2] / (2\sigma^2) \sim -\log p(A_{i:i=1\ldots n}) = \min!$

$\Leftrightarrow \Sigma(A_i-\mu)^2 = \min!$

$\Leftrightarrow d/d\mu\ \Sigma(A_i-\mu)^2 = 0$

$\Leftrightarrow \Sigma(A_i-\mu) = 0$

$\Leftrightarrow \mu = (\Sigma A_i) / n =$ Mittelwert

Bemerkung: σ^2 nicht länger relevant!

ML Schätzer

- Ein Schätzer Â ist ein ML Schätzer für eine Variable A unter Beobachtungen
  X_1, X_2, X_3, … , falls

$$p(X_1, X_2, X_3, \ldots \mid \hat{A}) = \max_A p(X_1, X_2, X_3, \ldots \mid A)$$

Erwartungstreuer Schätzer

- Ein Schätzer Â ist ein erwartungstreuer Schätzer (**unbiased**) für einen Parameter A unter Beobachtungen X_1, X_2, X_3, … , falls

$$\mathbf{E}\left(\hat{A}(X_1, X_2, X_3, \ldots)\right) = A$$

Der Mittelwertschätzer

- Behauptung: Der Mittelwertschätzer $\mu = (\Sigma A_i) / n$ ist erwartungstreu.

- Beweis:

$$\begin{aligned}\mathbf{E}\,[(\Sigma A_i) / n\,] &= (\Sigma\, \mathbf{E}A_i) / n \\ &= (\Sigma\, \mathbf{E}A) / n \\ &= \mathbf{E}A \\ &= \mu.\end{aligned}$$

Übung: Taxiproblem

In einer Stadt gibt es N Taxis, die mit fortlaufenden Nummern 1…N gekennzeichnet sind. Wir wollen N schätzen. Dafür setzen wir uns an eine Strassenkreuzung und notieren die Nummern der vorbeifahrenden Taxis.

Am Abend haben wir folgende Nummer notiert:

5, 7, 7, 7, 14, 84, 99, 125, 125, 126, 230, 281, 399, 412

Wie können wir N schätzen?

Ist die Schätzung ML?

Ist die Schätzung erwartungstreu?

Taxiproblem: ML Schätzer

- $\hat{N} = \max(n_i)$
- Das ist ML, denn
- Für alle $N<\max(n_i)$ gilt $P(n_{i:i=1\ldots m}|N)=0$
- Für alle $N>\max(n_i)$ gilt

  $$\begin{aligned} P(n_{i:i=1\ldots m}|N) &= 1-[1-N^{-1}]^m \\ &< 1-[1-\max(n_i)^{-1}]^m \\ &= P(n_{i:i=1\ldots m}|\hat{N}) \end{aligned}$$
- Dieser Schätzer ist nicht besonders gut, da er N systematisch unterschätzt!

Taxiproblem: Erwartungstreuer Schätzer

- $\hat{N} = \max(n_i) + \min(n_i) - 1$
- Erwartungstreu, denn

$$
\begin{aligned}
\mathbf{E}\hat{N} &= \mathbf{E}[\max(n_i) + \min(n_i) - 1] \\
&= \mathbf{E}[N - (N - \max(n_i)) + \min(n_i) - 1] \\
&= N - \mathbf{E}(N - \max(n_i)) + \mathbf{E}(\min(n_i) - 1) \\
&= N
\end{aligned}
$$

Übung

- Wir simulieren das Taxiproblem
  - Wie verhält sich der ML Schätzer?
  - Wie verhält sich der erwartungstreue Schätzer?
  - Können wir noch einen Schätzer mit besseren Eigenschaften finden?

Schätzung der Varianz einer Normalverteilung

- X_i sei $N(\mu,\sigma^2)$
- Fall 1: μ bekannt
  - Dann ist $s^2 = [\Sigma (X_i - \mu)^2]/n$ ein ML Schätzer
  - s^2 ist auch erwartungstreu
  - Beweis: Übung / Hausaufgabe
- Fall 2: μ unbekannt, setze $m = [\Sigma X_i]/n$
  - $s^2 = [\Sigma (X_i - m)^2]/n$ ist ein ML Schätzer
  - s^2 ist nicht erwartungstreu
  - $s_{-1}{}^2 = [\Sigma (X_i - m)^2]/(n-1)$ ist erwartungstreu

s^2 ist ML

- $p(X_{1:n}|\mu,\sigma^2) \sim \sigma^{-n} \exp[-\Sigma(X_i-\mu)^2/(2\sigma^2)]$
- m ist ML für μ unabhängig von σ^2
- $2 \log p(X_{1:n}|\sigma^2) \sim -n \log \sigma - \Sigma(X_i-\mu)^2/(2\sigma^2)$
- Maximum: Ableitung nach σ muss 0 sein:
- $\sigma^{-1} = \Sigma(X_i-\mu)^2/(n\sigma^3)$
- $\sigma^2 = s^2$

s_{-1}^2 ist erwartungstreu

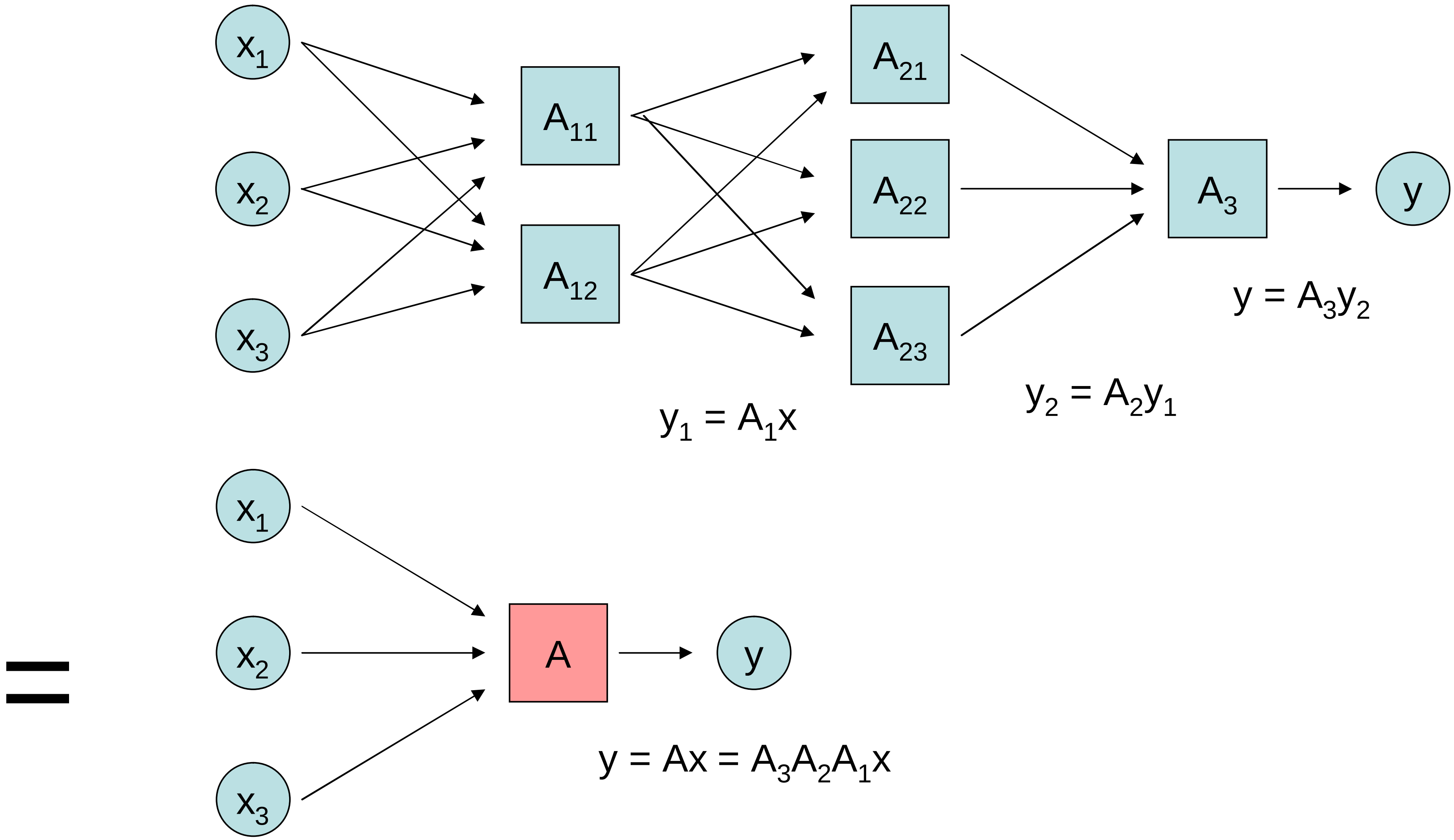
- Z.Z.: $\sigma^2 = \mathbf{E}\Sigma(X_i-m)^2/(n-1)$
- Können annehmen: $\mathbf{E}X_i = 0$

$$
\begin{aligned}
\Leftrightarrow (n-1)\sigma^2 &= \Sigma\, \mathbf{E}\,(X_i-m)^2 \\
&= \Sigma\, \mathbf{E}\,[(n-1)/n\ X_i]^2 + (n-1)\,\mathbf{E}\,[X_i/n]^2 \\
&= (n-1)^2/n\ \sigma^2 + (n-1)/n\ \sigma^2 \\
&= (n-1)\sigma^2 \quad \blacklozenge
\end{aligned}
$$

Lineare Regression

- Modell: $Y = bX$
- ML-Schätzer für b:
- $(\mathbf{Y}-b\mathbf{X})^T(\mathbf{Y}-b\mathbf{X}) = \min!$
- Ableitung nach b (Gradient):
- $0 = 2\mathbf{X}^T\mathbf{X}b-2\mathbf{X}^T\mathbf{Y}$
- $\underline{b = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\mathbf{Y}}$

Lineares Neuronales Netz



Übung

- Gegeben sind folgende Trainingsdaten
- Aufgabe: trainiere ein entsprechendes Neuronales Netz

X_1	X_2	Y
0.9	3.8	-1.1
1.9	2.4	1.8
3.1	3.6	3.6
0	1.8	-0.7
1.8	3.2	0.9
3.2	2.5	5.3
3.0	3.7	3.7

Lösung

- X \ Y = [2.1176; -0.7706]
- [X ones(7,1)] \ Y = [2.1137; -1.0376; 0.8591]

- Wahre Koeffizienten = [2; -1; 1]

Übung

Schätzung der Anzahl Fische in einem Teich

In einem Teich befinden sich N (unbekannt) Fische. Wir schätzen sie wie folgt: Wir fangen M Fische, markieren sie, und lassen sie wieder frei. Dann warten wir etwas und fangen dann n Fische und zählen, wie viele davon markiert sind (m). Jetzt schätzen wir $\hat{N} = M*n / m$.

- Zeigen Sie: der Schätzer ist ML
- Der Schätzer ist nicht erwartungstreu.

Übung

- Gegeben sind folgende Trainingsdaten
- Aufgabe: trainiere ein quadratisches Regressionsmodell (Ansatz: $y = ax^2+bx+c$)
- Ist das eine lineare Regression? Warum?

X	Y
0	0
1	0
2	1
3	2
4	5
5	9