

# Data Mining Tutorial

## Klassifikation II

Erich Schubert, Arthur Zimek

Ludwig-Maximilians-Universität München

2014-06-06 — KDD Übung

<b>Wetter</b>	<b>Schnee</b>	<b>Ski?</b>
Sonne	$< 50$	nein
Regen	$< 50$	nein
Regen	$\geq 50$	nein
Schnee	$\geq 50$	ja

<b>Wetter</b>	<b>Schnee</b>	<b>Ski?</b>
Schnee	$< 50$	nein
Sonne	$\geq 50$	ja
Schnee	$\geq 50$	ja
Regen	$< 50$	ja

A priori- und bedingte Wahrscheinlichkeiten:

<b>Wetter</b>	<b>Schnee</b>	<b>Ski?</b>	<b>Wetter</b>	<b>Schnee</b>	<b>Ski?</b>
Sonne	< 50	nein	Schnee	< 50	nein
Regen	< 50	nein	Sonne	≥ 50	ja
Regen	≥ 50	nein	Schnee	≥ 50	ja
Schnee	≥ 50	ja	Regen	< 50	ja

A priori- und bedingte Wahrscheinlichkeiten:

$$P(Ski) = 1/2$$

$$P(\neg Ski) = 1/2$$

Wetter	Schnee	Ski?	Wetter	Schnee	Ski?
Sonne	< 50	nein	Schnee	< 50	nein
Regen	< 50	nein	Sonne	≥ 50	ja
Regen	≥ 50	nein	Schnee	≥ 50	ja
Schnee	≥ 50	ja	Regen	< 50	ja

A priori- und bedingte Wahrscheinlichkeiten:

$$P(\textit{Wetter} = \textit{Sonne} | \textit{Ski}) = 1/4$$

$$P(\textit{Wetter} = \textit{Schnee} | \textit{Ski}) = 2/4$$

$$P(\textit{Wetter} = \textit{Regen} | \textit{Ski}) = 1/4$$

$$P(\textit{Wetter} = \textit{Sonne} | \neg \textit{Ski}) = 1/4$$

$$P(\textit{Wetter} = \textit{Schnee} | \neg \textit{Ski}) = 1/4$$

$$P(\textit{Wetter} = \textit{Regen} | \neg \textit{Ski}) = 2/4$$

Wetter	Schnee	Ski?	Wetter	Schnee	Ski?
Sonne	< 50	nein	Schnee	< 50	nein
Regen	< 50	nein	Sonne	≥ 50	ja
Regen	≥ 50	nein	Schnee	≥ 50	ja
Schnee	≥ 50	ja	Regen	< 50	ja

A priori- und bedingte Wahrscheinlichkeiten:

$$P(\text{Schnee} \geq 50 | \text{Ski}) = 3/4$$

$$P(\text{Schnee} < 50 | \text{Ski}) = 1/4$$

$$P(\text{Schnee} \geq 50 | \neg \text{Ski}) = 1/4$$

$$P(\text{Schnee} < 50 | \neg \text{Ski}) = 3/4$$

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	$1/2$	$1/4$	$2/4$	$1/4$	$3/4$	$1/4$
$\neg$ Ski	$1/2$	$1/4$	$1/4$	$2/4$	$1/4$	$3/4$

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	$1/2$	$1/4$	$2/4$	$1/4$	$3/4$	$1/4$
$\neg$ Ski	$1/2$	$1/4$	$1/4$	$2/4$	$1/4$	$3/4$

**A) Wetter=Sonne, Schnee  $\geq 50$**

$$\begin{aligned}
 & P(\text{Ski} | \text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50) \\
 = & \frac{P(\text{Wetter} = \text{Sonne} | \text{Ski}) \cdot P(\text{Schnee} \geq 50 | \text{Ski}) \cdot P(\text{Ski})}{P(\text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50)} \\
 = & \frac{\frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{2}}{P(\text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50)} \\
 = & \frac{\frac{3}{32}}{P(\text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50)}
 \end{aligned}$$

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	$1/2$	$1/4$	$2/4$	$1/4$	$3/4$	$1/4$
$\neg$ Ski	$1/2$	$1/4$	$1/4$	$2/4$	$1/4$	$3/4$

**A)** Wetter=Sonne, Schnee  $\geq 50$

$$\begin{aligned}
 & P(\neg \text{Ski} | \text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50) \\
 = & \frac{P(\text{Wetter} = \text{Sonne} | \neg \text{Ski}) \cdot P(\text{Schnee} \geq 50 | \neg \text{Ski}) \cdot P(\neg \text{Ski})}{P(\text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50)} \\
 = & \frac{\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{2}}{P(\text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50)} \\
 = & \frac{\frac{1}{32}}{P(\text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50)}
 \end{aligned}$$



	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	1/2	1/4	2/4	1/4	3/4	1/4
$\neg$ Ski	1/2	1/4	1/4	2/4	1/4	3/4

**A)** Wetter=Sonne, Schnee  $\geq 50$

$$P(\text{Ski} | \text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50) = \frac{\frac{3}{32}}{P(\dots)}$$

$$P(\neg \text{Ski} | \text{Wetter} = \text{Sonne}, \text{Schnee} \geq 50) = \frac{\frac{1}{32}}{P(\dots)}$$

$\Rightarrow$  Skifahren

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	$1/2$	$1/4$	$2/4$	$1/4$	$3/4$	$1/4$
$\neg$ Ski	$1/2$	$1/4$	$1/4$	$2/4$	$1/4$	$3/4$

**B) Wetter=Regen, Schnee  $< 50$**

$$\begin{aligned}
 & P(\text{Ski} | \text{Wetter} = \text{Regen}, \text{Schnee} < 50) \\
 &= \frac{P(\text{Wetter} = \text{Regen} | \text{Ski}) \cdot P(\text{Schnee} < 50 | \text{Ski}) \cdot P(\text{Ski})}{P(\text{Wetter} = \text{Regen}, \text{Schnee} < 50)} \\
 &= \frac{\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{2}}{P(\text{Wetter} = \text{Regen}, \text{Schnee} < 50)} \\
 &= \frac{\frac{1}{32}}{P(\text{Wetter} = \text{Regen}, \text{Schnee} < 50)}
 \end{aligned}$$

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	$1/2$	$1/4$	$2/4$	$1/4$	$3/4$	$1/4$
$\neg$ Ski	$1/2$	$1/4$	$1/4$	$2/4$	$1/4$	$3/4$

**B) Wetter=Regen, Schnee  $< 50$**

$$\begin{aligned}
 & P(\neg \text{Ski} | \text{Wetter} = \text{Regen}, \text{Schnee} < 50) \\
 &= \frac{P(\text{Wetter} = \text{Regen} | \neg \text{Ski}) \cdot P(\text{Schnee} < 50 | \neg \text{Ski}) \cdot P(\neg \text{Ski})}{P(\text{Wetter} = \text{Regen}, \text{Schnee} < 50)} \\
 &= \frac{\frac{2}{4} \cdot \frac{3}{4} \cdot \frac{1}{2}}{P(\text{Wetter} = \text{Regen}, \text{Schnee} < 50)} \\
 &= \frac{\frac{6}{32}}{P(\text{Wetter} = \text{Regen}, \text{Schnee} < 50)}
 \end{aligned}$$

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	$1/2$	$1/4$	$2/4$	$1/4$	$3/4$	$1/4$
$\neg$ Ski	$1/2$	$1/4$	$1/4$	$2/4$	$1/4$	$3/4$

**B) Wetter=Regen, Schnee  $< 50$**

$$P(\text{Ski} | \text{Wetter} = \text{Regen}, \text{Schnee} < 50) = \frac{1}{32}$$

$$P(\neg \text{Ski} | \text{Wetter} = \text{Regen}, \text{Schnee} < 50) = \frac{6}{32}$$

$\Rightarrow$  nicht Skifahren

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	$1/2$	$1/4$	$2/4$	$1/4$	$3/4$	$1/4$
$\neg$ Ski	$1/2$	$1/4$	$1/4$	$2/4$	$1/4$	$3/4$

C) Wetter=Schnee, Schnee  $< 50$

$$\begin{aligned}
 & P(\text{Ski} | \text{Wetter} = \text{Schnee}, \text{Schnee} < 50) \\
 &= \frac{P(\text{Wetter} = \text{Schnee} | \text{Ski}) \cdot P(\text{Schnee} < 50 | \text{Ski}) \cdot P(\text{Ski})}{P(\text{Wetter} = \text{Schnee}, \text{Schnee} < 50)} \\
 &= \frac{\frac{2}{4} \cdot \frac{1}{4} \cdot \frac{1}{2}}{P(\text{Wetter} = \text{Schnee}, \text{Schnee} < 50)} \\
 &= \frac{\frac{2}{32}}{P(\text{Wetter} = \text{Schnee}, \text{Schnee} < 50)}
 \end{aligned}$$

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	$1/2$	$1/4$	$2/4$	$1/4$	$3/4$	$1/4$
$\neg$ Ski	$1/2$	$1/4$	$1/4$	$2/4$	$1/4$	$3/4$

**C)** Wetter=Schnee, Schnee  $< 50$

$$\begin{aligned}
 & P(\neg \text{Ski} | \text{Wetter} = \text{Schnee}, \text{Schnee} < 50) \\
 &= \frac{P(\text{Wetter} = \text{Schnee} | \neg \text{Ski}) \cdot P(\text{Schnee} < 50 | \neg \text{Ski}) \cdot P(\neg \text{Ski})}{P(\text{Wetter} = \text{Schnee}, \text{Schnee} < 50)} \\
 &= \frac{\frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{2}}{P(\text{Wetter} = \text{Schnee}, \text{Schnee} < 50)} \\
 &= \frac{\frac{3}{32}}{P(\text{Wetter} = \text{Schnee}, \text{Schnee} < 50)}
 \end{aligned}$$

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	$\geq 50$	$< 50$
Ski	$1/2$	$1/4$	$2/4$	$1/4$	$3/4$	$1/4$
$\neg$ Ski	$1/2$	$1/4$	$1/4$	$2/4$	$1/4$	$3/4$

**C) Wetter=Schnee, Schnee  $< 50$**

$$P(\text{Ski} | \text{Wetter} = \text{Schnee}, \text{Schnee} < 50) = \frac{\frac{2}{32}}{P(\dots)}$$

$$P(\neg \text{Ski} | \text{Wetter} = \text{Schnee}, \text{Schnee} < 50) = \frac{\frac{3}{32}}{P(\dots)}$$

$\Rightarrow$  nicht Skifahren









