Data Mining Tutorial

Klassifikation II

Erich Schubert, Arthur Zimek

Ludwig-Maximilians-Universität München

2013-06-21 — KDD Übung

Naïve Bayes

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

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

A priori- und bedingte Wahrscheinlichkeiten:

Naïve Bayes

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

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

A priori- und bedingte Wahrscheinlichkeiten:

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

Naïve Bayes

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

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

A priori- und bedingte Wahrscheinlichkeiten:

$$
\begin{aligned}
P(Wetter = Sonne|Ski) &= 1/4 \\
P(Wetter = Schnee|Ski) &= 2/4 \\
P(Wetter = Regen|Ski) &= 1/4 \\
P(Wetter = Sonne|\neg Ski) &= 1/4 \\
P(Wetter = Schnee|\neg Ski) &= 1/4 \\
P(Wetter = Regen|\neg Ski) &= 2/4
\end{aligned}
$$

Naïve Bayes

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

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

A priori- und bedingte Wahrscheinlichkeiten:

$$P(Schnee \geq 50 | Ski) = 3/4$$
$$P(Schnee < 50 | Ski) = 1/4$$
$$P(Schnee \geq 50 | \neg Ski) = 1/4$$
$$P(Schnee < 50 | \neg Ski) = 3/4$$

Naive Bayes

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

Naive Bayes

		Wetter			Schnee	
	a priori	Sonne	Schnee	Regen	≥ 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 ≥ 50

$$P(Ski|Wetter = Sonne, Schnee \geq 50)$$

$$= \frac{P(Wetter = Sonne|Ski) \cdot P(Schnee \geq 50|Ski) \cdot P(Ski)}{P(Wetter = Sonne, Schnee \geq 50)}$$

$$= \frac{\frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{2}}{P(Wetter = Sonne, Schnee \geq 50)}$$

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

Naive Bayes

		Wetter			Schnee	
	a priori	Sonne	Schnee	Regen	≥ 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 ≥ 50

$$P(\neg Ski | Wetter = Sonne, Schnee \geq 50)$$

$$= \frac{P(Wetter = Sonne | \neg Ski) \cdot P(Schnee \geq 50 | \neg Ski) \cdot P(\neg Ski)}{P(Wetter = Sonne, Schnee \geq 50)}$$

$$= \frac{\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{2}}{P(Wetter = Sonne, Schnee \geq 50)}$$

$$= \frac{\frac{1}{32}}{P(Wetter = Sonne, Schnee \geq 50)}$$

Naive Bayes

		Wetter			Schnee	
	a priori	Sonne	Schnee	Regen	≥ 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 ≥ 50

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

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

$\Rightarrow$ Skifahren

Naive Bayes

		Wetter			Schnee	
	a priori	Sonne	Schnee	Regen	≥ 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

$$P(Ski|Wetter = Regen, Schnee < 50)$$

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

$$= \frac{\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{2}}{P(Wetter = Regen, Schnee < 50)}$$

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

Naive Bayes

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	≥ 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

$$P(\neg Ski | Wetter = Regen, Schnee < 50)$$

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

$$= \frac{\frac{2}{4} \cdot \frac{3}{4} \cdot \frac{1}{2}}{P(Wetter = Regen, Schnee < 50)}$$

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

Naive Bayes

		Wetter			Schnee	
	a priori	Sonne	Schnee	Regen	≥ 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

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

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

$\Rightarrow$ nicht Skifahren

Naive Bayes

		Wetter			Schnee	
	a priori	Sonne	Schnee	Regen	≥ 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

$$P(Ski|Wetter = Schnee, Schnee < 50)$$

$$= \frac{P(Wetter = Schnee|Ski) \cdot P(Schnee < 50|Ski) \cdot P(Ski)}{P(Wetter = Schnee, Schnee < 50)}$$

$$= \frac{\frac{2}{4} \cdot \frac{1}{4} \cdot \frac{1}{2}}{P(Wetter = Schnee, Schnee < 50)}$$

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

Naive Bayes

	a priori	Wetter			Schnee	
		Sonne	Schnee	Regen	≥ 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

$$P(\neg Ski | Wetter = Schnee, Schnee < 50)$$

$$= \frac{P(Wetter = Schnee | \neg Ski) \cdot P(Schnee < 50 | \neg Ski) \cdot P(\neg Ski)}{P(Wetter = Schnee, Schnee < 50)}$$

$$= \frac{\frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{2}}{P(Wetter = Schnee, Schnee < 50)}$$

$$= \frac{\frac{3}{32}}{P(Wetter = Schnee, Schnee < 50)}$$

Naive Bayes

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

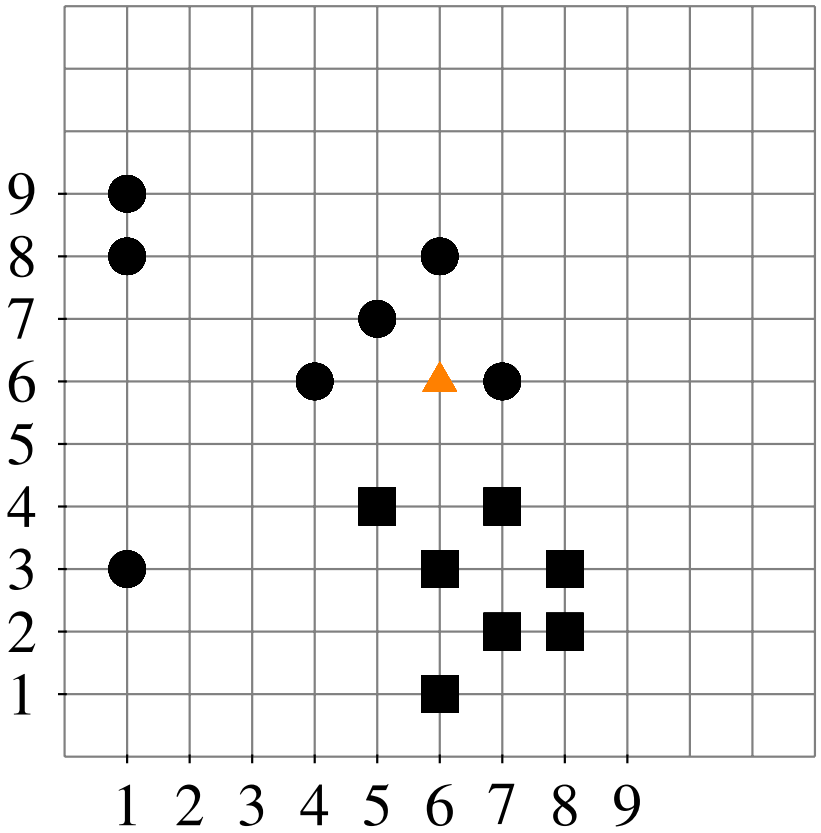
C) Wetter=Schnee, Schnee < 50

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

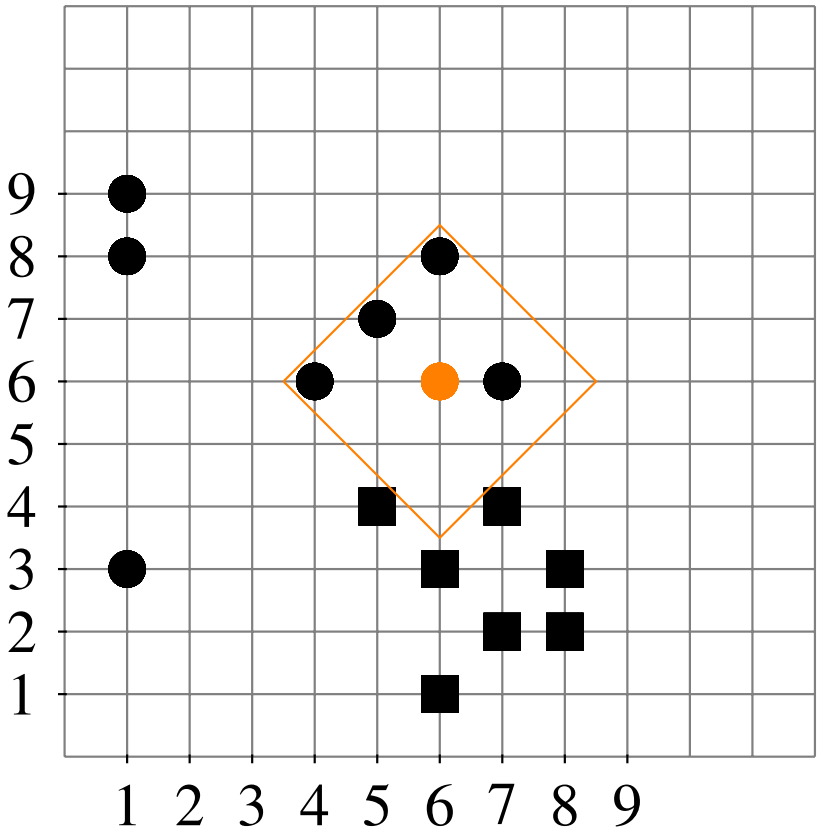
$$P(\neg Ski|Wetter = Schnee, Schnee < 50) = \frac{\frac{3}{32}}{P(\ldots)}$$

$\Rightarrow$ nicht Skifahren

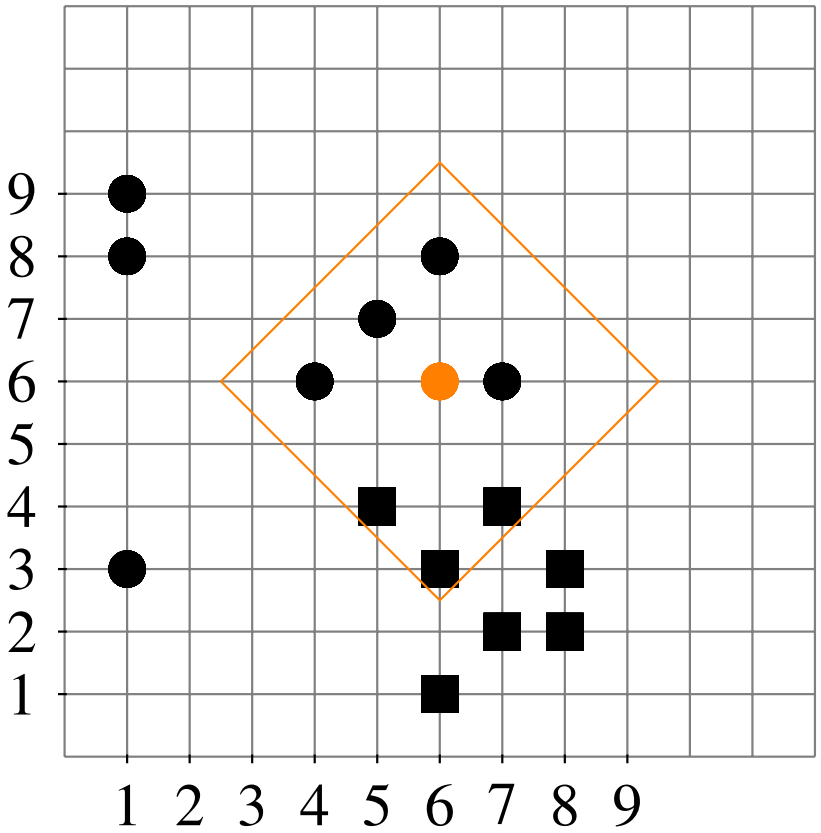
Nächste-Nachbarn Klassifikation



Nächste-Nachbarn Klassifikation



Nächste-Nachbarn Klassifikation



Nächste-Nachbarn Klassifikation