Pagrindiniai DT uždaviniai ir sąvokos

Duomenys apie orą

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Duomenų rinkinys (imtis), sudarytas iš 14 įrašų.

	Oras	*Temperatūra*	*Drėgnumas*	*Vėjuota*	*Žaisti*
1	saulėta	karšta	didelis	FALSE	ne
2	saulėta	karšta	didelis	TRUE	ne
3	debesuota	karšta	didelis	FALSE	taip
4	lietinga	šilta	didelis	FALSE	taip
5	lietinga	vėsu	normalus	FALSE	taip
6	lietinga	vėsu	normalus	TRUE	ne
7	debesuota	vėsu	normalus	TRUE	taip
8	saulėta	šilta	didelis	FALSE	ne
9	saulėta	vėsu	normalus	FALSE	taip
10	lietinga	šilta	normalus	FALSE	taip
11	saulėta	šilta	normalus	TRUE	taip
12	debesuota	šilta	didelis	TRUE	taip
13	debesuota	karšta	normalus	FALSE	taip
14	lietinga	šilta	didelis	TRUE	ne

Klasifikacijos taisyklės pagal oro duomenis

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Dabar pagal turimus duomenis pabandykime sukonstruoti taisykles, kurios leistų nuspėti "Žvaigždžių" elgesį, esant bet kokioms oro sąlygoms. Kitaip sakant, reikia *"išmokti"* apskaičiuoti kintamojo *Žaisti* reikšmę, priklausomai nuo likusiųjų kintamųjų reikšmių. Taisyklės galėtų būti tokios

> `Jei` *Oras=saulėta* `ir` *Drėgnumas=didelis* `tai` *Žaisti=ne*
> `Jei` *Oras=lietinga* `ir` *Vėjuota=TRUE* `tai` *Žaisti=ne*
> `Jei` *Oras=debesuota* `tai` *Žaisti=taip*
> `Jei` *Drėgnumas=normalus* `tai` *Žaisti=taip*
> `Kitais atvejais` *Žaisti=taip*

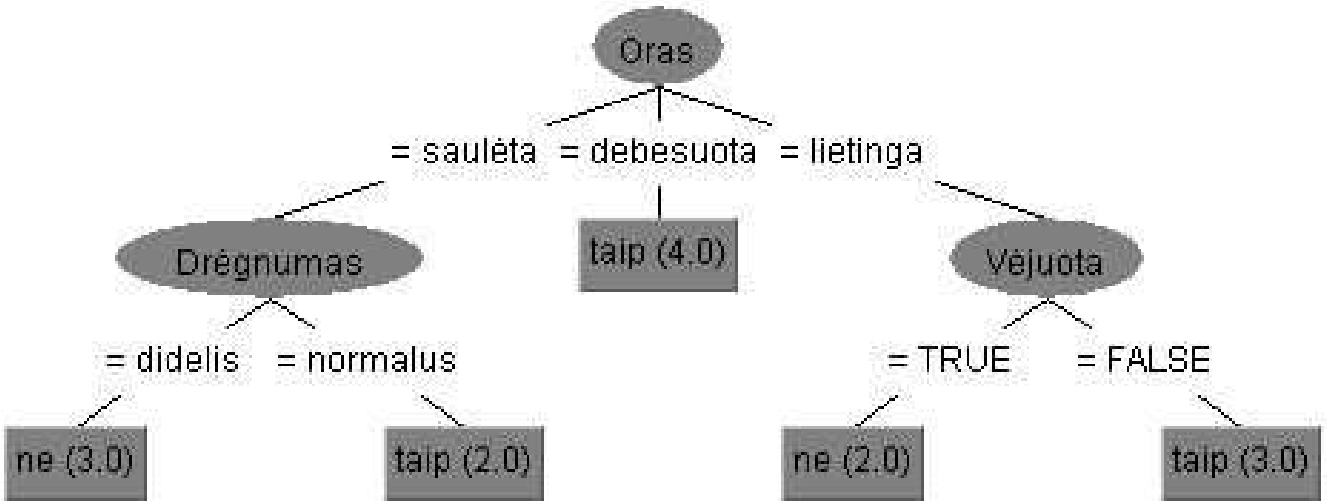
Atskirai paimta tasyklė

> `Jei` *Drėgnumas=normalus* `tai` *Žaisti=taip*

ne visada bus teisinga.

Sprendimų medis duomenims apie orą

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas



Asociacijos taisyklės pagal oro duomenis

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Asociaciacijos taisyklės nusako galimus sąryšius tarp įvairių kintamųjų. Pavyzdžiui

`Jei` *Temperatūra=vėsu* `tai` *Drėgnumas=normalus*
`Jei` *Oras=saulėta* `ir` *Žaisti=ne* `tai` *Drėgnumas=didelis*
`Jei` *Vėjuota=FALSE* `ir` *Žaisti=ne* `tai` *Oras=saulėta* `ir` *Drėgnumas=didelis*

Patikslinti duomenys apie orą

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Kintamuosius *(Oras, Temperatūra, Drėgnumas, Vėjuota, Žaisti)* pažymėkime (X_1, X_2, X_3, X_4, Y) ir išmatuokime oro temperatūrą bei drėgnumą.

	X_1	X_2	X_3	X_4	Y
1	saulėta	29	85	FALSE	ne
2	saulėta	27	90	TRUE	ne
3	debesuota	28	86	FALSE	taip
4	lietinga	21	96	FALSE	taip
5	lietinga	20	80	FALSE	taip
6	lietinga	18	70	TRUE	ne
7	debesuota	18	65	TRUE	taip
8	saulėta	22	95	FALSE	ne
9	saulėta	21	70	FALSE	taip
10	lietinga	24	80	FALSE	taip
11	saulėta	24	70	TRUE	taip
12	debesuota	22	90	TRUE	taip
13	debesuota	27	75	FALSE	taip
14	lietinga	22	91	TRUE	ne

Patikslinti duomenys apie orą

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Dabar kintamieji X_2 ir X_3 jau bus kiekybiniai (arba skaitiniai) , nes jų reikšmės nusako temperatūrą laipsniais Celsijaus skalėje ir drėgnumą procentais. Kokybinio kintamojo Y reikšmių aibė yra $\{taip,\ ne\}$. Taisyklės, pagal kurias klasifikuojami įrašai kintamojo Y atžvilgiu, galėtų būti tokios

$$\texttt{Jei } X_1 = \textit{saulėta} \texttt{ ir } X_3 > 84 \texttt{ tai } Y = ne$$
$$\texttt{Jei } X_1 = \textit{lietinga} \texttt{ ir } X_4 = \textit{TRUE} \texttt{ tai } Y = \textit{ne}$$
$$\texttt{Kitais atvejais } \ Y = \textit{taip}$$

Regos korekcija

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

	Amžius	*Regėjimas*	*Astigmatizmas*	*Ašarų kiekis*	*Lęšiai*
1	jaunas	trumparegis	ne	sumažėjęs	neskirti
2	jaunas	trumparegis	ne	normalus	minkšti
3	jaunas	trumparegis	taip	sumažėjęs	neskirti
4	jaunas	trumparegis	taip	normalus	kieti
5	jaunas	toliaregis	ne	sumažėjęs	neskirti
6	jaunas	toliaregis	ne	normalus	minkšti
7	jaunas	toliaregis	taip	sumažėjęs	neskirti
8	jaunas	toliaregis	taip	normalus	kieti
9	vidutinis	trumparegis	ne	sumažėjęs	neskirti
10	vidutinis	trumparegis	ne	normalus	minkšti
11	vidutinis	trumparegis	taip	sumažėjęs	neskirti
12	vidutinis	trumparegis	taip	normalus	kieti
13	vidutinis	toliaregis	ne	sumažėjęs	neskirti
14	vidutinis	toliaregis	ne	normalus	minkšti
15	vidutinis	toliaregis	taip	sumažėjęs	neskirti
16	vidutinis	toliaregis	taip	normalus	neskirti
17	vyresnis	trumparegis	ne	sumažėjęs	neskirti
18	vyresnis	trumparegis	ne	normalus	neskirti
19	vyresnis	trumparegis	taip	sumažėjęs	neskirti
20	vyresnis	trumparegis	taip	normalus	kieti
21	vyresnis	toliaregis	ne	sumažėjęs	neskirti
22	vyresnis	toliaregis	ne	normalus	minkšti
23	vyresnis	toliaregis	taip	sumažėjęs	neskirti
24	vyresnis	toliaregis	taip	normalus	neskirti

Lęšių skyrimo taisyklės

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

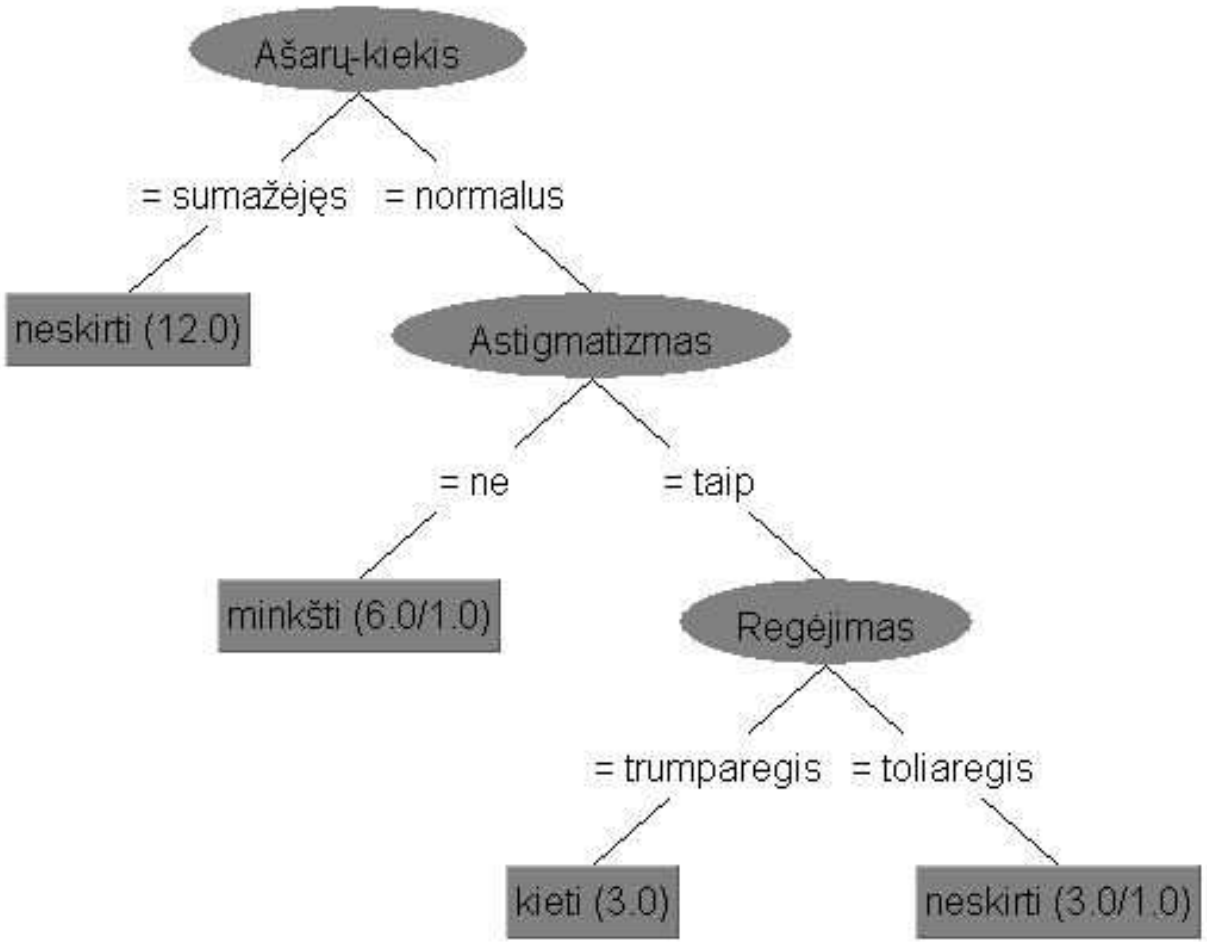
Iš viso yra 24 įrašai, nusakantys visas galimas šių parametrų reikšmių kombinacijas ($3 \times 2 \times 2 \times 2 = 24$). Tačiau jų pateikimo būdas vizualiai sunkiai aprėpiamas. O jeigu kintamųjų skaičius būtų didesnis ir įrašų turėtume ne 24, o tarkime 10000 ? Kitaip sakant, reikalingos kuo paprastesnės *taisyklės*, leidžiančios teisingai klasifikuoti visus ar bent jau didesnę dalį imties įrašų pagal kintamojo *Lešiai* reikšmes. Pavyzdžiui, tokia paprasta taisyklė

`Jei` *Ašarų kiekis=sumažėjęs* `tai` *Lęšiai=neskirti*

teisingai klasifikuoja 12 įrašų, bet likusieji lieka visai neklasifikuoti.

Sprendimų medis regos korekcijai

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas



Irisų duomenys

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Duomenys apie irisus (R.A.Fišeris, 1936)

	Taurėlapio ilgis (X_1)	Taurėlapio plotis (X_2)	Vainiklapio ilgis (X_3)	Vainiklapio plotis (X_4)	Iriso rūšis (Y)
1	5,1	3,5	1,4	0,2	Iris-setosa
2	4,9	3,0	1,4	0,2	Iris-setosa
3	4,7	3,2	1,3	0,2	Iris-setosa
4	4,6	3,1	1,5	0,2	Iris-setosa
5	5,0	3,6	1,4	0,2	Iris-setosa
...	...	...	...	...	...
50	5,0	3,3	1,4	0,2	Iris-setosa
51	7,0	3,2	4,7	1,4	Iris-versicolor
52	6,4	3,2	4,5	1,5	Iris-versicolor
53	6,9	3,1	4,9	1,5	Iris-versicolor
54	5,5	2,3	4,0	1,3	Iris-versicolor
55	6,5	2,8	4,6	1,5	Iris-versicolor
...	...	...	...	...	...
100	5,7	2,8	4,1	1,3	Iris-versicolor
101	6,3	3,3	6,0	2,5	Iris-virginica
102	5,8	2,7	5,1	1,9	Iris-virginica
103	7,1	3,0	5,9	2,1	Iris-virginica
104	6,3	2,9	5,6	1,8	Iris-virginica
105	6,5	3,0	5,8	2,2	Iris-virginica
...	...	...	...	...	...
150	5,9	3,0	5,1	1,8	Iris-virginica

Irisų klasifikacija

Išnagrinėjus visus 150 imties įrašų, galima būtų "išmokti" , pavyzdžiui, tokias irisų klasifikavimo taisykles:

$$
\begin{aligned}
&\texttt{Jei } X_3 < 2,45 \texttt{ tai } Y = Iris - setosa\\
&\texttt{Jei } X_2 < 2,10 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_2 < 2,45 \texttt{ ir } X_3 < 4,55 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_2 < 2,95 \texttt{ ir } X_4 < 1,35 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_3 > 2,45 \texttt{ ir } X_3 < 4,45 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_1 > 5,85 \texttt{ ir } X_3 < 4,75 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_2 < 2,55 \texttt{ ir } X_3 < 4,95 \texttt{ ir } X_4 < 1,55 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_3 > 2,45 \texttt{ ir } X_3 < 4,95 \texttt{ ir } X_4 < 1,55 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_1 > 6,55 \texttt{ ir } X_3 < 5,05 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_2 < 2,75 \texttt{ ir } X_4 < 1,65 \texttt{ ir } X_1 < 6,05 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_1 > 5,85 \texttt{ ir } X_1 < 5,95 \texttt{ ir } X_3 < 4,85 \texttt{ tai } Y = Iris - versicolor\\
&\texttt{Jei } X_3 > 5,15 \texttt{ tai } Y = Iris - virginica\\
&\texttt{Jei } X_4 > 1,85 \texttt{ tai } Y = Iris - virginica\\
&\texttt{Jei } X_4 > 1,75 \texttt{ ir } X_2 < 3,05 \texttt{ tai } Y = Iris - virginica\\
&\texttt{Jei } X_3 > 4,95 \texttt{ ir } X_4 < 1,55 \texttt{ tai } Y = Iris - virginica
\end{aligned}
$$

Rankraščio atpažinimas

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Ranka rašytas skaitmuo: nespalvotas 16 x 16 = 256 pikselių piešinys.

Taigi kiekvienas imties įrašas

$$(X_1, X_2, \ldots, X_{256}, Y)\,,$$

$X_i \in \{0, 1, \ldots, 255\}$ ir $Y \in \{0, 1, \ldots, 9\}$.

Kompiuterio našumas

Duomenys apie įvairias (hipotetines) kompiuterio konfigūracijas, siekiant išsiaiškinti kaip priklauso jo našumas, išreikštas sąlyginiais vienetais, nuo įvairių sistemos parametrų.

	Takto ilgis (ns)	Pagr.atm. min (Kb)	Pagr.atm. max(Kb)	Spart. atm.(Kb)	Kanalai min	Kanalai max	Našumas
	X_1	X_2	X_3	X_4	X_5	X_6	Y
1	125	256	6000	256	16	128	199
2	29	8000	32000	32	8	32	253
3	29	8000	32000	32	8	32	253
4	29	8000	32000	32	8	32	253
5	29	8000	16000	32	8	16	132
6	26	8000	32000	64	8	32	290
7	23	16000	32000	64	16	32	381
8	23	16000	32000	64	16	32	381
9	23	16000	64000	64	16	32	749
10	23	32000	64000	128	32	64	1238
...	...	...	...	...	...	...	...
207	125	2000	8000	0	2	14	41
208	480	512	8000	32	0	0	47
209	480	1000	4000	0	0	0	25

Kompiuterio našumo skaitinė prognozė

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Manydami, kad našumas (priklausomas kintamasis Y) tiesiškai priklauso nuo kintamųjų $X_1, X_2, X_3, X_4, X_5, X_6$, galėtume gauti, pavyzdžiui, tokį kintamojo Y įvertį:

$$\begin{aligned}\hat{Y} &= -66,481 + 0,066X_1 + 0,014X_2 + 0,0066X_3 \\ &\quad +0,494X_4 - 0,172X_5 + 1,20117X_6\,.\end{aligned}$$

Tai yra vadinamasis tiesinės regresijos modelis, o gautoji lygybė kartais dar vadinama tiesinės regresijos lygtimi.

Duomenys ir jų atributai

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Vieno populiacijos objekto stebėjimo (matavimo) rezultatas vadinamas įrašu. Visi įrašai sudaro imtį, o jų kiekis vadinamas imties dydžiu. Tuo atveju, kai matuojame tik vieną atributą (pvz., ūgį), įrašas turės vienintelę komponentę, o pats stebimasis dydis (ir pati imtis) vadinami vienmačiais. Jei mums rūpi kelios stebimojo objekto charakteristikos (pvz.,lytis, ūgis ir svoris), įrašą sudarys p komponenčių (paminėtuoju atveju p=3), o pats stebimasis dydis (ir imtis) vadinami p-mačiais. Pavyzdžiui, Fišerio irisų imtis yra penkiamatė ($p = 5$), o jos dydis lygus 150.
Komponentes sudaro įvairių tipų kintamieji (kitaip: atributai).

Kintamųjų tipai

Kintamojo tipas		Leistinos operacijos
Kategorinis	Vardinis	Įrašų, patekusių į kiekvieną kategoriją, skaičiaus radimas
	Ranginis	Įrašų, turinčių konkretų rangą, skaičiaus radimas. Rangų palyginimas (santykiai "daugiau", "mažiau")
Skaitinis	Intervalinis	Sudėtis, atimtis, daugyba, dalyba iš skaičiaus
	Santykinis	Visos matematinės operacijos

DT uždaviniai: kontroliuojamas mokymas

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Tarkime imties įrašai sudaryti iš nepriklausomo kintamojo X ir priklausomo kintamojo Y reikšmių: $(x_1, y_1), \ldots, (x_n, y_n)$. Beje kintamieji gali būti ir daugiamačiai. Tada žymėsime $X = (X_1, \ldots, X_k)$ ir $Y = (Y_1, \ldots, Y_m)$.

DT uždaviniai: kontroliuojamas mokymas

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Priklausomai nuo kintamojo Y tipo, skiriami tokie kontroliuojamo mokymo uždaviniai:

Klasifikavimas.

Kintamasis Y - kategorinis. Iš turimų imties duomenų reikia "išmokti" visoje populiacijoje nustatyti Y reikšmę pagal žinomą X reikšmę . Todėl iš tikrųjų yra konstruojamas kintamojo Y įvertis $\hat{Y} = f(X)$, stengiantis, kad kuo didesnei imties įrašų daliai jis būtų teisingas, t.y. $f(x_i) = y_i$. Kitaip sakant, turėdami įrašus (x_i, y_i) , mes galime *kontroliuoti* gautojo įverčio patikimumą.

DT uždaviniai: kontroliuojamas mokymas

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Priklausomai nuo kintamojo Y tipo, skiriami tokie kontroliuojamo mokymo uždaviniai:

Klasifikavimas.

Kintamasis Y - kategorinis. Iš turimų imties duomenų reikia "išmokti" visoje populiacijoje nustatyti Y reikšmę pagal žinomą X reikšmę . Todėl iš tikrųjų yra konstruojamas kintamojo Y įvertis $\hat{Y} = f(X)$, stengiantis, kad kuo didesnei imties įrašų daliai jis būtų teisingas, t.y. $f(x_i) = y_i$. Kitaip sakant, turėdami įrašus (x_i, y_i) , mes galime *kontroliuoti* gautojo įverčio patikimumą.

Skaitinė prognozė.

Šiuo atveju Y - skaitinis. Todėl, konstruojant įvertį $\hat{Y} = f(X)$, svarbios yra prielaidos apie funkcijos f analizines savybes ir kintamojo X komponenčių įtakos svorį. Kiek įvertis atitinka tikrąjį kintamąjį Y nusako vadinamoji klaidų (nuostolių) funkcija $L(Y, f(X))$. Šios funkcijos parinkimas taip pat įtakoja galutinį rezultatą.

DT uždaviniai: nekontroliuojamas mokymas

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Imtis susideda tik iš kintamojo X reikšmių $x_1, \ldots, x_n$. Beje kintamojo X dimensija k gali būti net didesnė už imties dydį n. Išskirsime tokius nekontroliuojamo mokymo uždavinius:

DT uždaviniai: nekontroliuojamas mokymas

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Imtis susideda tik iš kintamojo X reikšmių $x_1, \ldots, x_n$. Beje kintamojo X dimensija k gali būti net didesnė už imties dydį n. Išskirsime tokius nekontroliuojamo mokymo uždavinius:

Asociacijos taisyklių konstravimas. Stengiamasi įžvelgti kokius nors dėsningumus, t.y., kokias nors "dėsningas" kintamųjų reikšmes. Paprastai šis uždavinys kyla analizuojant didelės dimensijos kategorinių kintamųjų imtis. Tipiškas pavyzdys - vadinamasis pirkėjo krepšelio uždavinys: analizuojant prekybos centro pardavimų duomenis, stengiamasi nustatyti kokios prekės dažniausiai perkamos kartu.

DT uždaviniai: nekontroliuojamas mokymas

Duomenys apie orą
Patikslinti duomenys apie orą
Regos korekcija
Irisų duomenys
Rankraščio atpažinimas
Kompiuterio našumas
Duomenys ir jų atributai
Kintamųjų tipai
DT uždaviniai: kontroliuojamas mokymas
DT uždaviniai: nekontroliuojamas mokymas

Imtis susideda tik iš kintamojo X reikšmių $x_1, \ldots, x_n$. Beje kintamojo X dimensija k gali būti net didesnė už imties dydį n. Išskirsime tokius nekontroliuojamo mokymo uždavinius:

Asociacijos taisyklių konstravimas. Stengiamasi įžvelgti kokius nors dėsningumus, t.y., kokias nors "dėsningas" kintamųjų reikšmes. Paprastai šis uždavinys kyla analizuojant didelės dimensijos kategorinių kintamųjų imtis. Tipiškas pavyzdys - vadinamasis pirkėjo krepšelio uždavinys: analizuojant prekybos centro pardavimų duomenis, stengiamasi nustatyti kokios prekės dažniausiai perkamos kartu.

Klasterinė analizė. Tikslas : turimus imties įrašus suskirstyti į tam tikras "natūralias" grupes (klasterius). Klasterių skaičius gali ir nebūti žinomas iš anksto.