

# Bajeso klasifikatoriai

# Bajeso formulė

## Bajeso formulė

Naivusis Bajeso  
klasifikatorius

Kategoriniai kintamieji

Skaitiniai kintamieji

Bajeso tinklai

Paprastas Bajeso tinklo  
pavyzdys

Bajeso tinklo

konstravimas

Bajeso tinklas ligonių  
duomenims

Sąlyginiai skirstiniai  
ligonių duomenims

Diagnozės pavyzdys

$Y$  - klasės kintamasis,

$\mathbf{X} = (X_1, X_2, \dots, X_k)$  imties atributų vektorius.

**Apriorinę tikimybę**  $P(Y = y)$  galime įvertinti tos klasės mokymo imties įrašų dalimi. Svarbesnės yra **aposteriorinės tikimybės**  $P(Y = y | \mathbf{X} = \mathbf{x})$ . Tiriamąjį įrašą  $(\mathbf{x}', y')$  reikėtų priskirti tai klasei, kurios aposteriorinė tikimybė didžiausia, t.y.

$$y' = \underset{y}{\operatorname{argmax}} P(Y = y | \mathbf{X} = \mathbf{x}').$$

Apriorines ir aposteriorines tikimybes sieja Bajeso formulė

$$P(Y = y | \mathbf{X} = \mathbf{x}') = \frac{P(Y = y) P(\mathbf{X} = \mathbf{x}' | Y = y)}{P(\mathbf{X} = \mathbf{x}')}.$$

# Naivusis Bajeso klasifikatorius

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius

Kategoriniai kintamieji

Skaitiniai kintamieji

Bajeso tinklai

Paprastas Bajeso tinklo  
pavyzdys

Bajeso tinklo

konstravimas

Bajeso tinklas ligonių  
duomenims

Sąlyginiai skirstiniai  
ligonių duomenims

Diagnozės pavyzdys

Prielaida: žinomos klasės įrašo atributai yra nepriklausomi, t.y.

$$P(\mathbf{X}|Y = y) = \prod_{i=1}^k P(X_i|Y = y) .$$

Taigi įrašo su atributų reikšmėmis  $\mathbf{x}' = (x_1, x_2, \dots, x_k)$  klasifikavimui reikalingas aposteriorines tikimybes visoms klasėms  $y$  galime išreikšti taip

$$P(Y = y | \mathbf{X} = \mathbf{x}') = P(Y = y) \prod_{i=1}^k P(X_i = x_i | Y = y) \alpha ,$$

čia daugiklis  $\alpha = 1/P(\mathbf{X} = \mathbf{x}')$  nepriklauso nuo  $y$ . Todėl reikės maksimizuoti tik dešinėje lygybės pusėje esančią apriorinės tikimybės  $P(Y = y)$  ir sąlyginių tikimybių  $P(X_i = x_i | Y = y)$  sandaugą.

# Kategoriniai kintamieji

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius

[Kategoriniai kintamieji](#)

Skaitiniai kintamieji

Bajeso tinklai  
Paprastas Bajeso tinklo  
pavyzdys

Bajeso tinklo  
konstravimas  
Bajeso tinklas ligonių  
duomenims  
Sąlyginiai skirstiniai  
ligonių duomenims  
Diagnozės pavyzdys

Kategoriniams kintamiesiems  $X_i$  tikimybės  $P(X_i = x_i | Y = y)$  prilyginamos santykiui

$$P(X_i = x_i | Y = y) = \frac{n_y(i)}{n_y}.$$

Čia  $n_y$  - klasei  $y$  priklausančių imties įrašų skaičius,  $n_y(i)$  - skaičius tų klasės  $y$  įrašų, kuriems  $X_i = x_i$ . Kartais naudojamas  $m$ -įvertis

$$P(X_i = x_i | Y = y) = \frac{n_y(i) + mp}{n_y + m}.$$

Čia  $p$  - apriorinė atributo reikšmės  $X_i = x_i$  tikimybė klasėje  $y$ . Jei atributas  $X_i$  įgyja  $m_i$  skirtingų reikšmių, tai dažniausiai pasirenkama  $p = 1/m_i$ . Neneigiamas parametras  $m$  ( ekvivalenčios imties dydis ) valdo įverčio padėtį tarp nemodifikuoto santykio  $n_y(i)/n_y$  ir apriorinės tikimybės  $p$ . Kuo didesnis  $m$  tuo įvertis yra artimesnis  $p$ .

# Skaitiniai kintamieji

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius

Kategoriniai kintamieji

[Skaitiniai kintamieji](#)

Bajeso tinklai  
Paprastas Bajeso tinklo  
pavyzdys  
Bajeso tinklo  
konstravimas  
Bajeso tinklas ligonių  
duomenims  
Sąlyginiai skirstiniai  
ligonių duomenims  
Diagnozės pavyzdys

Tolydžiųjų skaitinių kintamųjų atveju galimi du sąlyginių tikimybių  
 $P(X_i = x_i | Y = y)$  skaičiavimo būdai.

# Skaitiniai kintamieji

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius

Kategoriniai kintamieji

[Skaitiniai kintamieji](#)

Bajeso tinklai  
Paprastas Bajeso tinklo  
pavyzdys  
Bajeso tinklo  
konstravimas  
Bajeso tinklas ligonių  
duomenims  
Sąlyginiai skirstiniai  
ligonių duomenims  
Diagnozės pavyzdys

Tolydžiųjų skaitinių kintamųjų atveju galimi du sąlyginių tikimybių  $P(X_i = x_i | Y = y)$  skaičiavimo būdai.

1. Skaitinį kintamąjį  $X_i$  galima diskretizuoti.

# Skaitiniai kintamieji

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius  
Kategoriniai kintamieji  
[Skaitiniai kintamieji](#)  
Bajeso tinklai  
Paprastas Bajeso tinklo  
pavyzdys  
Bajeso tinklo  
konstravimas  
Bajeso tinklas ligonių  
duomenims  
Sąlyginiai skirstiniai  
ligonių duomenims  
Diagnozės pavyzdys

Tolydžiųjų skaitinių kintamųjų atveju galimi du sąlyginių tikimybių  $P(X_i = x_i | Y = y)$  skaičiavimo būdai.

1. Skaitinį kintamąjį  $X_i$  galima diskretizuoti.
2. Tegul  $p_i(x|j)$  yra a.d.  $X_i$  sąlyginė tankio funkcija su sąlyga  $Y = y_j$ . Pasirinkę mažą  $\varepsilon > 0$ , turėsime

$$P(x_i \leq X_i \leq x_i + \varepsilon | Y = y_j) = \int_{x_i}^{x_i + \varepsilon} p_i(x|j) dx \approx \varepsilon \cdot p_i(x_i|j)$$

Todėl skaičiavimuose tikimybes  $P(X_i = x_i | Y = y_j)$  kartais tiesiog pakeisime tankio funkcijos reikšmėmis  $p_i(x_i|j)$ .

Dažniausiai pasirenkamas normaliojo dėsnio tankis, priklausantis nuo dviejų parametrų: vidurkio  $\mu_{ij}$  ir dispersijos  $\sigma_{ij}^2$ .

Jei parametrai  $\mu_{ij}$  ir  $\sigma_{ij}^2$  yra nežinomi, jie keičiami atitinkamais įverčiais: imties vidurkiu  $\bar{x}_{ij}$  ir dispersija  $s_{ij}^2$ .

# Bajeso tinklai

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius  
Kategoriniai kintamieji  
Skaitiniai kintamieji  
**Bajeso tinklai**  
Paprastas Bajeso tinklo  
pavyzdys  
Bajeso tinklo  
konstravimas  
Bajeso tinklas ligonių  
duomenims  
Sąlyginiai skirstiniai  
ligonių duomenims  
Diagnozės pavyzdys

Tarkime, kad imties kintamieji vaizduojami grafo viršūnėmis, o bet kurių dviejų kintamųjų tarpusavio priklausomybę ( jei tokia yra ) atspindi šias viršūnes jungianti briauna.

**Apibrėžimas.** *Sakysime, kad kintamieji sudaro Bajeso tinklą, jei tenkinamos dvi sąlygos:*

- 1) kintamuosius vaizduojantis grafas yra orientuotas ir neturi ciklų,*
- 2) jei žinomi viršūnės tėvai, tai pati viršūnė nepriklauso nuo visų likusių viršūnių, išskyrus jos palikuonis.*

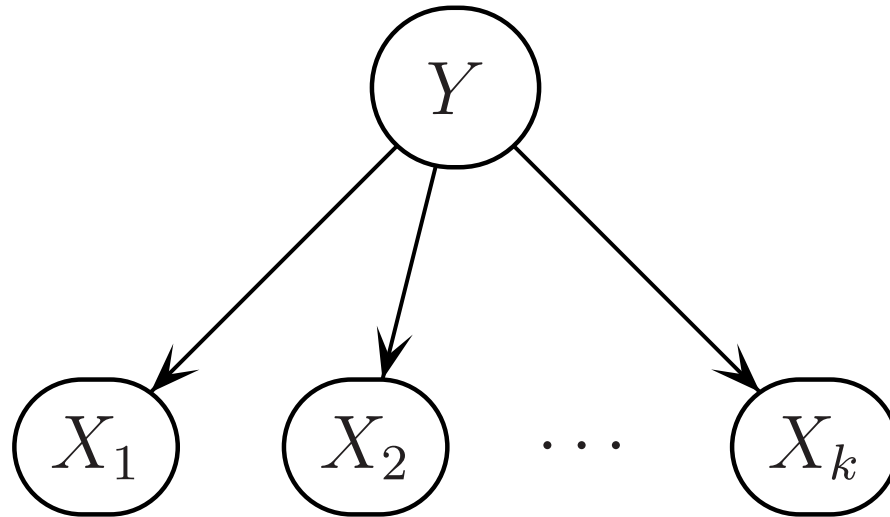
Bajeso tinklas nusakomas viršūnėse esančių kintamųjų skirstiniais. Jei viršūnė turi tėvus, tai ją atitinkantis skirstinys yra sąlyginis. Priešingu atveju - nesąlyginis.



## Paprastas Bajeso tinklo pavyzdys

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius  
Kategoriniai kintamieji  
Skaitiniai kintamieji  
Bajeso tinklai  
Paprastas Bajeso tinklo  
pavyzdys  
Bajeso tinklo  
konstravimas  
Bajeso tinklas ligonių  
duomenims  
Sąlyginiai skirstiniai  
ligonių duomenims  
Diagnozės pavyzdys

Naivusis Bajeso metodas taikomas esant prielaidai, kad žinomos klasės  $Y = y$  įrašo atributai  $\mathbf{X} = (X_1, X_2, \dots, X_k)$  yra nepriklausomi. Todėl jie sudaro Bajeso tinklą, kurį vaizduoja toks grafas



# Bajeso tinklo konstravimas

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius  
Kategoriniai kintamieji  
Skaitiniai kintamieji  
Bajeso tinklai  
Paprastas Bajeso tinklo  
pavyzdys  
[Bajeso tinklo  
konstravimas](#)  
Bajeso tinklas ligonių  
duomenims  
Sąlyginiai skirstiniai  
ligonių duomenims  
Diagnozės pavyzdys

Bajeso tinklo klasifikatoriaus konstravimas susideda iš dviejų etapų.

1. Nustatoma tinklo struktūra. Tam reikia nurodyti priklausomus kintamuosius.
2. Pagal turimą imtį randami visų viršūnių skirstiniai.

Pavyzdys. Tegul informacija apie pacientą nusakoma šešiais binariniais kintamaisiais  $(X_1, X_2, X_3, X_4, X_5, Y)$  :

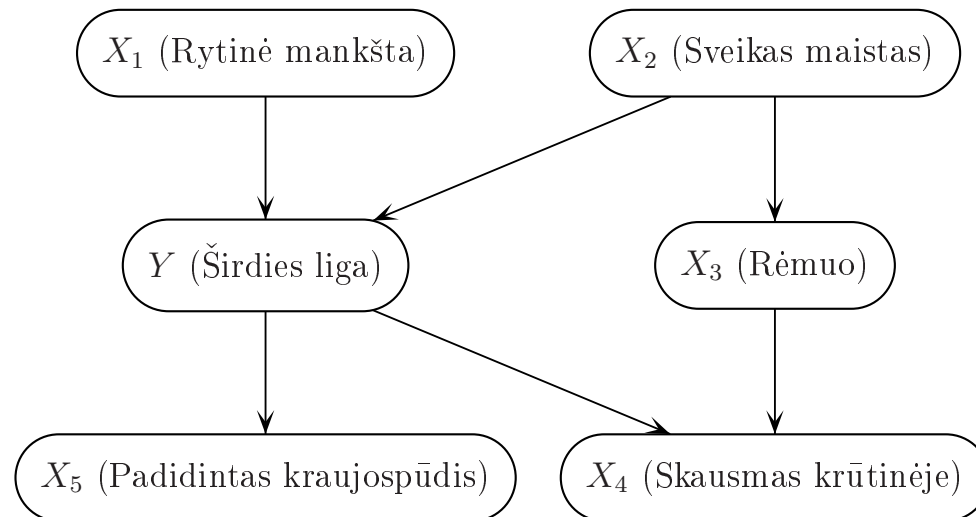
$X_1$  - Rytinė mankšta,  
 $X_2$  - Sveikas maistas,  
 $X_3$  - Rėmuo,  
 $X_4$  - Skausmas krūtinėje,  
 $X_5$  - Padidintas kraujospūdis,  
 $Y$  - Širdies liga.

Kiekvienas iš kintamųjų įgyja reikšmę 1, jei atitinkamas požymis yra, ir 0 priešingu atveju.

# Bajeso tinklas ligonių duomenims

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius  
Kategoriniai kintamieji  
Skaitiniai kintamieji  
Bajeso tinklai  
Paprastas Bajeso tinklo  
pavyzdys  
Bajeso tinklo  
konstravimas  
Bajeso tinklas ligonių  
duomenims  
Sąlyginiai skirstiniai  
ligonių duomenims  
Diagnozės pavyzdys

Atsižvelgiant į medikų nuomonę nustatyta požymių tarpusavio priklausomybė nusakoma grafu



Pagal turimus imties duomenis 70% pacientų rytais sportuoja ir tik 25% sveikai maitinasi. Todėl tėvų neturinčių kintamųjų  $X_1$  ir  $X_2$  skirstinius nusakys lygybės

$$P(X_1 = 0) = 1 - P(X_1 = 1) = 0,3,$$

$$P(X_2 = 0) = 1 - P(X_2 = 1) = 0,75.$$

# Sąlyginiai skirstiniai ligonių duomenims

Bajeso formulė  
Naivusis Bajeso  
klasifikatorius  
Kategoriniai kintamieji  
Skaitiniai kintamieji  
Bajeso tinklai  
Paprastas Bajeso tinklo  
pavyzdys  
Bajeso tinklo  
konstravimas  
Bajeso tinklas ligonių  
duomenims  
[Sąlyginiai skirstiniai  
ligonių duomenims](#)  
Diagnozės pavyzdys

Likusios keturios viršūnės turi tėvus. Tarsime, kad kintamųjų  $X_3$ ,  $X_4$ ,  $X_5$  ir  $Y$  sąlyginiai skirstiniai yra

				$X_4$	
		$X_3$	$Y$	0	1
$X_2$	0	0,15	0,85	0,9	0,1
	1	0,8	0,2	0,4	0,6
	0	0,15	0,85	0,6	0,4
	1	0,8	0,2	0,2	0,8

				$Y$	
		$X_1$	$X_2$	0	1
$Y$	0	0,8	0,2	0,25	0,75
	1	0,15	0,85	0,45	0,55
	0	0,8	0,2	0,55	0,45
	1	0,15	0,85	0,75	0,25

## Diagnozės pavyzdys

Turintis aukštą kraujospūdį pacientas reguliariai mankština ir valgo sveiką maistą. Ar jis serga?

$$\begin{aligned} & P(Y = 1 | X_5 = 1, X_2 = 1, X_1 = 1) \\ = & \frac{P(X_5 = 1 | Y = 1, X_2 = 1, X_1 = 1) P(Y = 1 | X_2 = 1, X_1 = 1)}{P(X_5 = 1 | X_2 = 1, X_1 = 1)} \\ = & \frac{P(X_5 = 1 | Y = 1) P(Y = 1 | X_2 = 1, X_1 = 1)}{\sum_{i \in \{0,1\}} P(X_5 = 1 | Y = i) P(Y = i | X_2 = 1, X_1 = 1)} \\ = & \frac{0,85 \cdot 0,25}{0,2 \cdot 0,75 + 0,85 \cdot 0,25} \approx 0,5862. \end{aligned}$$

Tikimybė, kad pacientas neserga bus

$$P(Y = 0 | X_5 = 1, X_2 = 1, X_1 = 1) \approx 1 - 0,5862 = 0,4138.$$