

# 1 BAYESOVKÁ KLASIFIKACE ZALOŽENÁ NA KOPULA FUNKCÍCH

## RYCHLÝ NÁHLED KAPITOLY



S *klasifikací* se setkáváme velmi často: učitel klasifikuje žáka podle jeho znalostí, lékař klasifikuje pacienta podle toho, jaké má pacient příznaky. Množství takových klasifikačních úloh lze automatizovat pomocí počítačů a metod *Dolování dat*, které se řadí do širšího oboru *Umělá inteligence*. Náplní této kapitoly je seznámit čtenáře s jednou takovou metodou, která byla poměrně nedávno navržena na Slezské univerzitě – metodou *bayesovské klasifikace založené na kopula funkcích*.

## CÍLE KAPITOLY



Po prostudování této kapitoly budete vědět:

- co je to bayesovská klasifikace,
- co je to Naivní Bayes,
- co jsou kopula funkce,
- jak se konstruuje a jak pracuje bayesovský klasifikátor založený na kopula funkcích a
- jak je výkonný takový klasifikátor ve srovnání s dalšími klasifikátory.

## KLÍČOVÁ SLOVA KAPITOLY



klasifikace, kopule, Bayesova věta, Naivní bayes

Klasifikace je problém, se kterým se lidé na celém světě setkávají každý den, ačkoliv si to často ani neuvědomují. Například doktoři v nemocnici klasifikují pacienty podle toho, jak jsou ohroženi určitou chorobou, nebo učitel rozhoduje, jakou známku dát studentům na základě výsledků testů.

Díky velkému množství dat, která se v současné době uchovávají (např. data vznikající na sociálních sítích, finančních trzích, databáze pacientů v nemocnicích, apod.), lze značně

množství klasifikačních problémů řešit automatizovaně pomocí počítačů a tzv. *klasifikačních metod*.

Hlavním požadavkem pro využití takových metod je mít k dispozici tzv. *klasifikační data* k danému problému, které si lze představit jako tabulku, kde jeden řádek odpovídá jednomu objektu, a v každém z těchto řádků jsou ve sloupcích obsaženy všechny relevantní informace právě pro jeden klasifikovaný objekt (tzv. *vstupní atributy*) a také jeho *třída*. Např. řádek odpovídá pacientovi a ve sloupcích jsou pak informace jako výška, váha, krevní tlak, cholesterol v krvi apod. a také třída přiřazená tomuto pacientovi, např. informace o tom zda pacient trpí či netrpí danou chorobou. Cílem klasifikační metody je poté vytvořit *model* popisující souvislosti mezi vstupními atributy a třídou. Takový model lze pak využít tak, že objeví-li se nový pacient, stačí model aplikovat na vstupní atributy tohoto pacienta a dostaneme předpověď, danou tímto modelem, toho, zda pacient trpí či netrpí danou chorobou, tzn. dostaneme diagnózu pacienta aniž by pacienta musel vyšetřovat lékař.

Mezi velmi populární klasifikační metody patří tzv. *bayesovské klasifikátory*, které se úspěšně využívají např. pro detekci spamu či při rozpoznávání obrazu. Tyto klasifikátory jsou představeny v následující kapitole.



## KONTROLNÍ OTÁZKA

1. Co je to klasifikační metoda a jak vypadají klasifikační data?

### 1.1 Bayesovské klasifikátory

Nechť  $\Omega = \{\omega_1, \dots, \omega_m\}$  je množina obsahující  $m$  tříd. Např.  $m = 2$  a  $\omega_1 = \text{riziko infarktu}$  a  $\omega_2 = \text{bez rizika}$ . Obecným cílem klasifikace je poté přiřadit každému objektu popsanému vstupními atributy (vektorem)  $x \in \mathbb{R}^d$  nějakou třídu z  $\Omega$ , např. je-li  $d = 4$  a jsou-li  $x_1$  až  $x_4$  vstupní atributy popisující výšku, váhu, krevní tlak a cholesterol v krvi u daného pacienta, potom cílem klasifikace může být přiřadit každému pacientovi s danými hodnotami atributů  $x_1$  až  $x_4$  buď třídu  $\omega_1$  nebo  $\omega_2$ .

*Bayesovský klasifikátor* přiřazuje objektu popsanému atributy  $x$  třídu  $\omega_i$ , pokud platí

$$g_i(x) > g_j(x) \quad \text{pro všechny } j \neq i$$

kde  $g_i: [0, \infty)^d \rightarrow \mathbb{R}$ ,  $i = 1, \dots, m$  jsou tzv. *diskriminační funkce*, které jsou definovány jako

$$g_i(x) = \mathbb{P}(\omega_i|x),$$

tedy odpovídají pravděpodobnosti, že objekt  $x$  patří do třídy  $\omega_i$ . Jinými slovy, Bayesovský klasifikátor přiřazuje  $x$  do té třídy, do které patří s největší pravděpodobností. Jelikož se ale

hodnota výrazu  $\mathbb{P}(\omega_i|x)$  obtížně odhaduje, využívá se tvar diskriminační funkce přeepsaný pomocí známého Bayesova pravidla, tedy ve tvaru

$$g_i(x) = \mathbb{P}(\omega_i|x) = \frac{f(x|\omega_i)\mathbb{P}(\omega_i)}{\sum_{j=1}^m f(x|\omega_j)\mathbb{P}(\omega_j)},$$

kde funkce  $f(x|\omega_i)$  je hustota pravděpodobnosti vektoru vstupních atributů za předpokladu, že objekt patří do třídy  $\omega_i$ , a  $\mathbb{P}(\omega_i), i = 1, \dots, m$  jsou apriorní pravděpodobnosti tříd z  $\Omega$ . Jelikož jmenovatel v předchozím výrazu je stejný pro všechna  $g_i$ , lze  $g_i$  zjednodušit na

$$g_i(x) = f(x|\omega_i) + \mathbb{P}(\omega_i)$$

aniž by to ovlivnilo výsledek klasifikace (daný dříve uvedenou nerovností).

Chceme-li tedy vybudovat model pro bayesovskou klasifikaci, je potřeba určit pro každou třídu v datech jednu diskriminační funkci  $g_i$ . Podíváme-li se na její zjednodušenou definici, vidíme, že se skládá ze součtu dvou výrazů. Máme-li klasifikační data, která obsahují např. 2 třídy, a tyto třídy jsou zastoupeny v datech stejně, např. 50 pacientů s rizikem infarktu a 50 bez rizika infarktu, odhad apriorní pravděpodobnosti  $\mathbb{P}(\omega_i)$  značený  $\bar{\mathbb{P}}(\omega_i)$  se nejčastěji určuje jako poměrné zastoupení třídy v datech, tedy v našem příkladu  $\bar{\mathbb{P}}(\omega_1) = \bar{\mathbb{P}}(\omega_2) = \frac{50}{100} = \frac{1}{2}$ . Složitějším problémem je určit odhad hustoty  $f(x|\omega_i)$ . Nejjednodušší způsob, jakým se k odhadu hustoty  $f(x|\omega_i)$  přistupuje, lze nalézt v přístupu nejpobulárnějšího bayesovského klasifikátoru, který je nazýván jako *Naivní Bayes*, a který je představen v následující kapitole.

## KONTROLNÍ OTÁZKA



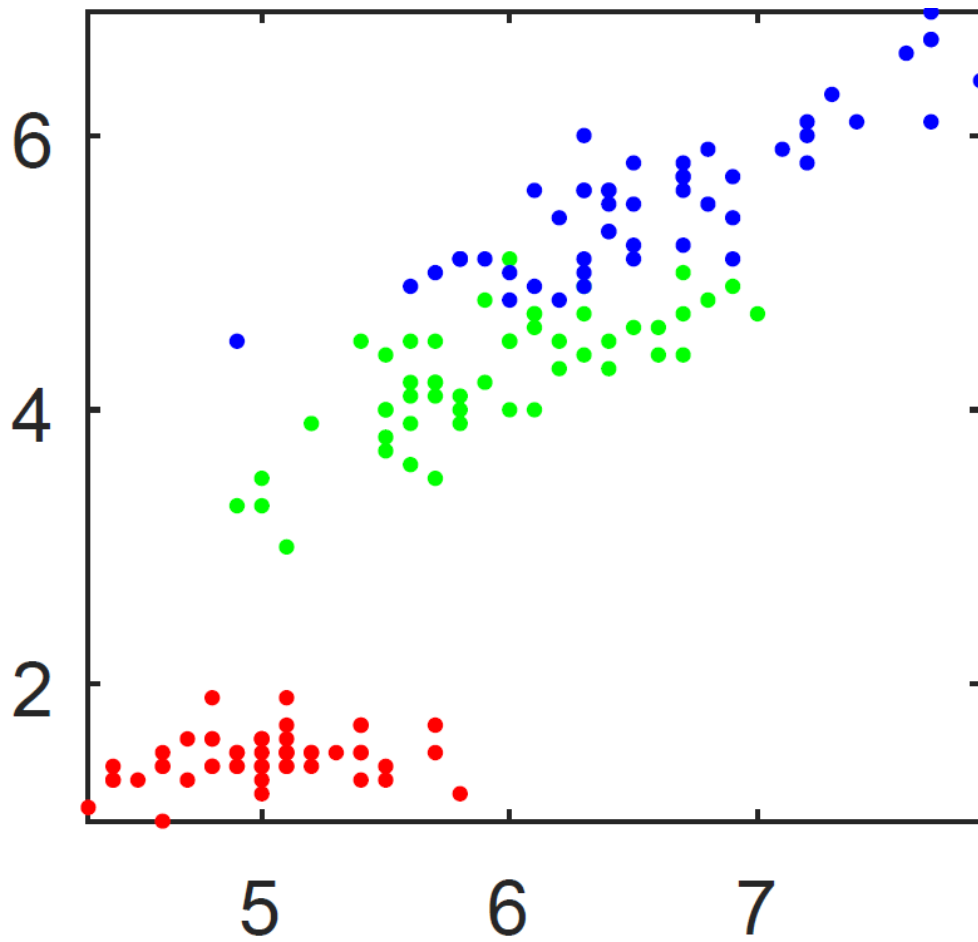
1. Jak pracuje bayesovský klasifikátor?
2. Co je to diskriminační funkce?
3. Jakou úlohu v ní hraje apriorní pravděpodobnost třídy a jak se její hodnota odhaduje?

## 1.2 Naivní Bayes

*Naivní Bayes* využívá zjednodušujícího předpokladu, že všechny vstupní atributy jsou navzájem nezávislé. Matematicky vyjádřeno, tedy že platí

$$f(x|\omega_i) = f_1(x_1|\omega_i)f_2(x_2|\omega_i) \dots f_d(x_d|\omega_i).$$

To znamená, že sdružená hustota  $f(x|\omega_i)$  je dána součinem marginálních hustot  $f_j(x_j|\omega_i)$  pro  $j = 1, \dots, d$ . Ilustrujme si nyní funkčnost tohoto klasifikátoru na datech zobrazených na Obrázek 1.

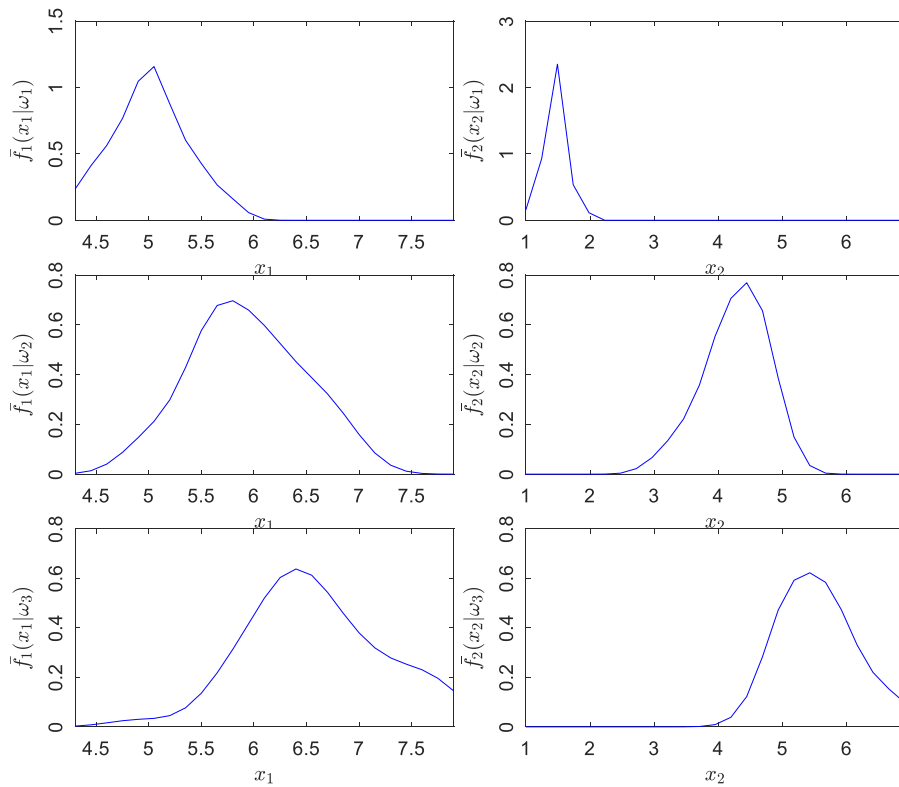


**Obrázek 1: Data obsahující dva vstupní atributy (vodorovná a svislá osa) a tři třídy (červená, zelená a modrá) po 50ti objektech.**

Pro tato data je tedy  $d = 2$  a  $m = 3$ . Jelikož jsou všechny třídy v datech zastoupeny stejně, tzn. 50 červených, 50 zelených a 50 modrých objektů, odhadneme apriorní pravděpodobnosti jako  $\bar{\mathbb{P}}(\omega_1) = \bar{\mathbb{P}}(\omega_2) = \bar{\mathbb{P}}(\omega_3) = \frac{50}{150} = \frac{1}{3}$ . Pro odhad sdružené hustoty  $f(x|\omega_i)$  využijeme předpoklad o nezávislosti vstupních atributů a budeme tedy její odhad  $\bar{f}(x|\omega_i)$  počítat z

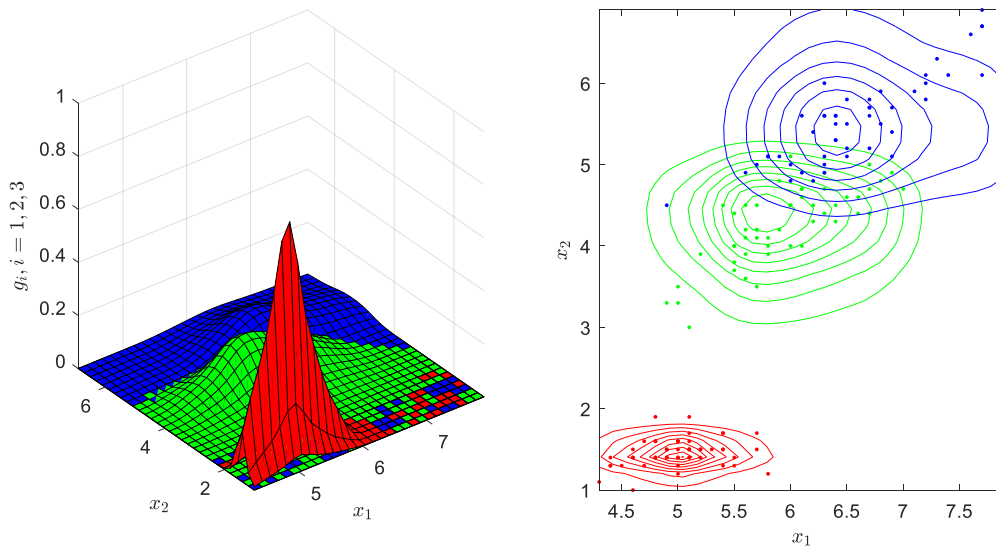
$$\bar{f}(x|\omega_i) = \bar{f}_1(x_1|\omega_i)\bar{f}_2(x_2|\omega_i) \dots \bar{f}_d(x_d|\omega_i),$$

kde  $\bar{f}_j(x_j|\omega_i)$  je odhad hustoty rozdělení  $j$ -tého vstupního atributu,  $j = 1, \dots, d$ , pro třídu  $\omega_i$ ,  $i = 1, \dots, m$ . Na základě daných dat spočteme odhady  $\bar{f}_j(x_j|\omega_i)$  pro všechna  $i = 1, 2, 3$  a  $j = 1, 2$ . Např. s využitím tzv. *hladkého jádrového odhadu* (kernel smooth estimation – viz Górecki et al. (2016)) získáme odhady zobrazené na obrázku níže.



**Obrázek 2: Odhady  $\bar{f}_j(x_j|\omega_i)$  pro všechna  $i = 1, 2, 3$  a  $j = 1, 2$  na základě dat s Obrázek 1**

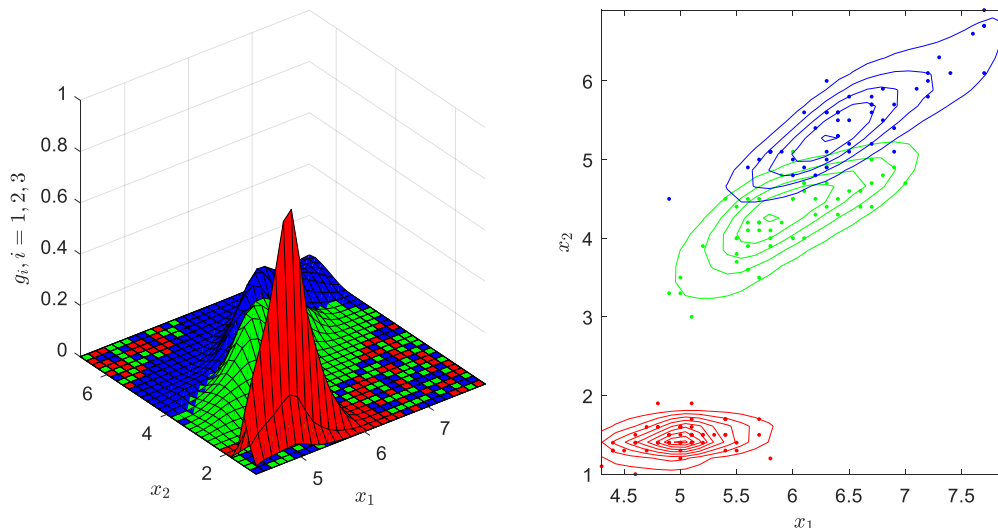
Tyto odhady pak již určují  $\bar{f}(x|\omega_i)$ , čímž je tedy dána i diskriminační funkce  $g_i$  pro  $i = 1, \dots, 3$ , viz Obrázek 3.



**Obrázek 3: (vlevo)  $g_1$  (červená)  $g_2$  (zelená) a  $g_3$  (modrá). (vpravo) Kontury (vrstevnice) funkcí  $g_1, g_2$  a  $g_3$ .**

Nový objekt  $x$  pak klasifikujeme do takové třídy  $\omega_i$ , pro kterou je diskriminační funkce  $g_i(x)$  největší. Např. pro objekt (bod) s  $x_1 = 6$  a  $x_2 = 4$  je jasně vidět, že největší hodnotu v tomto bodě nabývá funkce  $g_2$ . Klasifikátor by tedy zařadil tento objekt do třídy  $\omega_2$  (zelená).

Všimněme si ale, že odhad sružené hustoty v určitém smyslu příliš neodpovídá daným datům – zvláště pro zelené a modré objekty lze dobře vidět, že objekty jsou jakoby nataženy na vedlejší diagonále (zdola vlevo nahoru doprava), a tudíž by takovéto rozložení objektů měla reflektovat i sružená hustota, tedy měla být více nakloněna/natažena na vedlejší diagonále, např. tak jak je naznačeno na Obrázek 4. Naivní Bayes však předpokládá nezávislost vstupních atributů, což takovéto naklonění neumožňuje a sružené hustoty jsou vždy natočeny ve směru os (viz Obrázek 3). Tento předpoklad (často považovaný za naivní – z toho plyne i jméno tohoto klasifikátoru) poté také může způsobovat horší klasifikační schopnosti konstruovaného modelu. Pro odstranění tohoto nedostatku je možno využít tzv. *kopula* funkcí (nebo jednodušeji *kopulí*), které budou stručně představeny v následující části textu.



Obrázek 4: Diskriminační funkce porušující předpoklad nezávislosti vstupních atributů.



### KONTROLNÍ OTÁZKA

1. Na jakém předpokladu stojí Naivní Bayes?
  2. Kterou část diskriminační funkce tento předpoklad ovlivňuje?
-

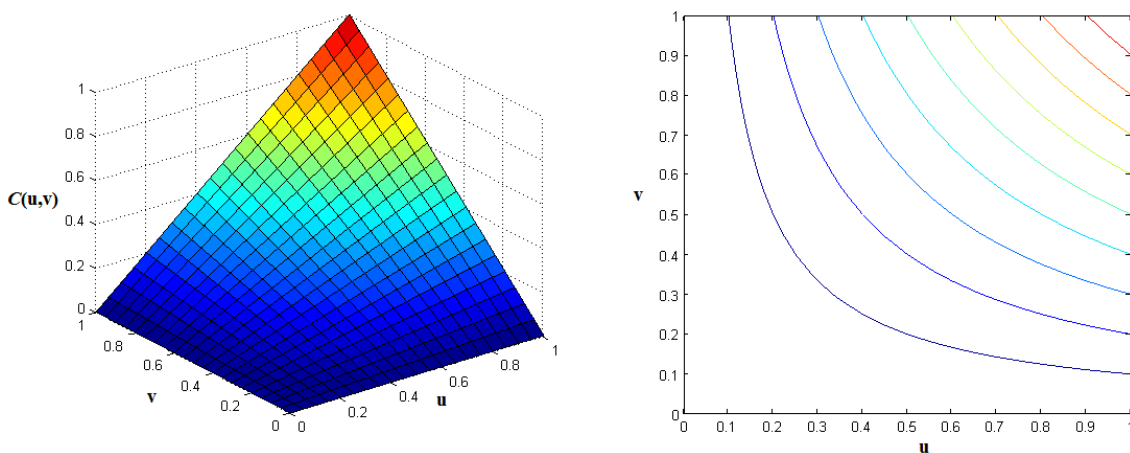
## 1.3 Kopule

### DEFINICE



*Kopule* (z latinského *copula* (česky *spoj, spojení*)) je sdružená distribuční funkce mající všechny jednorozměrné marginální distribuční funkce stejnoměrně rozdělené na jednotkovém intervalu.

Příklad dvourozměrné kopule lze vidět na Obrázek 5.



**Obrázek 5:** (vlevo) Distribuční funkce tzv. kopule nezávislosti dané  $C(u, v) = u * v$ . (vpravo) Kontura této distribuční funkce.

Kopule tedy tvoří nějakou speciální třídu sdružených distribučních funkcí. Co tuto třídu činí zajímavou, je Sklarova věta.

### VĚTA



*Sklarova věta (1959)*. Nechť  $F$  je  $d$ -rozměrná distribuční funkce s jednorozměrnými marginály  $F_1, \dots, F_d$ . Pak existuje kopule  $C: [0,1]^d \rightarrow [0,1]$  taková, že platí

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \quad (1)$$

pro všechna  $(x_1, \dots, x_d) \in \mathbb{R}^d$ . Tato funkce  $C$  je jednoznačně určena, pokud  $F_1, \dots, F_d$  jsou spojitě. Naopak, pokud je funkce  $C$  kopule a  $F_1, \dots, F_d$  jsou jednorozměrné distribuční funkce, potom funkce daná (1) je sdruženou distribuční funkcí s marginály  $F_1, \dots, F_d$ .

Hustotu sdružené distribuční funkce  $F$  danou (1) lze vyjádřit ( $d$ -krát derivujeme postupně podle  $x_1$  až  $x_d$ ) jako

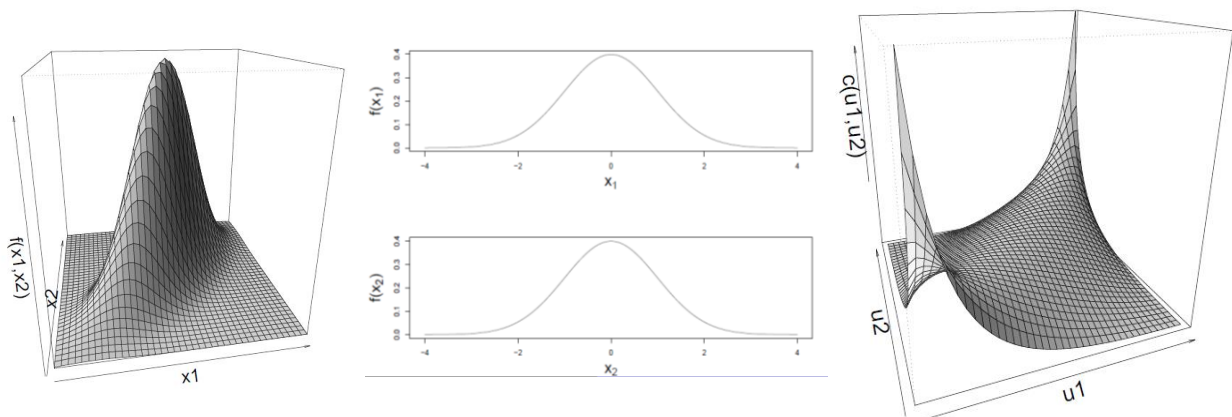
$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i), \quad (2)$$

kde  $f_i(x_i)$  je hustota  $F_i(x_i)$  pro všechna  $i = 1, \dots, d$  a

$$c(u_1, \dots, u_d) = \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d} \quad (3)$$

hustota kopule  $C$ .

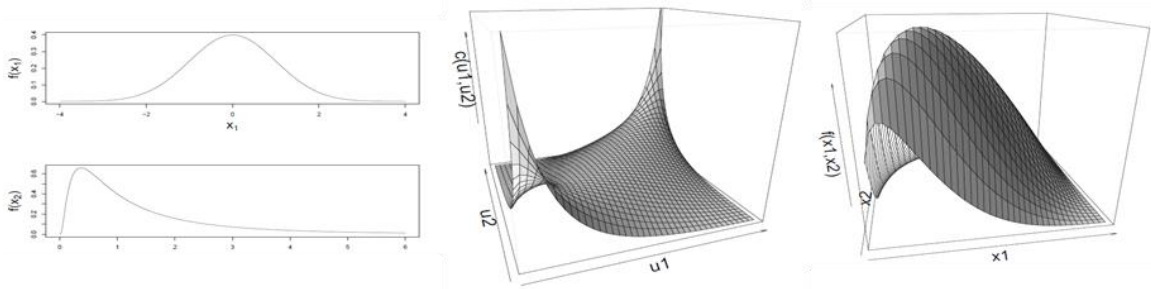
Ilustrujme si význam Sklarovy věty na následujícím příkladu. Máme-li např. dvě náhodné veličiny  $(X_1, X_2)$  rozdělené podle dvourozměrného standardního normálního rozdělení  $F$  (jeho hustota  $f$  je na Obrázek 6 vlevo), pak Sklarova věta říká, že toto  $F$  lze jednoznačně rozložit do dvou komponent: 1) na jednorozměrné marginály (viz Obrázek 6 uprostřed) a 2) na kopuli (viz Obrázek 6 vpravo). Jelikož jednorozměrné marginály neobsahují žádnou informaci o vztazích modelovaných proměnných  $X_1$  a  $X_2$  mezi sebou, je veškerá tato informace obsažena právě v této kopuli. Toto nám například umožňuje studovat závislost mezi proměnnými *zcela odděleně* od jednorozměrných marginálních rozdělení, což se velmi často používá např. ve finančnictví nebo hydrologii.



**Obrázek 6: (vlevo) Hustota dvourozměrného standardního normálního rozdělení. (uprostřed) Hustoty jednorozměrných marginálů rozdělení zobrazeného vlevo. (vpravo) Hustota kopule rozdělení zobrazeného vlevo.**

Druhý způsob, jak využít Sklarovu větu je ke konstrukci nových sdružených distribučních funkcí. Např. vezmeme-li jiné marginální distribuční funkce než ty z Obrázek 6, a to např. ty z Obrázek 7 vlevo, a spojíme-li je pomocí Sklarovy věty s kopulí z Obrázek 6 vpravo (je to ta stejná co je zobrazena i na Obrázek 7 uprostřed), pak dostaneme novou sdruženou distribuční funkci, jejíž hustota je zobrazena na Obrázek 7 vpravo. Tímto je možno budovat velmi flexibilní sdružené distribuční funkce, které pak lze použít k modelování dat s komplikovanou strukturou.





**Obrázek 7: Hustoty jednorozměrných marginály, kde druhý marginál není normální. (uprostřed) Hustota normální kopule. (vpravo) Hustota nová sdružené distribuční funkce daná marginály vlevo a kopulí uprostřed skrze Sklarovu větu.**

V praxi se používají kopule, jejichž vlastnosti jsou dány jedním či více parametry. Mluví se pak o tzv. *parametrických rodinách* kopulí. Představme si nyní několik takových významných rodin.

### 1.3.1 ELIPTICKÉ KOPULE

*Eliptické kopule* vycházejí z tzv. eliptických rozdělení, viz Nelsen (2006). Nejpoužívanějšími rodinami jsou rodiny gaussovské a studentovské a jejich vyjádření vychází přímo ze Sklarovy věty.

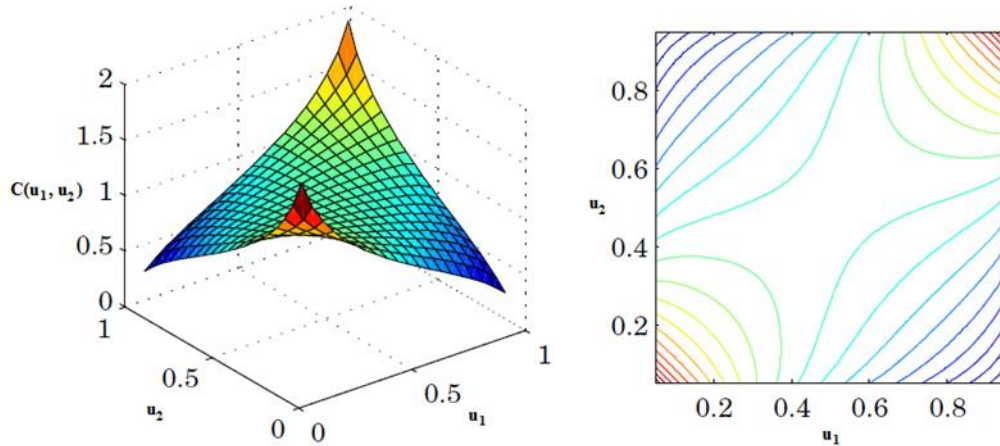
#### GAUSSOVSKÉ KOPULE

*Gaussovskou kopuli*  $C_{\Sigma}^{Ga}$  lze vyjádřit jako

$$C_{\Sigma}^{Ga}(u_1, \dots, u_d) = \Phi_{\Sigma}(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)),$$

kde  $\phi$  je jednorozměrná standardní normální distribuční a  $\Phi_{\Sigma}$  je sdružená normální distri

buce s korelační maticí  $\Sigma$  a vektorem středů rovným 0. Dvourozměrnou gaussovskou kopuli lze vidět na Obrázek 8. Její odvození, které vychází přímo ze Sklarovy věty, jsme si již ukázali v Kapitole 1.3.



**Obrázek 8: Hustota (vlevo) a kontura (vpravo) gaussovské kopule s parametrem korelační matice  $\Sigma$  rovným 0,3.**

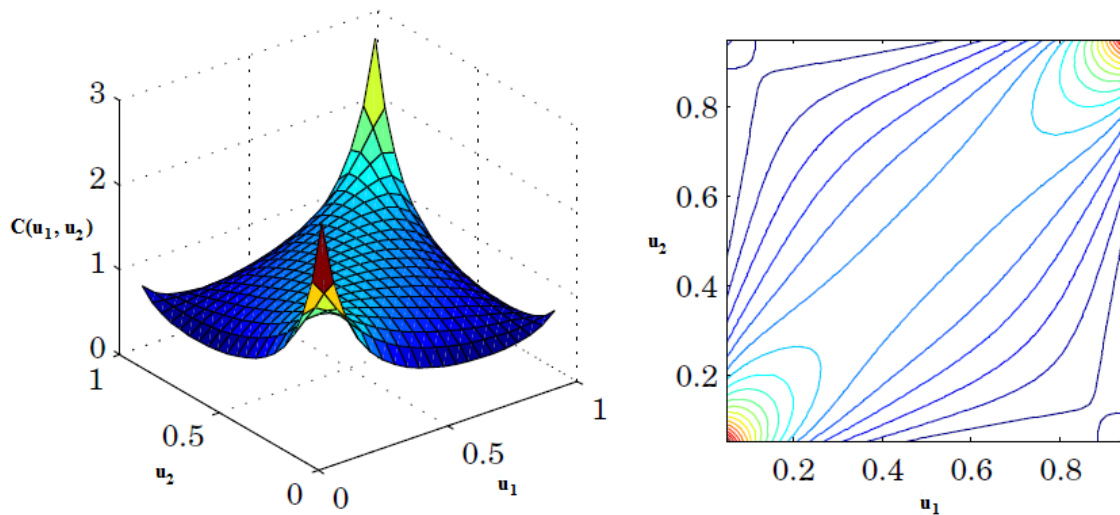
### STUDENTOVSKÉ T-KOPULE

Studentovskou  $t$ -kopuli  $C_{v,\Sigma}^t$  lze vyjádřit jako

$$C_{v,\Sigma}^t(u_1, \dots, u_d) = t_{v,\Sigma}(t_v^{-1}(u_1), \dots, t_v^{-1}(u_d)),$$

kde  $t_v$  je jednorozměrná studentovská distribuční funkce s  $v$  stupni volnosti a kde  $t_{v,\Sigma}$  je sdužená studentovská distribuční funkce s korelační maticí  $\Sigma$  a  $v$  stupni volnosti.

Dvourozměrnou studentovskou kopuli lze vidět na Obrázek 9. Její odvození je stejné jako pro gaussovskou kopuli.



**Obrázek 9: Hustota (vlevo) a kontura (vpravo) studentovské kopule s parametrem korelační matice  $\Sigma$  rovným 0,3 a stupni volnosti  $v$  rovné 3.**

### 1.3.2 ARCHIMEDOVSKÉ KOPULE

Při vytváření *archimedovských kopulí* se ze Sklarovy věty nevyhází (na rozdíl od eliptických kopulí), ale vychází se z nějaké funkční formy, po které se poté chce, aby splňovala vlastnosti kopule. Pro konstrukci archimedovských kopulí se vychází z tzv. *archimedovského generátoru*.

#### DEFINICE

*Archimedovský generátor* je spojitá a nerostoucí funkce  $\psi: [0, \infty] \rightarrow [0, 1]$ , která splňuje  $\psi(0) = 1, \psi(\infty) = \lim_{t \rightarrow \infty} \psi(t) = 0$ , a která je přísně klesající na  $[0, \inf\{t \mid \psi(t) = 0\}]$ .

*Archimedovská kopule* je pak jakákoli kopule  $C$ , která je vyjádřitelná ve tvaru

$$C(u_1, \dots, u_d) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)), \quad u_1, \dots, u_d \in [0, 1]^d$$

kde  $\psi$  je archimedovský generátor a  $\psi^{-1}: [0, 1] \rightarrow [0, \infty]$  je definována jako  $\psi^{-1}(s) = \inf\{t: \psi(t) = s\}, s \in [0, 1]$ .

Tabulka 1 představuje tři často používané rodiny archimedovských kopulí parametrizované jedním parametrem.

**Tabulka 1: Archimedovské rodiny s jedním parametrem  $\theta$**

Rodina	Rozsah parametru $\theta$	$\psi(t)$
Claytonova	$[-1, \infty) \setminus \{0\}$	$(1 + \theta t)^{-1/\theta}$
Gumbelova	$[1, \infty)$	$\exp(-t^{1/\theta})$
Frankova	$(-\infty, \infty) \setminus \{0\}$	$-\frac{1}{\theta} \log(1 + \exp(-t)(\exp(-\theta) - 1))$

Zdroj: Górecki et al. (2016)

#### CLAYTONOVY KOPULE

Rodina Claytonových kopulí je hojně využívána ve finančnictví ke studiím tzv. *korelačních rizik*, viz Nelsen (2006). Kopuli z této rodiny s parametrem  $\theta = 1$  lze vidět na Obrázek 10a), kde můžeme vidět velké hustoty v dolním rohu, která je typická právě pro data pocházející z finančních trhů.

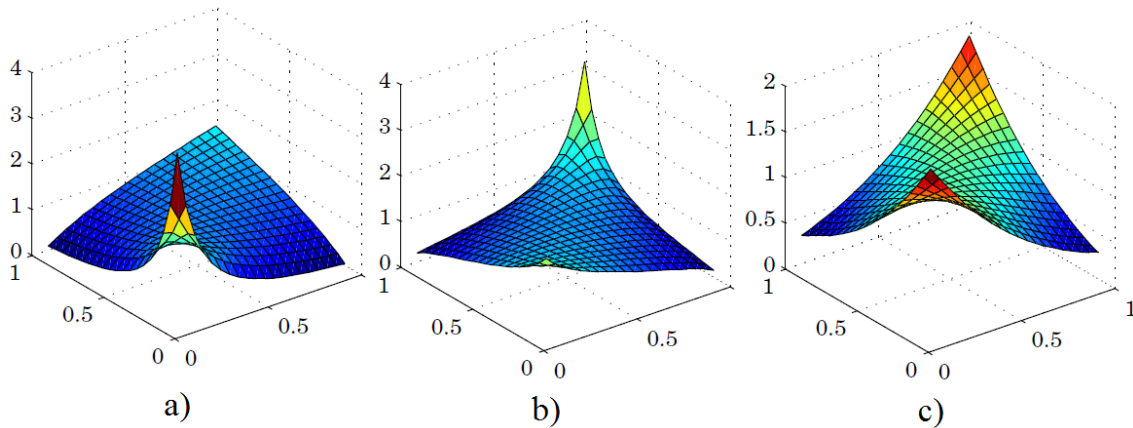
#### GUMBELOVY KOPULE

Rodina Gumbelových kopulí je schopna modelovat asymetrickou závislost, avšak opačně než Claytonovy kopule, viz Obrázek 10b) (velká hustota je tentokrát v horním rohu).



## FRANKOVA KOPULE

Dvourozměrná Frankova kopule je na rozdíl o dvou předchozích rodin podle vedlejší diagonály symetrická, viz Obrázek 10c (stejná hustota jak v horním, tak dolním rohu).



**Obrázek 10: Hustota třech archimedovských kopulí: a) Claytonova, b) Gumbelova a c) Frankova.**

Nyní jsme se tedy s kopulemi a některými jejich rodinami seznámili a můžeme se vrátit k zpět k bayesovské klasifikaci, která však nyní již bude využívat i teorii související s kopulemi.



### KONTROLNÍ OTÁZKA

1. Co je to kopule?
2. Co říká Sklarova věta?
3. Jaké jsou příklady rodin kopulí a co určuje jejich vlastnosti?

## 1.4 Bayesovský klasifikátor založený na kopulích

Zopakujme si, že za předpokladu, že  $F$  dána (1) je absolutně spojitá sdružená distribuční funkce s marginály  $F_1, \dots, F_d$ , její hustotu (značenou  $f$ ) lze vyjádřit jako

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{k=1}^d f_k(x_k), \quad (4)$$

kde  $c(u_1, \dots, u_d)$  je hustota kopule  $C(u_1, \dots, u_d)$  a  $f_k$  je hustota  $F_k, k = 1, \dots, d$ . Užitím (4) lze  $f(x|\omega_i)$  z Kapitoly 1.1 přepsat do tvaru

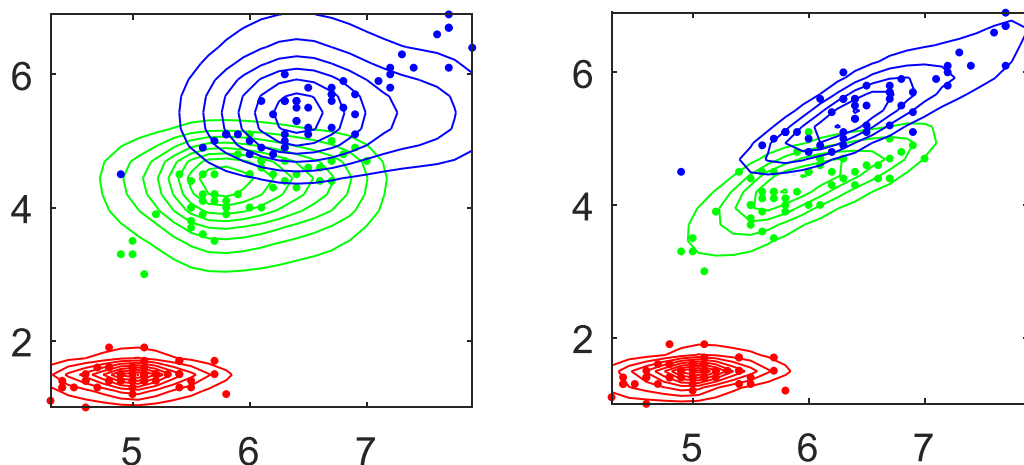
$$f(x|\omega_i) = c(F_1(x_1|\omega_i), \dots, F_d(x_d|\omega_i)|\omega_i) \prod_{k=1}^d f_k(x_k|\omega_i),$$

Diskriminační funkci  $g_i$  lze tedy přepsat na

$$g_i(x) = c(F_1(x_1|\omega_i), \dots, (F_d(x_d|\omega_i)|\omega_i) \prod_{k=1}^d f_k(x_k|\omega_i) \mathbb{P}(\omega_i).$$

Všimněme si, že funkce  $g_i$  se nyní skládá ze tří ingrediencí: podmíněné hustoty kopule  $c(\cdot|\omega_i)$ , podmíněné hustoty jednorozměrných marginálů  $f_1(\cdot|\omega_i), \dots, f_d(\cdot|\omega_i)$  a apriorní pravděpodobnosti tříd  $\mathbb{P}(\omega_i)$ . Chceme-li tedy vytvořit *bayesovský klasifikátor založený na kopulích*, je potřeba pro každou třídu v datech určit vždy tyto tři ingredience. Více detailů k tomuto tématu čtenář nalezne v Górecki at al. (2016).

Výhoda těchto klasifikátorů je ve větší flexibilitě vytvořených modelů, což je umožněno právě využitím kopulí, jak lze vidět na Obrázek 11. Zatímco vlevo lze vidět model daný Naivním Bayesem, který příliš neodpovídá roztaženosti dat po vedlejší diagonále, což je způsobeno předpokladem nezávislosti vstupních atributů, vpravo je zobrazen model daný bayesovským klasifikátorem založeným na gaussovských kopulích, u něhož lze vidět daleko lepší přizpůsobení se datům (distribuční funkce jsou nakloněny/roztaženy ve směru vedlejší diagonály – zvláště pro zelené a modré objekty), z čehož poté často plyne i lepší výkonnost těchto klasifikátorů, jak bude ilustrováno níže.



**Obrázek 11: (vlevo) Model daný Naivním Bayesem. (vpravo) Model daný bayesovským klasifikátorem založeným na gaussovských kopulích.**

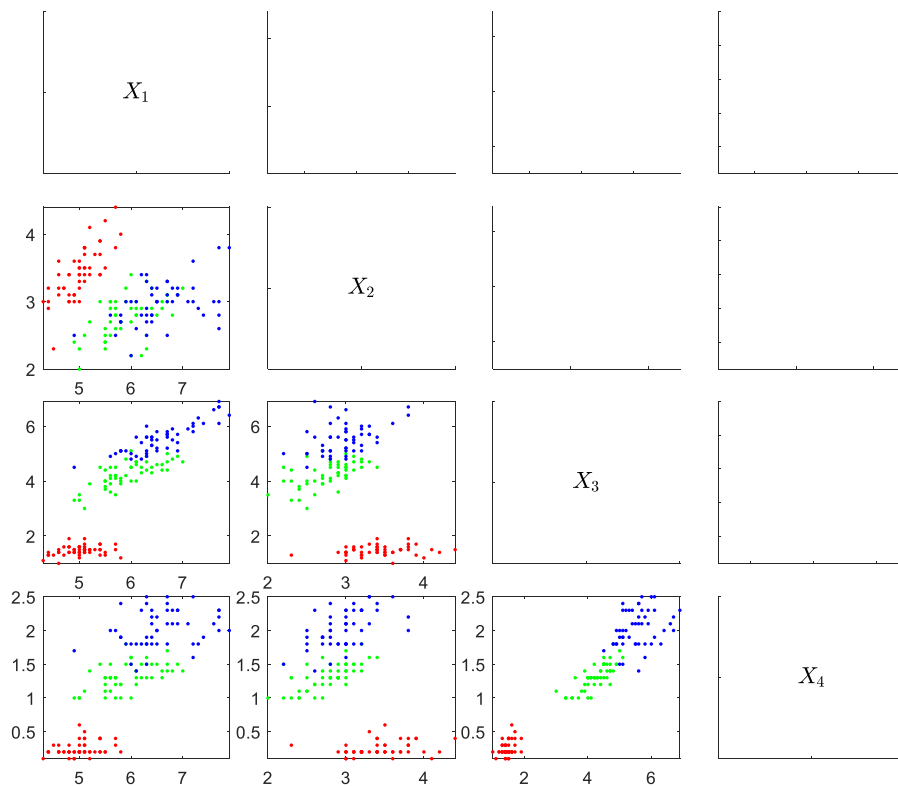
### KONTROLNÍ OTÁZKA



1. Jakou hrají kopule roli v Bayesovských klasifikátorech založených na kopulích?
2. Co v nich kopule přinášejí oproti Naivnímu Bayesovi?

## 1.5 Ilustrace bayesovských klasifikátorů založených na kopulích pro vícerozměrná data

V této části se podíváme, jak model daný *bayesovským klasifikátorem založeným na kopulích* (zkráceně BKZK) vypadá pro data obsahující 4 vstupní atributy, tedy  $d = 4$ . Jako data nám poslouží známá datová sada Iris volně dostupná z databáze University of Carolina, Irvine (<https://archive.ics.uci.edu/ml/datasets/Iris>). Tato data obsahují tři třídy, tedy  $m = 3$ , každá obsahující padesát kosatců (druh květin) a tyto tři třídy se jmenují  $\omega_1 = \text{Setosa}$ ,  $\omega_2 = \text{Versicolour}$   $\omega_3 = \text{Virginica}$ . Vstupní atributy odpovídají délkám a šířkám dvou typů listů (sepální a petální list) těchto květin, podle kterých se daný kosatec zařazuje do jedné z uvedených tří tříd. Data jsou zobrazeny na Obrázek 12.

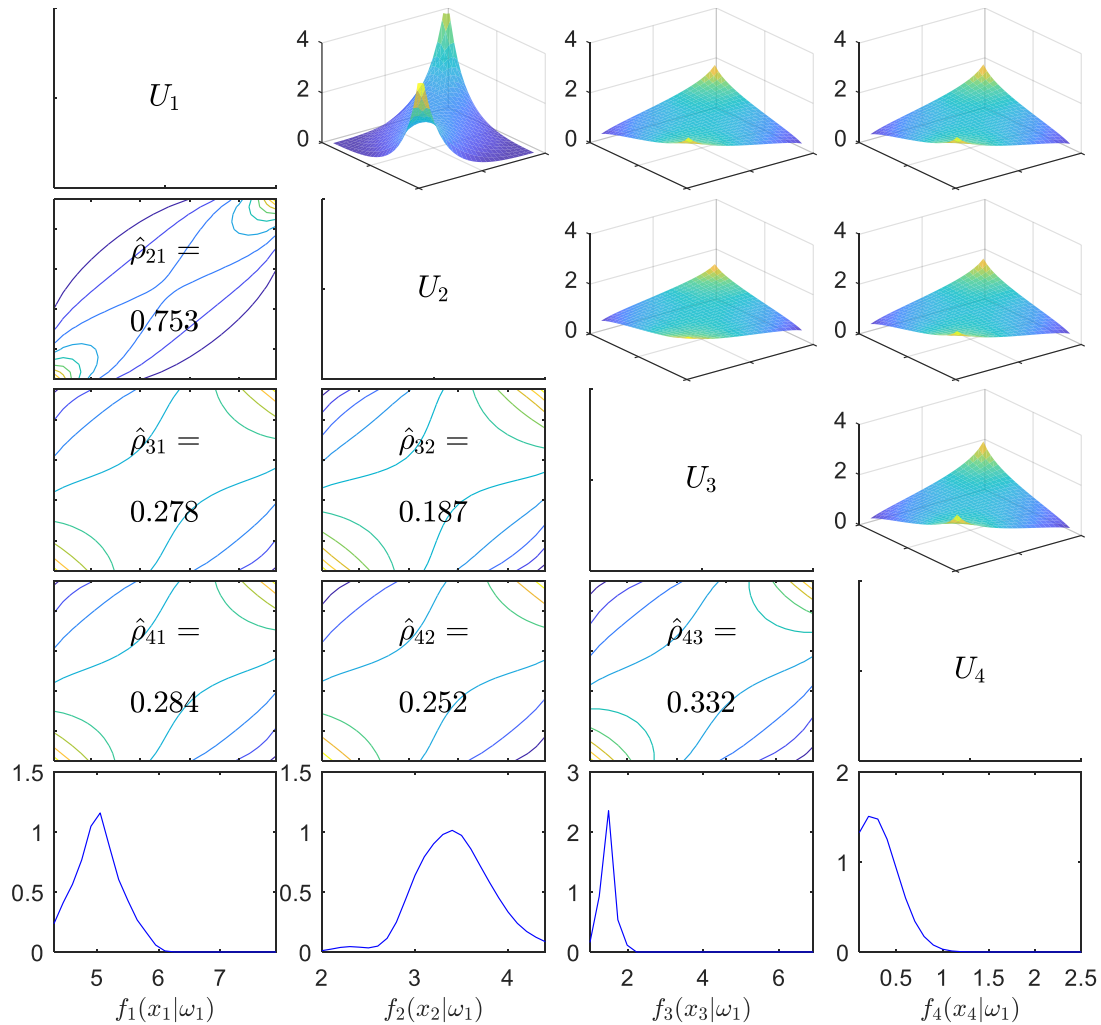


**Obrázek 12:** Data Iris se 4 vstupními atributy  $X_1, \dots, X_4$ . Data jsou zobrazeny zvlášť pro každou dvojici vstupních atributů, např. nejlépe nahoře jsou data odpovídající páru  $X_1$  a  $X_2$ . Třídám  $\omega_1, \omega_2$  a  $\omega_3$  pak odpovídá postupně červená, zelená a modrá barva.

### 1.5.1 MODEL S GAUSSOVSKOU KOPULÍ

Při určování modelu pro tato data BKZK počítá odhad výše zmíněných tří ingrediencí (apriorní pravděpodobnost třídy, kopuli a jednorozměrné marginály) pro každou třídu v datech. Výpočet odhadu apriorních pravděpodobností již byl popsán v Kapitole 1.2, takže jen zopakujeme, že je to  $\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = \mathbb{P}(\omega_3) = \frac{50}{150} = \frac{1}{3}$  (každá třída v datech je zastoupena 50ti objekty). Nyní pro každou třídu postupně odhadneme dvě zbývající ingredience - kopuli a

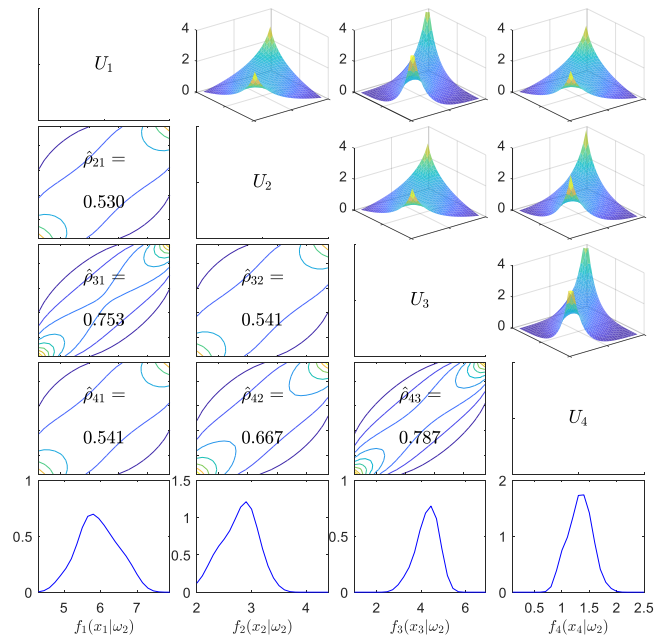
jednorozměrné marginály. Jednorozměrné marginály budeme odhadovat pomocí již dříve zmíněné metody hladkého jádrového odhadu (kernel smooth estimation). Kopuli budeme vybírat z rodiny *gaussovských* kopulí a její parametry určíme pomocí metody maximální věrohodnosti. Všechny výpočty jsou realizovány v systému MATLAB (hladké jádrové odhady funkcí ksdensity a odhad kopule funkcí copulafit). Výsledný odhad kopule a jednorozměrných marginálů je postupně zobrazen pro všechny tři třídy  $\omega_1$ ,  $\omega_2$  a  $\omega_3$  na Obrázek 13, Obrázek 14 a Obrázek 15. Jen dodejme, že proměnné  $U_1, \dots, U_4$  v těchto obrázcích odpovídají transformovaným vstupním atributům  $X_1, \dots, X_4$ , kde transformace, který je závislá na odhadech jednorozměrných marginálů, je daná  $U_i = \bar{F}_i(X_i)$  pro všechna  $i = 1, \dots, 4$ .



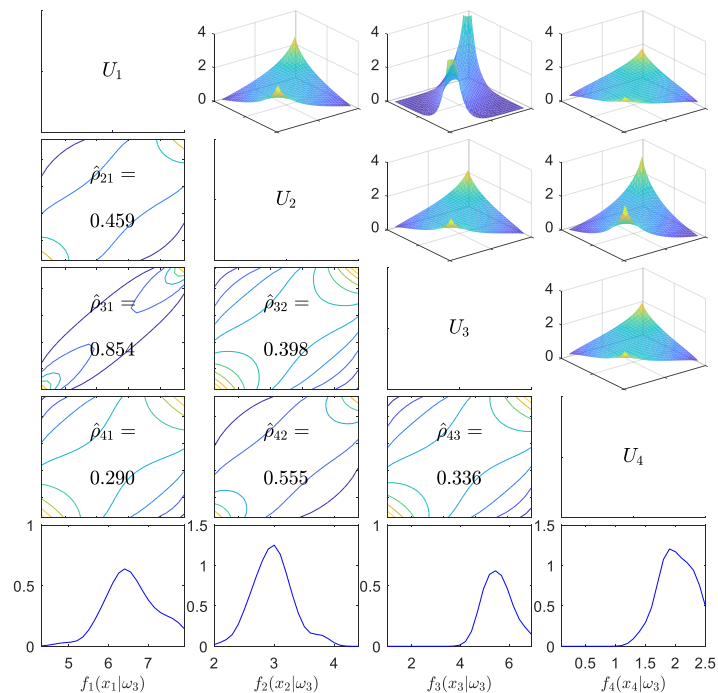
**Obrázek 13: Gaussovská čtyřrozměrná kopule a jednorozměrné marginály pro (červenou) třídu  $\omega_1 = \text{Setosa}$ . Hustoty jednorozměrných marginálů jsou zobrazeny úplně dole. Odhady parametrů ( $\rho$ ) kopule jsou zobrazeny pod diagonálou. Nad diagonálou jsou zobrazeny všechny dvourozměrné marginály odhadnuté čtyřrozměrné kopule. Pod diagonálou jsou navíc ještě zobrazeny kontury těchto dvourozměrných marginálů.**

Jak lze z obrázku vidět, pro každý pár transformovaných vstupních atributů je odhadnutá hodnota korelačního parametru nenulová, např. pro pár  $U_1$  a  $U_2$  je to  $\hat{\rho}_{21} = 0,753$ , což odpovídá

velmi silné závislosti mezi vstupními atributy  $X_1$  a  $X_2$ . Předpoklad nezávislosti, který požaduje Naivní Bayes, by zde tedy byl zcela nevhodný.



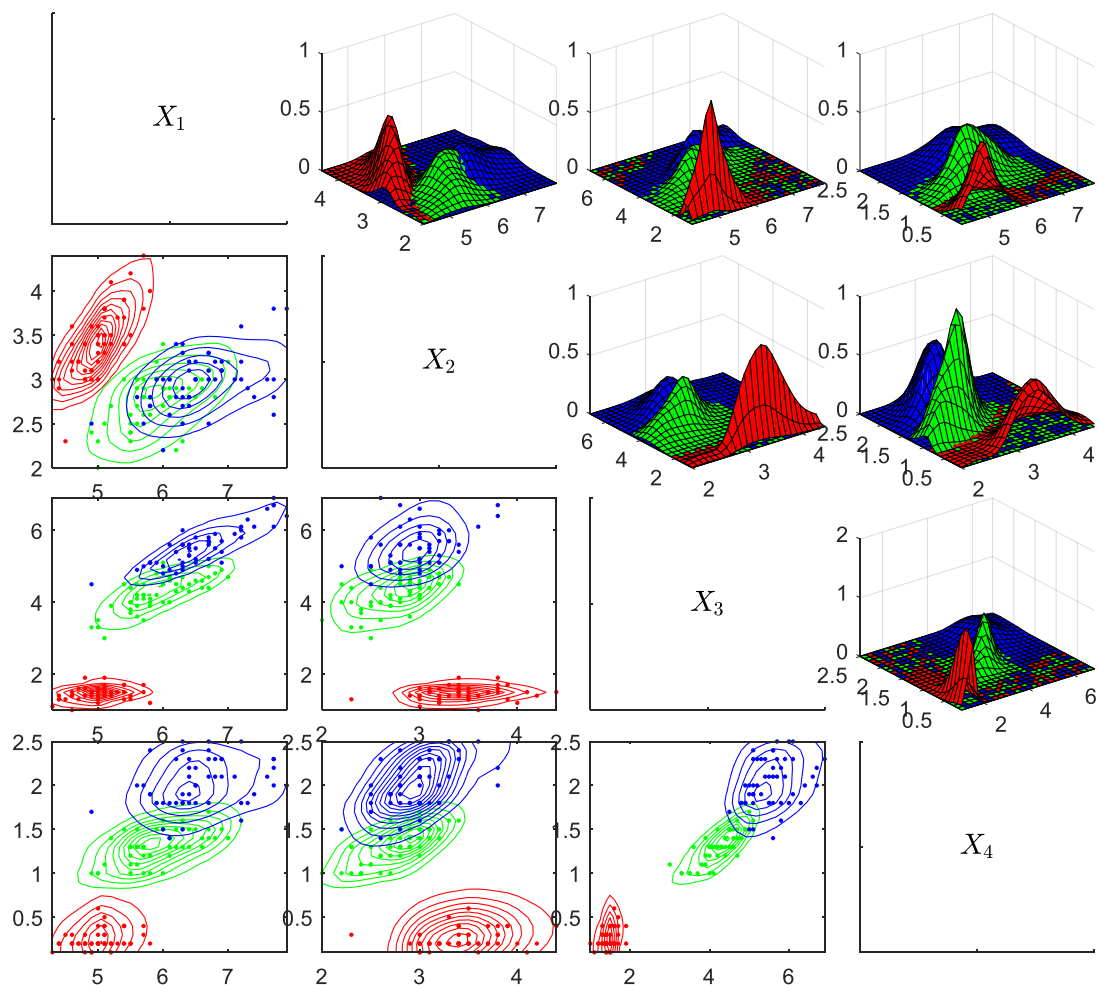
**Obrázek 14: Gaussovská čtyřrozměrná kopule a jednorozměrné marginály pro (zelenou) třídu  $\omega_2 = \text{Versicolour}$ . Zobrazené veličiny mají stejný význam jako na Obrázek 13.**



**Obrázek 15: Gaussovská čtyřrozměrná kopule a jednorozměrné marginály pro (zelenou) třídu  $\omega_3 = \text{Virginica}$ . Zobrazené veličiny mají stejný význam jako na Obrázek 13.**

Na základě všech tří ingrediencí pro všechny tři třídy lze již pak zobrazit celý model klasifikátoru, tak jak je vidět na Obrázek 16.

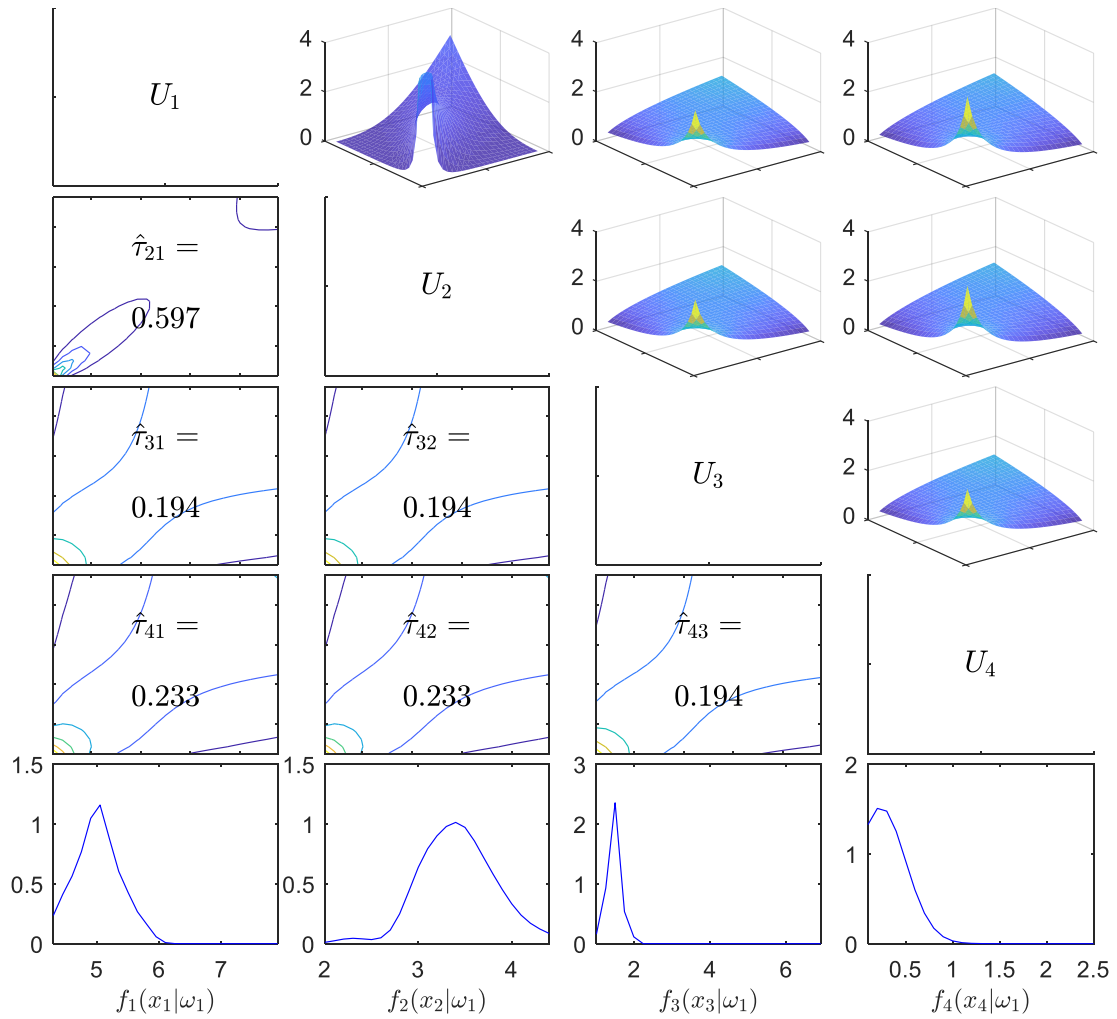




**Obrázek 16: Dvourozměrné marginály diskriminačních funkcí  $g_1$  (červená)  $g_2$  (zelená) a  $g_3$  (modrá) (nad diagonálou zobrazeny přímo, pod diagonálou jejich kontury) daná modely na Obrázek 13, Obrázek 14 a Obrázek 15.**

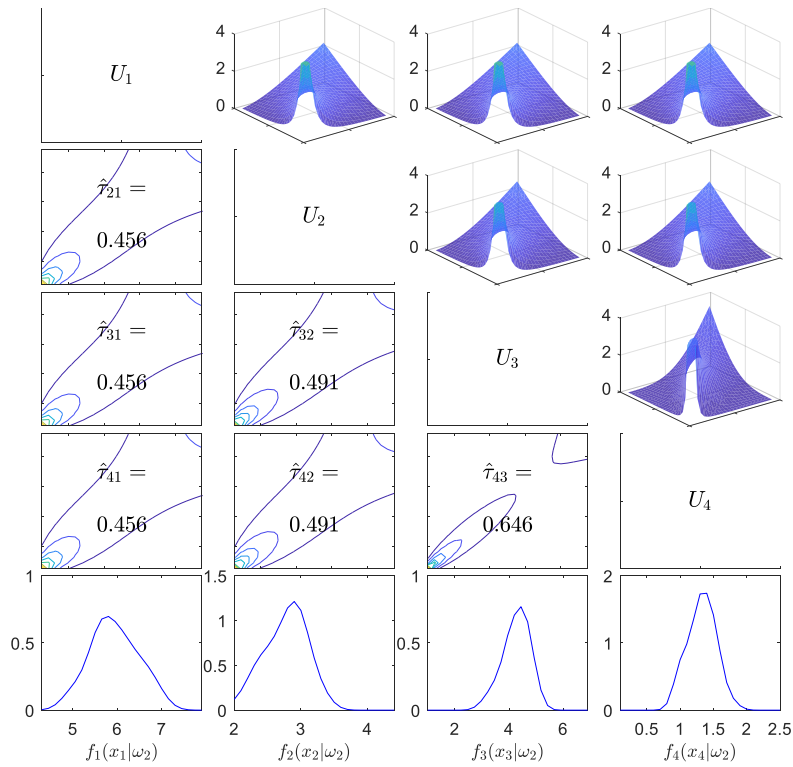
### 1.5.2 MODEL S HIERARCHICKOU ARCHIMEDOVSKOU KOPULÍ

Nyní si ukažme model získaný BKZK na stejných datech (Iris) a s použitím stejných postupů jako v předchozí kapitole, avšak za předpokladu, že kopule bude vybírána z rodiny Claytonových hierarchických archimedovských kopulí. *Hierarchická archimedovská kopule* vznikne tak, že se více archimedovských kopulí vnoří do sebe (jako vnoření více funkcí do sebe je myšleno např.  $f(g(x))$ , kde  $f$  a  $g$  jsou dvě funkce)). Dodejme, že, jelikož funkce pro odhad hierarchických archimedovských kopulí není v systému MATLAB implementována, využili jsme k tomuto účelu balíček HACopula vyvinutý na Slezské univerzitě, který je volně dostupný na GitHubu (<https://github.com/gorecki/HACopula>). Výsledný odhad kopule a jednorozměrných marginálů pro všechny třídy v datech je zobrazen na Obrázek 17, Obrázek 18 a Obrázek 20 Obrázek 19.

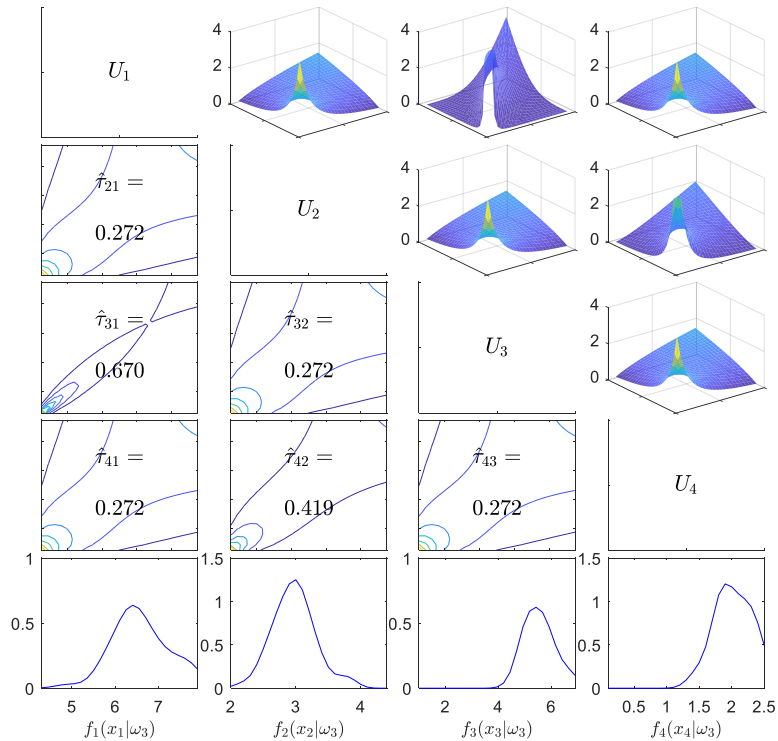


**Obrázek 17: Claytonova čtyřrozměrná hierarchická archimedovská kopule a jednorozměrné marginály pro (červenou) třídu  $\omega_1 = \text{Setosa}$ . Zobrazené veličiny mají stejný význam jako na Obrázek 13.**

Z obrázku je jasně vidět rozdíl v typu kopule. Na rozdíl od gaussovské kopule je totiž Claytonova kopule asymetrická vzhledem k vedlejší diagonále – pro všechny páry je to víc než evidentní – daleko větší hustota v levém dolním rohu než v pravém horním. Na dalších dvou obrázcích je zobrazena kopule a jednorozměrné marginály pro zbylé dvě třídy.

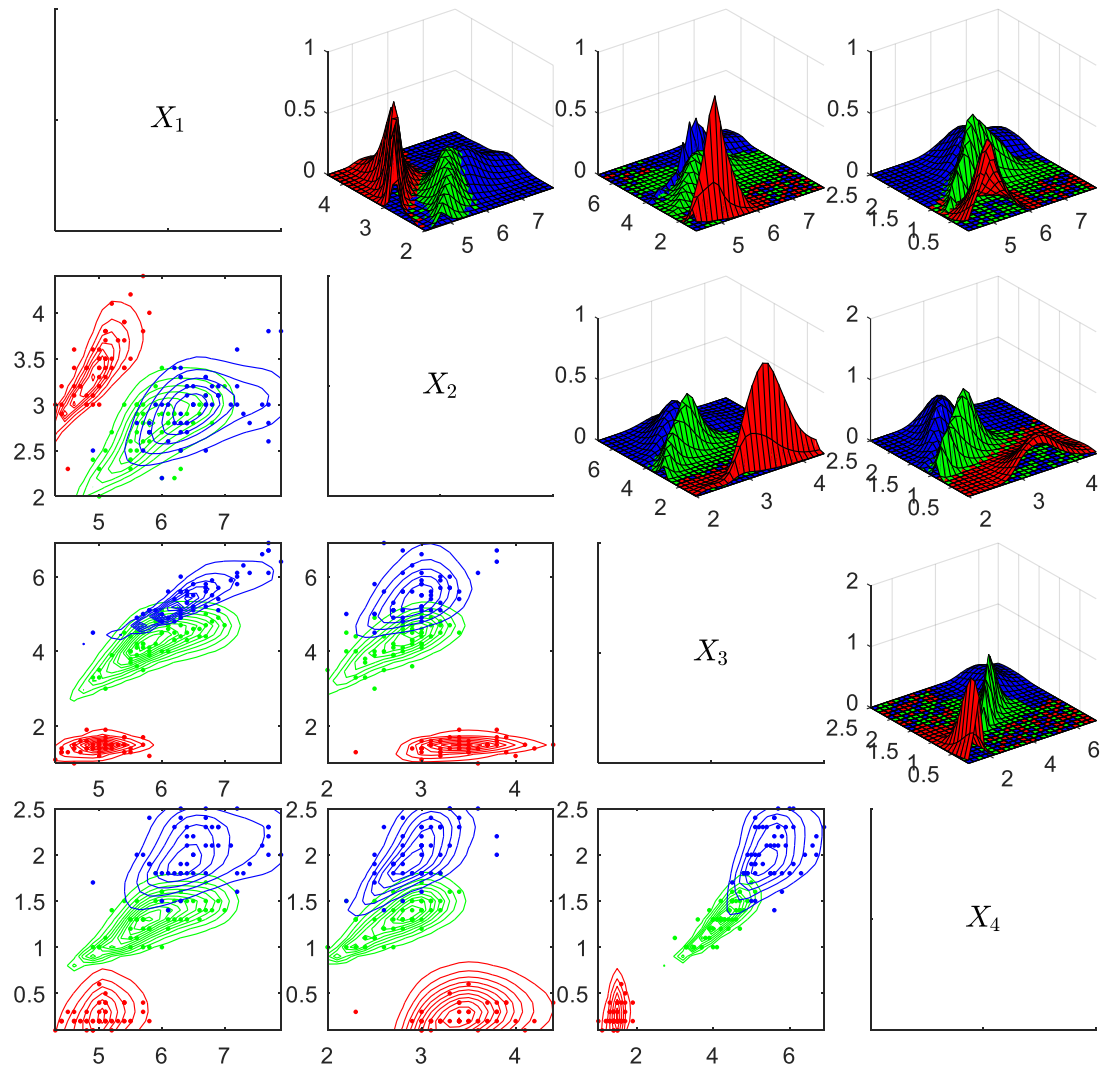


**Obrázek 18:** Claytonova čtyřrozměrná hierarchická archimedovská kopule a jednorozměrné marginály pro (zelenou) třídu  $\omega_2 = \text{Versicolour}$ . Zobrazené veličiny mají stejný význam jako na Obrázek 13.



**Obrázek 19:** Claytonova čtyřrozměrná hierarchická archimedovská kopule a jednorozměrné marginály pro (modrou) třídu  $\omega_3 = \text{Virginia}$ . Zobrazené veličiny mají stejný význam jako na Obrázek 13.

Stejně jako u předchozího příkladu využívajícího gaussovské kopule (Kapitola 1.5.1) lze na základě všech tří ingrediencí pro všechny tři třídy zobrazit celý model klasifikátoru, tak jak je vidět na Obrázek 20.



**Obrázek 20: Dvourozměrné marginály diskriminačních funkcí  $g_1$  (červená)  $g_2$  (zelená) a  $g_3$  (modrá) (nad diagonálou zobrazeny přímo, pod diagonálou jejich kontury) daná modely na Obrázek 17, Obrázek 18 a Obrázek 20 Obrázek 19.**

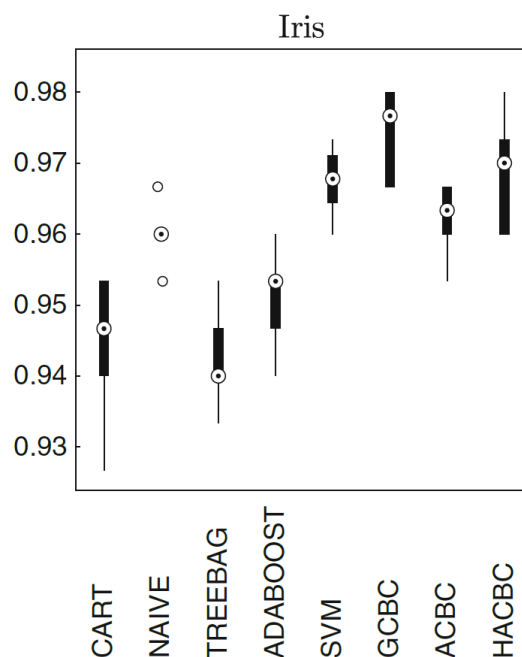
Vzhledem k již výše zmíněné asymetrii Claytonových kopulí lze nyní pozorovat hruškovitý tvar kontur diskriminačních funkcí, kde špička hrušky je vlevo dole a vpravo nahoře je široká část hrušky. Volba rodiny kopulí použitých v BKZK tedy vnáší další možnosti pro flexibilitu bayesovských klasifikačních modelů.

Díky takto flexibilním modelům můžou BKZK v klasifikační výkonnosti jak překonat Naivní Bayes, tak konkurovat i velmi složitým klasifikátorům, jako jsou např. stroje opěrných vektorů (support vector machines) či náhodné lesy, jak je ilustrováno v následující kapitole.

## 1.6 Výkonnost bayesovského klasifikátoru založeného na kopulích

V této části porovnáme výkonnost BKZK využívající různé rodiny kopulí s dalšími běžně používanými klasifikátory, jako jsou klasifikační a regresní stromy (zkratka CART), náhodné lesy založené na baggingu (TREEBAG) a boostingu (ADABOOST), stroje opěrných vektorů (SVM) a také do tohoto porovnání zahrneme Naivní Bayes (NAIVE). Z BKZK zahrneme BKZK využívající gaussovské kopule (značeno GCBC), archimedovské kopule (ACBC) a hierarchické archimedovské kopule (HACBC) – značení je ve shodě s článkem Górecki et al. (2016), ze kterého jsou všechny výsledky převzaty.

Výkonnost těchto osmi klasifikátorů budeme měřit tzv. *přesností* klasifikátoru, což je poměr počtu správně klasifikovaných objektů ke všem klasifikovaným objektům (tedy nějaká hodnota z intervalu  $[0, 1]$ , která se považuje za lepší, čím je vyšší). Toto měření přesnosti provedeme pomocí tzv. *10ti-násobné křížové validace*, kdy jsou data (opět použijeme data Iris, nám již známa z předchozích kapitol) náhodně rozdělena na 10 částí, a 10krát se opakuje proces, kdy 9 částí se použije pro vytvoření klasifikačního modelu a zbývající část se použije pro změření přesnosti (porovná se, kolikrát model zařadil objekt do třídy, do které skutečně objekt patřil). Získaných 10 hodnot přesnosti se poté průměruje. Jelikož tento proces zahrnuje prvek náhody, vše zopakujeme zase 10 krát, tzn. naše výsledky obsahují 10 hodnot přesnosti pro každý z osmi zvažovaných klasifikátorů. Tyto výsledky jsou zobrazeny pomocí boxplotu na Obrázek 21. Poznamenejme, že kolečko s tečkou uprostřed značí v medián z naměřených 10ti hodnot.



**Obrázek 21: Přesnost osmi klasifikátorů pro data Iris (převzato z Górecki et al. (2016)).**

Jak lze vidět, vzhledem k mediánu naměřených přesností se dva typy BKZK, konkrétně GCBC a HACBC, umístili na prvních dvou místech. Pro Naivní Bayes (NAIVE), tedy BKZK, který vlastně kopule nevyužívá (předpoklad nezávislosti), je medián přesnosti až na pátém místě. Toto měření tedy potvrzuje, že BKZK 1) mohou překonat i složitější klasifikátory jako

SVM nebo náhodné lesy (ADABOOST a TREEBAG) a 2) zakomponováním kopule místo předpokladu nezávislosti vstupních atributů (obsažené v Naivním Bayesovi) lze přesnost klasifikátoru zvýšit. Pro více detailů k tomuto porovnání odkážeme laskavého čtenáře na článek Górecki et al. (2016).



## SHRNUTÍ KAPITOLY

V této kapitole jsme si představili bayesovskou klasifikaci. Nejdříve jsme se seznámili s jednoduchým přístupem k tomuto typu klasifikace skrze Naivní Bayes a poté jsme se seznámili s poměrně nedávno navrženým bayesovským klasifikátorem založeným na kopula funkcích, což proběhlo včetně seznámení se s kopula funkcemi a několika jejich významnými parametrickými rodinami. Na závěr jsme provedli porovnání výše zmíněných klasifikátorů s dalšími typy často používaných klasifikátorů, ze kterého vyplývá praktická použitelnost představených bayesovských klasifikátorů založených na kopulích.



## OTÁZKY

1. Co je to klasifikační metoda a jak vypadají klasifikační data?
2. Jak pracuje bayesovský klasifikátor?
3. Co je to diskriminační funkce?
4. Jakým způsobem se v ní objeví apriorní pravděpodobnost třídy a jak se její hodnota odhaduje?
5. Na jakém předpokladu stojí Naivní Bayes?
6. Kterou část diskriminační funkce tento předpoklad ovlivňuje?
7. Co je to kopule?
8. Co říká Sklarova věta?
9. Jaké jsou příklady rodin kopulí a co určuje jejich vlastnosti?
10. Z čeho se skládá model diskriminační funkce v bayesovském klasifikátoru založeném na kopulích?



## ODPOVĚDI

1. Je to metoda, která na základě klasifikačních dat vytvoří model, na základě kterého lze poté automatizovaně klasifikovat nové objekty, u kterých třída není známa. Klasifikační data jsou data, ve kterých je každý objekt popsán vstupními atributy a je pro něho známa jeho třída.
2. Bayesovský klasifikátor zařazuje objekty do třídy na základě hodnot všech diskriminačních funkcí pro daný objekt a to tak, že objekt zařadí do třídy odpovídající nejvyšší z těchto hodnot.

3. Diskriminační funkce pro třídu  $\omega_i$  a pro daný objekt  $x$  odpovídá  $\mathbb{P}(\omega_i|x)$ , tedy pravděpodobnosti, že daný objekt patří do třídy  $\omega_i$ .
  4. Apriorní pravděpodobnost třídy  $\mathbb{P}(\omega_i)$  se v zápisu diskriminační funkce objeví po jejím přepisu pomocí známého Bayesova pravidla. Její hodnota se tradičně odhaduje jako relativní zastoupení třídy  $\omega_i$  v datech.
  5. Na předpokladu nezávislosti vstupních atributů.
  6. Zásadně se zjednoduší výpočet sdružené hustoty  $f(x|\omega_i)$ .
  7. Kopule je sdružená distribuční funkce mající všechny své jednorozměrné marginály stejnoměrně rozdělené na jednotkovém intervalu.
  8. Sklarova věta říká, že každou obecnou sdruženou distribuční funkci lze rozložit na kopuli a jednorozměrné marginály. Pokud jsou tyto marginály spojitě, je tento rozklad jednoznačný.
  9. Např. eliptické (gaussovské a studentovské) a archimedovské (Claytonovy, Frankovy nebo Gumbelovy) rodiny. Vlastnosti jednotlivých kopulí v těchto rodinách pak jejich udávají parametry.
  10. Z kopule, jednorozměrných marginálů a apriorní pravděpodobnosti třídy.
-