Téma 47

Zadání:

Kybernetika, teorie systémů a umělá inteligence. Základy teorie informace: základní pojmy, signál, kódování, informace. Sdružená a podmíněná entropie a její vlastnosti. Střední vzájemná informace. Komunikační kanál a jeho kapacita. Kódy a kódování. Princip maxima entropie. Rozhodování za neurčitosti a rizika, statistické a Bayesovské rozhodování. Základy teorie her, pravidlo minimaxu.

1 Teorie informace

Zpráva - jakákoliv posloupnost rozlišitelných znaků

Symboly - rozlišitelné prvky ve zprávě (grafické znázornění - *znaky*)

Abeceda - množina všech symbolů

Příklad: zpráva: $a\ b\ c\ c\ a\ b\ d\ a\ b\ d\ d\ c\ b\ a\ c$
délka zprávy: $n = 15$
abeceda: $A = \{a, b, c, d\}$
počet symbolů abecedy: $s = 4$

Signál - materiální nositel zprávy

Kódování - transformace zprávy vyjádřené pomocí jedné abecedy na zprávu vyjádřenou pomocí druhé abecedy

Informace - vztah mezi symboly zprávy a okolním světem (existuje více definic informace)

2 Informace a Entropie

Počet možných zpráv N délky n nad abecedou s celkovým počtem symbolů s je

$$N = s^n \tag{1}$$

Příklad:
$A = \{0, 1\}$, $s = 2$, $n = 6 => N = 2^6 = 64$
$A = \{a, b, c, d\}$, $s = 5$, $n = 6 => N = 5^6 = 15625$

2.1 Hartleyova míra informace

Vybereme-li jednu konkrétní zprávu ze všech možných kombinací (při dané abecedě a počtu znaků zprávy), redukujeme tím neurčitost. Čím větší bude počet možných alternativ (větší počet znaků zprávy a více symbolů abecedy), tím větší neurčitost tímto výběrem odstraníme. Množství informace ve zprávě (je rostoucí funkcí počtu alternativ) závisí na velikosti odstraněné neurčitosti výběrem zprávy ze všech možných alternativ. (tzn. čím delší je zpráva a čím více je znaků zprávy, tím je ve zprávě obsaženo větší množství informace.)
Hartleyovu míru informace vypočteme jako

$$I = K \ln N = Kn \ln s \tag{2}$$

2.2 Entropie

Obecně je entropie definována jako míra neurčitosti systému.

Výskyty symbolů ve zprávě mohou mít různou pravděpodobnost a proto se množství informace určuje jako velikost Entropie (střední hodnota informace na jeden symbol zprávy)

$$E = \frac{I}{n} = -K \sum_{i=1}^{s} p_i \ln p_i \tag{3}$$

kde konstanta $K = \frac{1}{\ln 2}$ (tedy např. $A = \{0, 1\}$, $n = 1 => E = K(\frac{1}{2} \ln \frac{1}{2} + \frac{1}{2} \ln \frac{1}{2}) = Kln2 = 1$.

Rovnice (3) se dá po dosazení K a za použití faktu, že $\frac{\ln x}{\ln y} = \log_y x$ upravit na

$$E = -\sum_{i=1}^{s} p_i \log_s p_i \quad [bits/symbol] \tag{4}$$

2.2.1 Vlastnosti entropie

1. $H(X) = 0$ tehdy a jen tehdy, jsou-li všechny pravděpodobnosti kromě jedné rovny nule a jedna pravděpodobnost rovna jedné.

2. Entropie dosahuje svého maxima, jsou-li všechny pravděpodobnosti stejné ($P_1 = P_2 = ... = P_s = \frac{1}{s}$)

3. Je spojitou funkcí pravděpodobnosti na intervalu $0 \leq P \leq 1$

Všimněme si, že $I = n \cdot H_{max} = n \cdot \log_2 s$ tedy Hartleyova míra informace udává maximální množství informace, které můžeme přenést n symboly z abecedy o s symbolech.

Příklad:
Jestliže X může nabývat hodnot 0 a 1, entropie X je definována jako $H(X) = -P(X = 0) \log_2 P(X = 0) - P(X = 1) \log_2 P(X = 1)$ Pravděpodobnost že X=0 se dá vyjádřit $P(X = 0) = 1 - P(X = 1)$ a tedy entropie se dá brát jako funkce jedné proměné P(X=1) (viz.2.2.1).

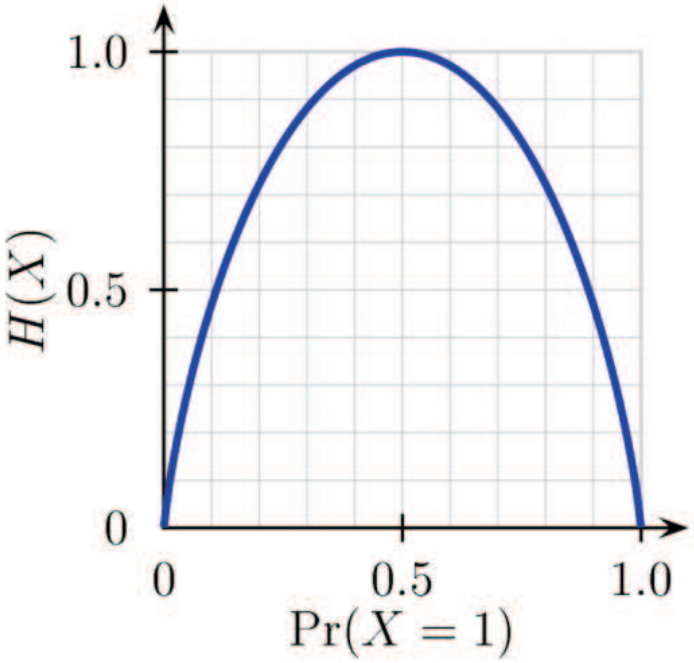


Figure 1: Funkce entropie v závislosti na $P(X = 1)$.

2.3 Sdružená a podmíněná pravděpodobnost

Sdružená pravděpodobnost dvou a více jevů $P(X_1, X_2, ..., X_n)$ je pravděpodobnost, že dané jevy nastanou zároveň. Marginální pravděpodobnost (pravděpodobnost jen jednoho z nich) se pak dá vyjádřit jako $P(X_i) = \sum_j P(X_i, X_j)$ (pro dva jevy).

Podmíněná pravděpodobnost dvou jevů $P(X|Y)$ je pravděpodobnost, že při jevu Y nastane i jev X. Z toho vyplývá, že při nezávislosti obou jevů je $P(X|Y) = P(X)$.

Sdruženou pravděpodobnost lze vyjádřit též jako

$$P(X, Y) = P(X) \cdot P(Y|X) \tag{5}$$

Příklad 1:
Mějme dva závislé jevy X, jehož výsledkem může být prvek množiny $\{1, 2, 3\}$, a Y, jehož výsledkem může být prvek množiny $\{a, b\}$. Pravděpodobnosti výsledků daných jevů můžeme vidět v následující tabulce.

Table 1: Tabulka pravděpodobností závislých jevů X a Y

Y/X	1	2	3
a	0,1	0,2	0,3
b	0,2	0,1	0,1

Příklady sdružených pr.: $P(1, a) = 0,1$; $P(2, b) = 0,1$; $P(3, a) = 0,3$
Příklady podmíněné pr.: $P(a|1) = \frac{P(1,a)}{P(1)} = \frac{0,1}{0,1+0,2} = \frac{1}{3}$; $P(3|b) = \frac{0,1}{0,2+0,1+0,1} = 0,25$

Příklad 2:
Mějme stejný příklad jako Příklad 2, ale jevy X a Y jsou nezávislé.

Table 2: Tabulka pravděpodobností nezávislých jevů X a Y

	X		Y
1	0,1	a	0,9
2	0,3	b	0,1
3	0,6		

Příklady sdružených pr.: $P(1,a) = 0,1 \cdot 0,9 = 0,09$; $P(2,b) = 0,3 \cdot 0,1 = 0,03$; $P(3,a) = 0,6 \cdot 0,9 = 0,54$
Příklady podmíněné pr.: $P(a|1) = P(a) = 0,9$; $P(3|b) = P(3) = 0,6$

2.4 Sdružená entropie

Mějme dva signály $x_i = \{x_1, ..., x_s\}$ a $y_j = \{y_1, ..., y_r\}$ pak sdružená entropie těchto signálů je

$$H(X,Y) = \sum_{i=1}^{s}\sum_{j=1}^{r} P(x_i, y_j) \log_2 P(x_i, y_j), \tag{6}$$

kde $P(x_i, y_j)$ je sdružená pravděpodobnost (pravděpodobnost výskytu obou znaků zároveň). Zobecněním dostaneme vztah:

$$H(X_1, X_2, ..., X_q) = \sum_{i=1}^{s}\sum_{j=1}^{r}...\sum_{k=1}^{m} P(x_{1i}, x_{2j}, ..., x_{qk}) \log_2 P(x_{1i}, x_{2j}, ..., x_{qk}) \tag{7}$$

Pro sdruženou entropii platí nerovnost

$$H(X,Y) \leq H(X) + H(Y). \tag{8}$$

2.5 Podmíněná entropie

Podmíněná entropie zdroje signálu Y daná zdrojem signálu $X = x_i$ je definována

$$H(Y|X = x_i) = -\sum_{j=1}^{s} P(y_j|x_i) \log_2 P(y_j|x_i) P(y_j|x_i). \tag{9}$$

Podmíněná entropie Y při daném X (psáno $H(Y|X)$), je pak definována jako vážený průměr $H(Y|X = x_i)$, tj.

$$H(Y|X) = \sum_i P(x_i) H(Y|X = x_i) = -\sum_i P(x_i) \sum_j P(y_j|x_i) \log_2 P(y_j|x_i). \tag{10}$$

Při použití (5) lze vztah zjednodušit na

$$H(Y|X) = -\sum_i \sum_j P(y_j, x_i) \log_2 P(y_j|x_i). \tag{11}$$

Vztah mezi sdruženou a podmíněnou entropii (odvození viz.[1])je

$$H(X,Y) = H(Y) + H(X|Y). \tag{12}$$

2.6 Entropie spojité veličiny (diferenciální entropie)

$$H(X) = \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx, \tag{13}$$

kde $f(x)$ je hustota pravděpodobnosti spojité veličiny.

Zatímco entropie diskrétní veličiny je absolutní mírou neurčitosti, ve spojité verzi je entropie relativní mírou neurčitosti vzhledem ke zvolenému systému souřadnic.

2.6.1 Sdružená entropie spojitých veličin

I pro spojitou veličinu definujeme sdruženou entropii

$$H(X) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \log_2 f(x,y) dx dy, \tag{14}$$

2.6.2 Podmíněná entropie spojitých veličin

I pro spojitou veličinu definujeme sdruženou entropii

$$H(X) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \log_2 \frac{f(x,y)}{g(x)} dx dy \tag{15}$$

$$H(X) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \log_2 \frac{f(x,y)}{h(x)} dx dy, \tag{16}$$

kde $f(x)$ je sdružená hustota pravděpodobnosti veličin x, y a $g(x), h(y)$ jsou marginální hustoty pravděpodobnosti veličin x a y, pro které platí

$$g(x) = \int_{-infty}^{infty} f(x,y) dy \tag{17}$$

$$h(y) = \int_{-infty}^{infty} f(x,y) dx \tag{18}$$

[1] Odvození $H(X,Y) = H(Y) + H(X|Y)$:

$$H(X,Y) = -\sum_i \sum_j P(x_i, y_j) \log_2 P(x_i, y_j) = -\sum_i \sum_j P(x_i, y_j) \log_2 P(x_i) P(y_j|x_i) = -\sum_i \sum_j P(x_i, y_j) \log_2 P(x_i)$$

$$-\sum_i \sum_j P(x_i, y_j) \log_2 P(y_j|x_i) = \sum_i P(x_i) \log_2 P(x_i) + H(Y|X) = H(X) + H(Y|X)$$

2.7 Střední vzájemná informace

Předpokládejme, že máme dvě diskrétní náhodné veličiny, které jsou určitým způsobem závislé. Platí:

$$H(X) - H(X|Y) = T(X:Y), \tag{19}$$

kde $T(X:Y)$ je střední vzájemná informace (transinformace, transm....) Lze též psát $T(X:Y) = H(X) + H(Y) - H(X,Y)$

Vlastnosti střední vzájemné informace:

1. $T(X:Y) \geq 0$
2. $T(X:Y) = T(Y:X)$
3. $T(X:Y) = 0$ pro statisticky nezávislé veličiny, neboť $H(X,Y) = H(X) + H(Y)$
4. Platí-li $T(X:Y) = min\{H(X), H(Y)\}$, pak jsou a vázány funkční závislostí $y = F(x)$, kde $F(x)$ je monotónní funkce.

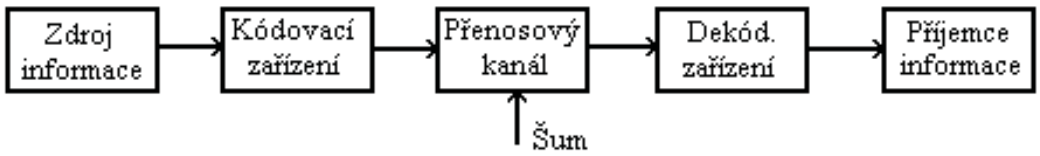
Normovaná střední vzájemná informace

$$t(X:Y) = \frac{T(X:Y)}{min\{H(X), H(Y)\}} \quad \in (0,1) \tag{20}$$

$$T(X:Y) = H(X) + H(X) - H(X,Y) = \sum_{i=1}^{s}\sum_{j=1}^{r} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_i)P(y_j)}. \tag{21}$$

3 Komunikační kanál a jeho kapacita

Struktura komunikačního kanálu:



3.1 Diskrétní kanál a jeho kapacita

$U = u_i$ množina přístupných hodnot signálu na vstupu kanálu
$Y = y_j$ množina přípustných hodnot signálu ma výstupu kanálu

Jsou-li U a Y konečná, pak trojice $(U, Y, [P(y_j/u_i)])$ resp. $(U, Y, [P(u_i/y_j)])$ představuje popis diskrétního kanálu, kde $[P(y_j/u_i)]$ je matice podmíněných pravděpodobností:

$$(P(y_i|u_j)) = \begin{pmatrix} P(y_1|u_1) & P(y_2|u_1) & ... & P(y_m|u_1) \\ P(y_1|u_2) & P(y_2|u_2) & ... & P(y_m|u_2) \\ ... & ... & & ... \\ P(y_1|u_n) & P(y_2|u_n) & ... & P(y_m|u_n) \end{pmatrix} \tag{22}$$

Obdobně lze vyjádřit matici $(P(u_i|y_j))$. Pomocí matic $(P(y_j|u_i))$ či $(P(u_i|y_j))$ lze porovnávat vlastnosti kanálů jen obtížně. K tomu lze použít střední vzájemnou informaci $T(U:Y) = H(U) - H(U|Y) = H(Y) - H(Y|U)$. Entropie $H(U|Y)$ resp. $H(Y|U)$ udává ztrátu informace způsobenou přenosem. Bude-li $H(Y|U) = H(U|Y) = 0$, je kanál bez šumu.

Rychlost přenosu informace:

$$R = v_u T(U:Y) \quad [bit/s], \tag{23}$$

kde v_u je rychlost přenosu jednotlivých symbolů:

$$v_u = \frac{1}{\bar{\tau_u}}, \tag{24}$$

kde $\bar{\tau_u}$ je střední doba přenosu jednoho symbolu. Rychlost fyzikálního přenosu v_u nemůžeme z hlediska teorie informace ovlivňovat. Střední vzájemná informace závisí nejen na pravděpodobnostech $P(y_j/u_i)$, ale také pravděpodobnostech $P(u_i)$. Tyto lze měnit vhodným kódováním a zvyšovat tak rychlost přenosu informace (kód se tak vlastně přizpůsobí vlastnostem kanálů).
Maximální rychlost přenosu C se nazývá **kapacita kanálu**:

$$C = v_u \mathbf{max}_{P(U)}(T(U:Y)), \tag{25}$$

kde maximum střední vzájemné informace hledáme přes všechna možná rozložení pravděpodobnosti P(U). Pro kanál bez šumu je T(U:Y) = H(U), tedy

$$C = v_u \mathbf{max}_{P(U)}(H(U)), \tag{26}$$

V případě rovnoměrného rozložení $P(u1) = P(u2) = ... = P(us) = 1/s$, kde s je počet možností vstupního signálu, dostáváme:

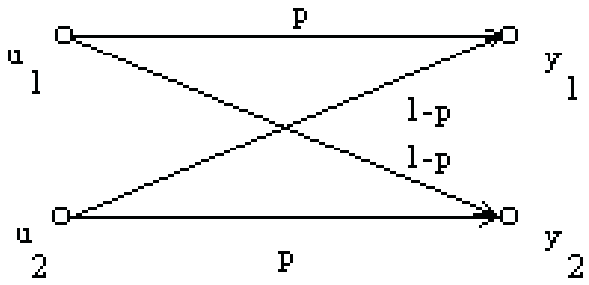
$$H(U) = \log_2 s$$

a tedy

$$C = v_u \log_2 s. \tag{27}$$

Symetrický kanál vzhledem ke vstupu (šum stejným způsobem ovlivní přenos každého vstupního symbolu) - pak jsou všechny řádky $(P(y_j|u_i))$ permutacemi čísel $P_1, P_2, ..., P_s$. Obdobně, pro symetrický kanál vzhledem k výstupu, platí předpoklady i pro sloupce matice $(P(y_j|u_i))$.
Kapacita symetrického kanálu je

$$C = v_u(\log_2 r + \sum_{j=1}^{r} P_j \log_2 P_j) \tag{28}$$



Binární symetrický kanál:

$$(P(y_i|u_j)) = \begin{pmatrix} P & 1-P \\ 1-P & P \end{pmatrix} \tag{29}$$

$$C = \log_2 + P\log_2 P + (1-P)\log_2(1-P) = 1 + P\log_2 P + (1-P)\log_2(1-P) \tag{30}$$

Shannon:
Jestliže je entropie zdroje menší než kapacita kanálu, je možné najít takový kód, který umožní přenášet daným kanálem zprávy, které generuje zdroj tak, aby pravděpodobnost chyby byla libovolně malá.

Obrácená věta:
Je-li entropie zdroje větší než kapacita kanálu, není možné najít takový kód, který by umožňoval přenášet tímto kanálem zprávy, které generuje zdroj tak, aby pravděpodobnost chyby byla libovolně malá.

3.2 Spojitý kanál a jeho kapacita

Je to obtížnější úloha, jednoduše řešitelná jen pro některé typy signálů. Zabývejme se tedy kapacitou kanálu při signálech stochastických s normálním rozložením (tyto mají největší entropii mezi signály se stejnou disperzí σ^2) Signál, který projde nf filtrem, nebo signál s různým rozložením, bude mít rozložení blízké normálnímu.
Předpokládejme, že signál i šum jsou ergodické náhodné procesy s nulovou střední hodnotou.

$$T(U:Y) = \frac{1}{2}\log_2(\frac{P+N}{N}) \tag{31}$$

Je-li f_h horní mezní frekvencí užitečného signálu, bude počet vzorků za jednotku času $2f_h$ a pro kapacitu kanálu bude platit

$$C = f_k \log_2(\frac{P+N}{N}) \tag{32}$$

Výraz

$$\frac{P+N}{N} = 1 + \frac{P}{N} = s \tag{33}$$

se interpretuje jako počet rozlišitelných úrovní výkonu signálu (= počet symbolů abecedy). Počet rozlišitelných amplitudových úrovní bude:

$$S_A = \frac{\sqrt{P+N}}{\sqrt{N}} \tag{34}$$

tj. poměr efektivní hodnoty součtu "signál + šum" k efektivní hodnotě šumu. (Pozor!! Závěry platí jen pro nezávislé normální a ergodické procesy!!)

4 Kódy a kódování

Kódování transformace zprávy z vyjádření v jedné abecedě do abecedy jiné

Dekódování inverzní operace ke kódování

Kód předpis, který určité skupině symbolů jedné abecedy jednoznačně přiřadí určitou skupinu symbolů z jiné abecedy

4.1 Redundance

Poměrná entropie:

$$h = \frac{H}{H_{max}} \tag{35}$$

Redundance:

$$r = 1 - h = \frac{H_{max} - H}{H_{max}} \cdot 100\% \tag{36}$$

Délku zprávy označme n, délku zprávy zakódované optimálním kódem s entropií H_{max} označme n_0, pak platí že

$$I = nH = n_0 H_{max} \tag{37}$$

a tedy

$$h = \frac{H}{H_{max}} = \frac{n_0}{n} \tag{38}$$

Redundance vzniká u *zdroje* (generované symboly nemají rovnoměrné rozdělení) nebo *kódováním* (nerovnoměrnost se ještě zvýší)

Redundance zdroje:

$$r_z = 1 - \frac{H_0}{H_{max}} = 1 - \frac{H_0}{\log_2 s} \tag{39}$$

kde H_0 je entropie zdroje a $\log_2 s$ je maximální entropie zdroje.

Redundance způsobená kódováním:

$$r_k = 1 - \frac{H_1}{H_0}, \tag{40}$$

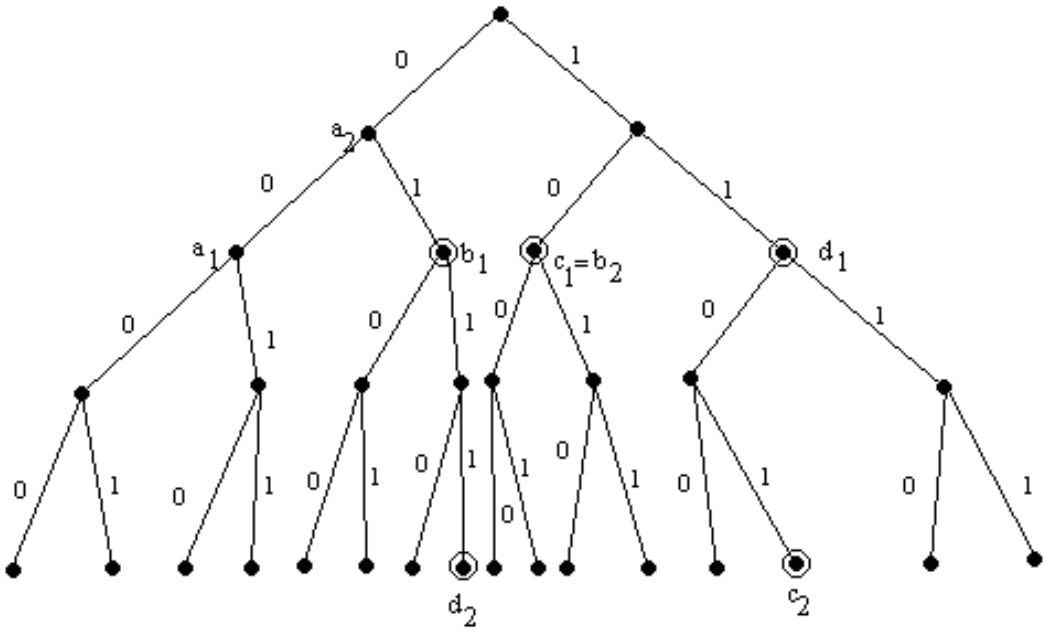
kde H_1 je entropie zakódované zprávy.

Celková redundance:

$$r_c = r_z + r_k - r_z r_k = 1 - \frac{H_1}{\log_2 s} \tag{41}$$

Je-li redundance rovna nule, pak je množství přenesené informace maximální jaké lze dosáhnout.

4.2 Prefixovost kódu



Pokud budou kódová slova uspořádána tak, že žádné slovo neleží na cestě od jiného slova ke kořeni grafu (žádné slovo není prefixem jiného), má uvedený kód prefixovou vlastnost.

Příklad:
Kód 1 není prefixový, kód 2 je.

symbol	a	b	c	d
kód 1	00	01	10	11
kód 2	0	10	1101	0111

4.3 Hammingova vzdálenost, detekce a korekce chyb

Hammingova vzdálenost ρ dvou kódových slov = počet míst, v nichž se kódová slova liší.

Příklad:

$\rho(000, 001) = \rho(000, 010) = 1$
$\rho(000, 101) = 2$
$\rho(100, 011) = 3$

Hammingova vzdálenost charakterizuje odolnost kódů proti poruchám a schopnost identifikovat, popř. opravit chyby. Pro kód bez redundance (vzdálenost dvou slov $\rho = 1$) nelze počítat s identifikací chyby. Pokud minimální vzdálenost dvou kódových slov je $\rho = 2$, lze zjišťovat chyby v jednom symbolu, při $\rho = 3$ je již možné opravovat chyby v jednom řádu. Obecně platí, že jestli pro kód s $\rho \geq n + 1$ lze identifikovat n chyb a pro kódu s $\rho \geq 2n + 1$ lze opravit n chyb.

5 Rozhodování za neurčitosti a rizika, statistické a Bayesovské rozhodování - jsem bohužel nikde nenašel (rozhodně ne v přednáškách v KUI)

Reference

Přednášky predmětu Kybernetika a Umělá Inteligence,
ČVUT FEL
http://cyber.felk.cvut.cz/gerstner/teaching/kui/index.htm

Bělohlávek R., *Přednášky predmětu Principy informatiky 2*,
Univerzita Palackého, Katedra informatiky.
http://www.inf.upol.cz/belohlavek/InformationTheoryAndApplications.pdf

Wikipedie (internetové encyklopedii), http://cs.wikipedia.org