

Лекция 2: Пропускная способность канала. Блочные коды.

Алексей Фролов

al.frolov@skoltech.ru

Сколковский институт науки и технологий (Сколтех)



Современные методы теории информации, оптимизации и управления

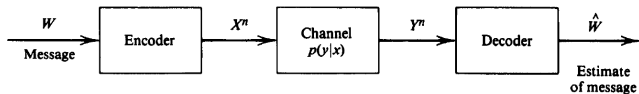
Сочи, Россия

2–23 августа, 2020

- 1 Дискретные каналы без памяти
- 2 Теорема кодирования
- 3 Гауссовский канал
- 4 Многопользовательские каналы
- 5 Блочные коды
- 6 Границы на параметры кодов
- 7 Линейные коды

- 1 Дискретные каналы без памяти
- 2 Теорема кодирования
- 3 Гауссовский канал
- 4 Многопользовательские каналы
- 5 Блочные коды
- 6 Границы на параметры кодов
- 7 Линейные коды

Дискретные каналы без памяти



Определение

Дискретный канал без памяти определяется так:

- ▶ $\mathcal{X}$ – входной алфавит (конечный);
- ▶ $\mathcal{Y}$ – выходной алфавит (конечный);
- ▶ матрица переходных вероятностей $P(y|x)$.

Определение

Канал без памяти, если

$$P(y^n|x^n) = \prod_{i=1}^n P(y_i|x_i).$$

Определение

$$C = \max_{P_X} \{I(X; Y)\},$$

где максимум берется по всем входным распределениям P_X .

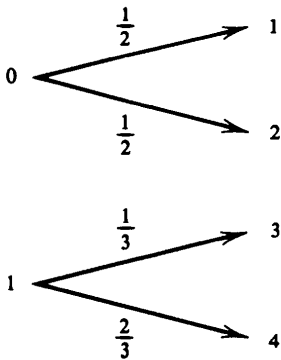
Пропускная способность – это наибольшая скорость (в битах за одно использование канала), при которой информация может быть передана со сколь угодно малой вероятностью ошибки.

Пример (Бесшумный канал)

0 → 0

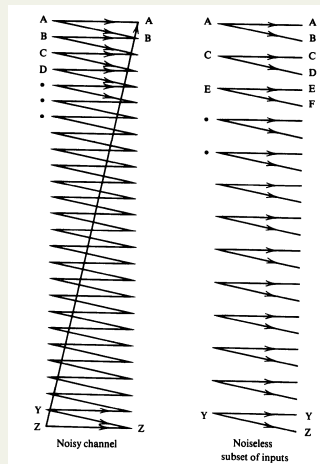
1 → 1

Пример (Канал с непересекающимися выходами)



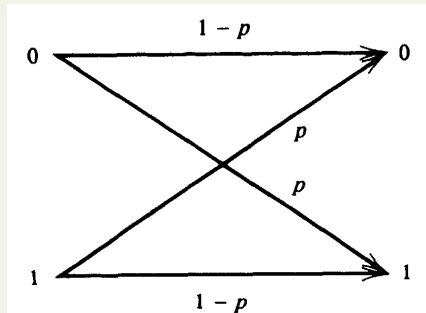
Пример (Шумная печатная машинка)

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(Y) - 1 \\ &= \log 26 - 1 = \log 13 \end{aligned}$$



Пример (Двоичный симметричный канал)

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_{x \in \mathcal{X}} P(x) H(Y|X = x) \\ &= H(Y) - \sum_{x \in \mathcal{X}} P(x) h(p) \\ &\leq 1 - h(p). \end{aligned}$$



Пример (Двоичный стирающий канал)

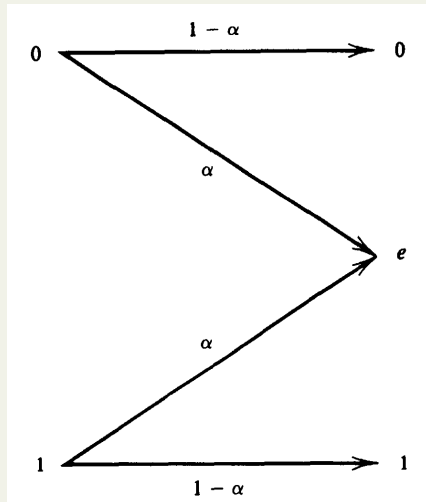
$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_{x \in \mathcal{X}} P(x) H(Y|X = x) \\ &= H(Y) - h(\alpha) \end{aligned}$$

Пусть E – это индикатор $I_{\{Y=e\}}$ и $\pi = \Pr(X = 1)$.

$$\begin{aligned} H(Y) &= H(Y, E) = H(E) + H(Y|E) \\ &= h(\alpha) + (1 - \alpha)h(\pi). \end{aligned}$$

Таким образом,

$$I(X; Y) = (1 - \alpha)h(\pi).$$



Пример

$$P(y|x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Предположим, что все строки в матрице переходных вероятностей являются перестановками друг друга, тогда (через $\mathbf{r}$ обозначим произвольную строку)

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H(\mathbf{r}) \\ &\leq \log |\mathcal{Y}| - H(\mathbf{r}). \end{aligned}$$

Предположим, что суммы по столбцам одинаковы и равны c , тогда $P(x) = 1/|\mathcal{X}|$ приводит к равномерному распределению/ на Y , т.е.

$$P(y) = \sum_{x \in \mathcal{X}} P(y|x)p(x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} P(y|x) = \frac{c}{|\mathcal{X}|}.$$

Определение

Канал называется симметричным, если строки матрицы переходных вероятностей являются перестановками друг друга и столбцы также являются перестановками друг друга.

Канал называется слабо симметричным, если строки матрицы переходных вероятностей являются перестановками друг друга, и суммы по столбцам одинаковы.

Пример (Слабо симметричный канал)

$$P(y|x) = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix}$$

Теорема

Для слабо симметричного канала,

$$C = \log |\mathcal{Y}| - H(\mathbf{r})$$

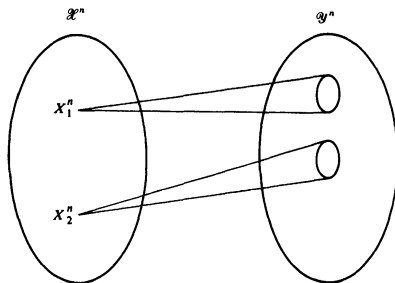
достигается на равномерном входном распределении.

- ▶ $C \geq 0$;
- ▶ $C \leq \log |\mathcal{X}|$;
- ▶ $C \leq \log |\mathcal{Y}|$;
- ▶ $I(X; Y)$ непрерывная функция $P(x)$;
- ▶ $I(X; Y)$ выпукла по $P(x)$;

Локальный минимум является глобальным и максимум существует и конечен.

- 1 Дискретные каналы без памяти
- 2 Теорема кодирования
- 3 Гауссовский канал
- 4 Многопользовательские каналы
- 5 Блочные коды
- 6 Границы на параметры кодов
- 7 Линейные коды

Основная идея: канал имеет подмножество входов, которые производят непересекающиеся последовательности на выходе.



- ▶ $\{1, 2, \dots, M\}$ – множество сообщений;
- ▶ $W \in \{1, 2, \dots, M\}$ – сообщение;
- ▶ $X^n(W) = \text{enc}(W)$ – кодовое слово;
- ▶ $Y^n \sim P(y^n|x^n)$ – принятая последовательность;
- ▶ $\hat{W} = \text{dec}(Y^n)$ – правило декодирования.

Определение (Код)

(M, n) код для канала $(\mathcal{X}, P(y|x), \mathcal{Y})$ состоит из:

- ▶ множества сообщений $\{1, 2, \dots, M\}$;
- ▶ функции кодирования $\text{enc}: \{1, 2, \dots, M\} \rightarrow \mathcal{X}^n$;
- ▶ функции декодирования $\text{dec}: \mathcal{Y}^n \rightarrow \{1, 2, \dots, M\}$.

Определение (Вероятность ошибки)

$$\lambda_i = \Pr(\text{dec}(Y^n) \neq i | X^n = X^n(i)).$$

$$\lambda^{(n)} = \max_{i \in \{1, 2, \dots, M\}} \lambda_i$$

$$P_e^{(n)} = \frac{1}{M} \sum_{i=1}^M \lambda_i$$

Определение

Скорость $R(M, n)$ кода определяется след образом

$$R = \frac{\log M}{n}.$$

Определение

Скорость R достижима, если существует последовательность кодов $(2^{Rn}, n)$, таких что $\lambda^{(n)} \rightarrow 0$ as $n \rightarrow \infty$.

Определение

Пропускная способность канала – супремум по всем достижимым скоростям.

$$A_{\varepsilon}^{(n)} = \left\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \right. \\ \left| -\frac{1}{n} \log P(x^n) - H(X) \right| < \varepsilon \\ \left| -\frac{1}{n} \log P(y^n) - H(Y) \right| < \varepsilon \\ \left. \left| -\frac{1}{n} \log P(x^n, y^n) - H(X, Y) \right| < \varepsilon \right\},$$

где $P(x^n, y^n) = \prod_{i=1}^n P(x_i, y_i)$.

Теорема

Пусть (X^n, Y^n) – это последовательности длины n , выбранные независимо и из $P(x^n, y^n)$, тогда

- 1 $\Pr((x^n, y^n) \in A_\epsilon^{(n)}) \rightarrow 1$ при $n \rightarrow \infty$;
- 2 $|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$;
- 3 Если $\hat{X}^n$ и $\hat{Y}^n$ независимые с распределениями, совпадающими с маргинальными распределениями $P(x^n, y^n)$, тогда

$$\Pr((\hat{x}^n, \hat{y}^n) \in A_\epsilon^{(n)}) \leq 2^{-n(I(X;Y)-3\epsilon)}$$

and

$$\Pr((\hat{x}^n, \hat{y}^n) \in A_\epsilon^{(n)}) \geq (1 - \epsilon)2^{-n(I(X;Y)+3\epsilon)}$$

Теорема

- ▶ Все скорости меньше C достижимы. В частности для любой скорости $R < C$, существует последовательность $(2^{Rn}, n)$ кодов с $\lambda^{(n)} \rightarrow 0$.
- ▶ Обратно, любая последовательность $(2^{Rn}, n)$ кодов с $\lambda^{(n)} \rightarrow 0$ удовлетворяет $R \leq C$.

Определение (Ансамбль кодов)

$$\mathcal{C} = \begin{bmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ x_1(2) & x_2(2) & \dots & x_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(2^{Rn}) & x_2(2^{Rn}) & \dots & x_n(2^{Rn}) \end{bmatrix}$$

Каждый элемент выбирается независимо из $P(x)$.

Получатель заявляет, что был передан индекс $\hat{W}$, если выполняются следующие условия:

- ▶ пара $(X^n(\hat{W}), Y^n)$ является совместно типичной;
- ▶ нет другого i , такого что пара $(X^n(i), Y^n)$ является совместно типичной.

Пусть $\Pr(\mathcal{E})$ – это средняя (по выбору кодовой книги) вероятность ошибки.

$$\begin{aligned}\Pr(\mathcal{E}) &= \sum_{\mathcal{C}} P(\mathcal{C}) P_e^{(n)}(\mathcal{C}) \\&= \sum_{\mathcal{C}} P(\mathcal{C}) \frac{1}{2^{Rn}} \sum_{w=1}^{2^{Rn}} \lambda_w(\mathcal{C}) \\&= \frac{1}{2^{Rn}} \sum_{w=1}^{2^{Rn}} \sum_{\mathcal{C}} P(\mathcal{C}) \lambda_w(\mathcal{C}) \\&= \sum_{\mathcal{C}} P(\mathcal{C}) \lambda_1(\mathcal{C}) \text{ (симметрия)} \\&= \Pr(\mathcal{E} | W = 1).\end{aligned}$$

Определим

$$E_i = \{(X^n(i), Y^n) \in A_\varepsilon^n\}.$$

$$\begin{aligned} \Pr(\mathcal{E} | W = 1) &\leq \Pr(E_1^c) + \sum_{i=2}^{2^{Rn}} \Pr(E_i) \\ &\leq \varepsilon + (2^{Rn} - 1) 2^{-n[I(X; Y) - 3\varepsilon]} \end{aligned}$$

Таким образом, если $R < I(X; Y)$, то можно выбрать ε и n , такие что $\Pr(\mathcal{E})$ меньше, чем ε' .

Цепь Маркова:

$$W \rightarrow X^n(W) \rightarrow Y^n \rightarrow \hat{W}.$$

Converse.

Неравенство Фано

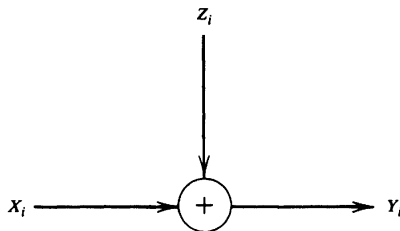
$$H(W|Y^n) \leq 1 + P_e^{(n)} nR$$

и $I(X^n; Y^n) \leq nC$

$$\begin{aligned} nR &= H(W) = H(W|Y^n) + I(W; Y^n) \\ &\leq H(W|Y^n) + I(X^n(W); Y^n) \\ &\leq 1 + P_e^{(n)} nR + I(X^n(W); Y^n) \\ &\leq 1 + P_e^{(n)} nR + nC. \end{aligned}$$



- 1 Дискретные каналы без памяти
- 2 Теорема кодирования
- 3 Гауссовский канал**
- 4 Многопользовательские каналы
- 5 Блочные коды
- 6 Границы на параметры кодов
- 7 Линейные коды



$$Y_i = X_i + Z_i, \quad Z_i \sim N(0, N).$$

Ограничение на мощность

$$\frac{1}{n} \sum_{i=1}^n x_i^2 \leq P.$$

$$SNR = \frac{P}{N} \quad (\text{линейный})$$

$$SNR = 10 \log_{10} \frac{P}{N} \quad (\text{dB})$$

Определение

Пропускная способность гауссовского канала при ограничении P определяется как

$$C = \max_{p(x): \mathbb{E}[X^2] \leq P} I(X; Y).$$

Теорема

$$C = \frac{1}{2} \log(1 + SNR).$$

Доказательство.

$$\begin{aligned} I(X; Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X + Z|X) \\ &= h(Y) - h(Z|X) \\ &= h(Y) - h(Z) \\ &= h(Y) - \frac{1}{2} \log 2\pi e N \end{aligned}$$

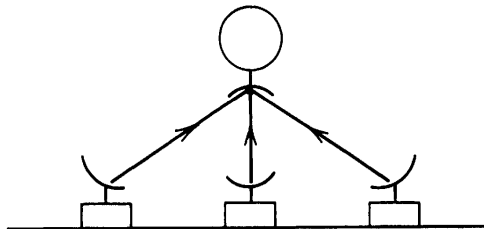
$$\mathbb{E}[Y^2] = \mathbb{E}[(X + Z)^2] = \mathbb{E}[X^2] + \mathbb{E}[Z^2] + 2\mathbb{E}[X]\mathbb{E}[Z] = P + N$$

Таким образом,

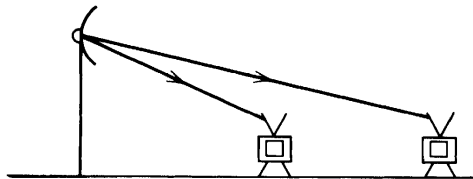
$$h(Y) = \frac{1}{2} \log 2\pi e (P + N).$$



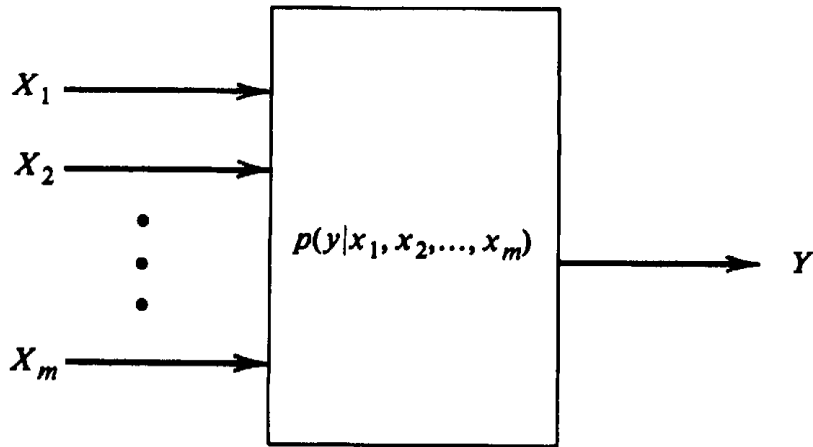
- 1 Дискретные каналы без памяти
- 2 Теорема кодирования
- 3 Гауссовский канал
- 4 Многопользовательские каналы**
- 5 Блочные коды
- 6 Границы на параметры кодов
- 7 Линейные коды



(a) MAC



(b) Широковещательный канал



Пусть $S \subseteq \{1, 2, \dots, m\}$, $R(S) = \sum_{i \in S} R_i$ и $X(S) = \{X(i), i \in S\}$.

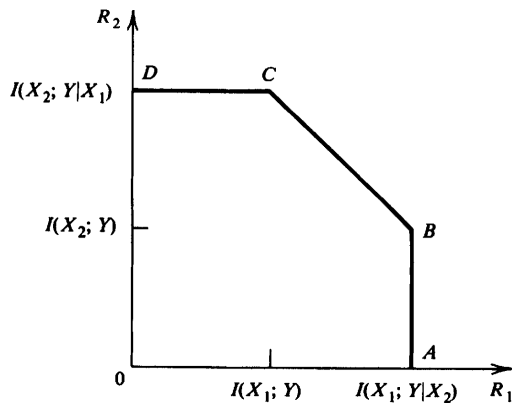
Теорема

Область пропускной способности многопользовательского канала – это выпуклая оболочка

$$R(S) \leq I(X(S); Y | X(S^c)) \quad \forall S \subseteq \{1, 2, \dots, m\}.$$

для $P_1(x_1)P_2(x_2) \dots P_m(x_m)$.

$$\begin{aligned}R_1 &\leq I(X_1; Y|X_2) \\R_2 &\leq I(X_2; Y|X_1) \\R_1 + R_2 &\leq I(X_1, X_2; Y)\end{aligned}$$



$$Y = X_1 + X_2 + Z.$$

$$C(x) = \frac{1}{2} \log(1 + x).$$

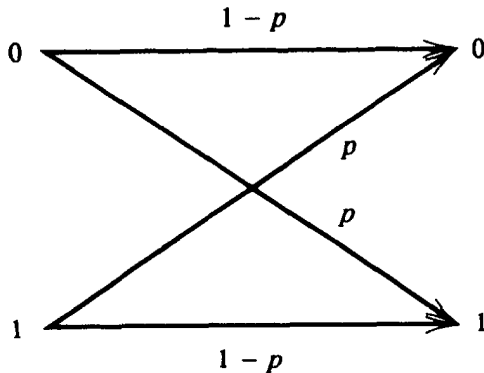
$$R_1 \leq C\left(\frac{P_1}{N}\right)$$

$$R_2 \leq C\left(\frac{P_2}{N}\right)$$

$$R_1 + R_2 \leq C\left(\frac{P_1 + P_2}{N}\right)$$

- ▶ Подсчет пропускной способности для ДНК-каналов
- ▶ Граница случайного кодирования для канала массового некоординированного множественного доступа (случай конечной длины)

- 1 Дискретные каналы без памяти
- 2 Теорема кодирования
- 3 Гауссовский канал
- 4 Многопользовательские каналы
- 5 Блочные коды
- 6 Границы на параметры кодов
- 7 Линейные коды



- ▶ $\{1, 2, \dots, M\}$ – множество сообщений;
- ▶ $Q = \{0, \dots, q - 1\}$;
- ▶ $\mathbf{x} = \Psi(i) \in Q^n$ – кодовое слово;
- ▶ $\mathcal{C} = \{\mathbf{x} = \Psi(i), i = 1, \dots, M\}$ – код;
- ▶ кодовая книга;
- ▶ $\mathbf{y} \sim P(y^n | x^n)$ – принятая последовательность;
- ▶ $\hat{i} = \Psi^{-1}(\mathbf{y})$ – правило декодирования.
- ▶ $R = \frac{\log_q M}{n} = \frac{k}{n}$.

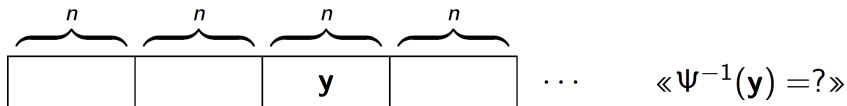
$$\psi : \begin{cases} 00 \rightarrow 00001 \\ 01 \rightarrow 01010 \\ 10 \rightarrow 10111 \\ 11 \rightarrow 11100 \end{cases}$$

$$\mathbf{y} = 10101$$

$\mathbf{x}$	$P(\mathbf{y} \mathbf{x})$
00001	$p^2(1-p)^3$
01010	p^5
10111	$p(1-p)^4$
11100	$p^2(1-p)^3$

$$p^5 < p^2(1-p)^3 < p(1-p)^4$$

$$\Rightarrow \mathbf{x} = 10111, \mathbf{i} = 10$$



Декодирование по максимуму правдоподобия:

- 1 $x = \arg \max_{x \in \mathcal{C}} P(y|x);$
- 2 $i = \Psi^{-1}(x).$

Лемма

Пусть $\mathcal{C} = \{x_i\}$, $p < 0.5$ и $P(y|x) = \max_i P(y|x_i)$, тогда

$$d(y, x) = \min_i d(y, x_i),$$

где $d(y, x)$ обозначает число позиций, в которых y и x отличаются.

Определение

Пусть $\alpha, \beta \in Q^n$.

$$d(\alpha, \beta) = |\{i : \alpha(i) \neq \beta(i)\}|.$$

Пример

$$\alpha = 01101$$

$$\beta = 00111$$

$$d(\alpha, \beta) = 2.$$

Определение

- ▶ $||\alpha|| = d(\alpha, \mathbf{0})$ – вес α ;
- ▶ $|\alpha| = \sum_{i=1}^n \alpha_i q^{n-i}$ – номер (лексикографический порядок) α ;

Определение

$$B_r(\alpha) = \{\beta \in Q^n : d(\alpha, \beta) \leq r\}$$

and

$$S_r(\alpha) = \{\beta \in Q^n : d(\alpha, \beta) = r\}$$

$$|S_r(\alpha)| = \binom{n}{r} (q-1)^r$$

и

$$|B_r(\alpha)| = \sum_{i=0}^r |S_r(\alpha)| = \sum_{i=0}^r \binom{n}{i} (q-1)^i$$

Определение

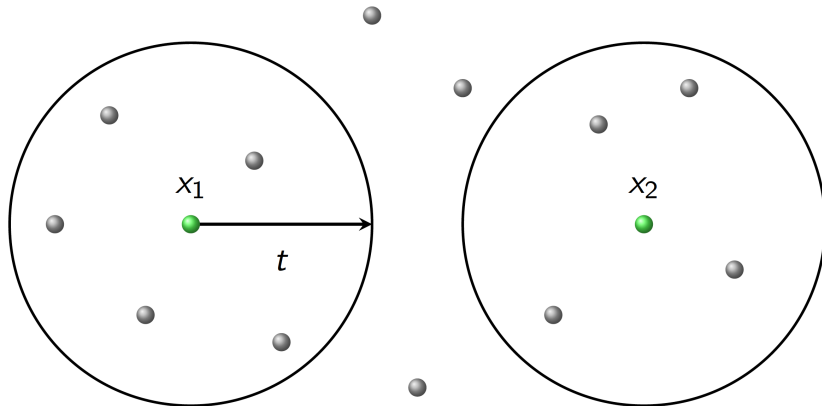
- ▶ Код $C \subseteq Q^n$;
- ▶ Минимальное расстояние

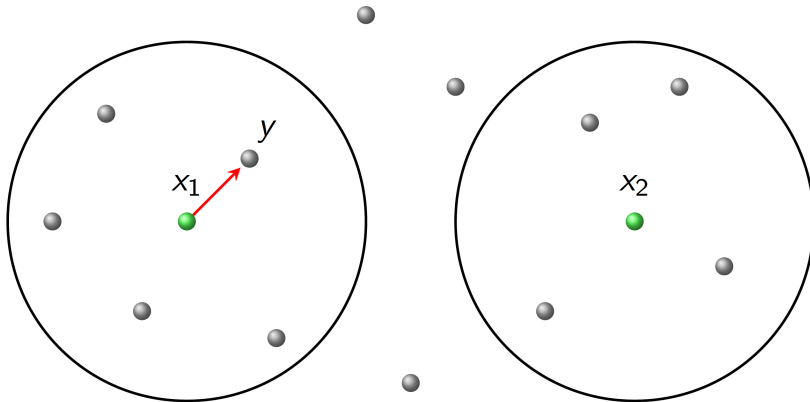
$$d(C) = \min_{a, b \in C; a \neq b} d(a, b).$$

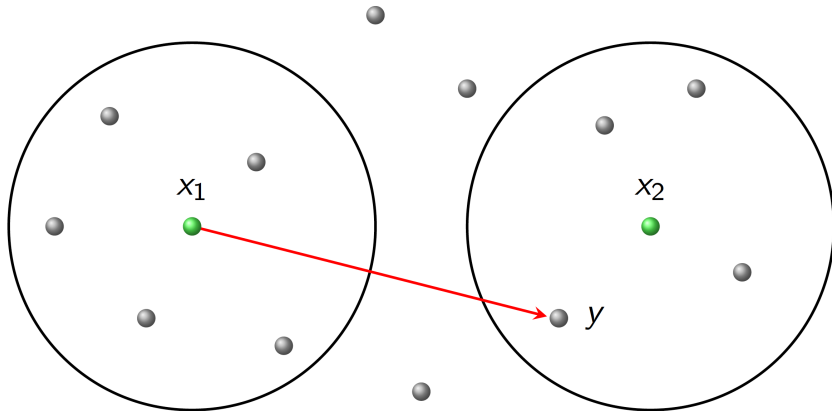
Теорема

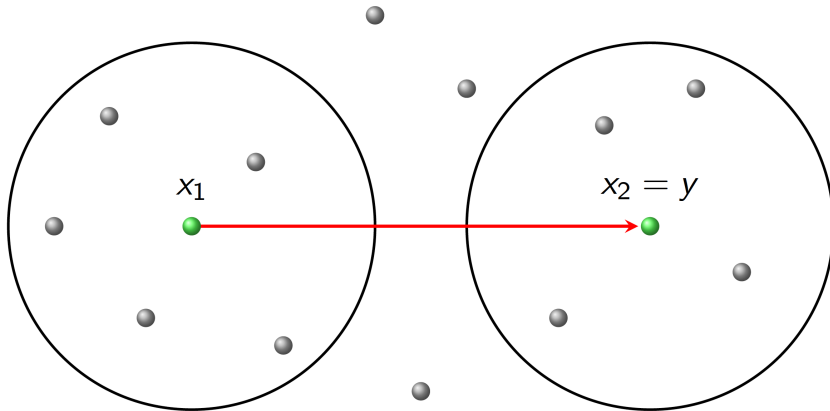
Пусть код C может исправить любые t ошибок, тогда

$$d(C) \geq 2t + 1.$$









Теорема

Пусть $d(C) = d$, тогда

$$t = \left\lfloor \frac{d-1}{2} \right\rfloor$$

и

$$s = d - 1.$$

- 1 Дискретные каналы без памяти
- 2 Теорема кодирования
- 3 Гауссовский канал
- 4 Многопользовательские каналы
- 5 Блочные коды
- 6 Границы на параметры кодов**
- 7 Линейные коды

Определение

$$A_q(n, d) = \max_{\mathcal{C} \subseteq Q^n, d(\mathcal{C})=d} |\mathcal{C}|.$$

Отметим, что максимизация мощности и скорости при фиксированном n – одинаковые задачи.

Далее будем опускать индекс q при $q = 2$.

Пусть $\alpha \in Q^n$. Введем обозначение

$$V_t = V_q(t) = |B_t(\alpha)| = \sum_{i=0}^t \binom{n}{i} (q-1)^i.$$

Теорема (Граница Хэмминга)

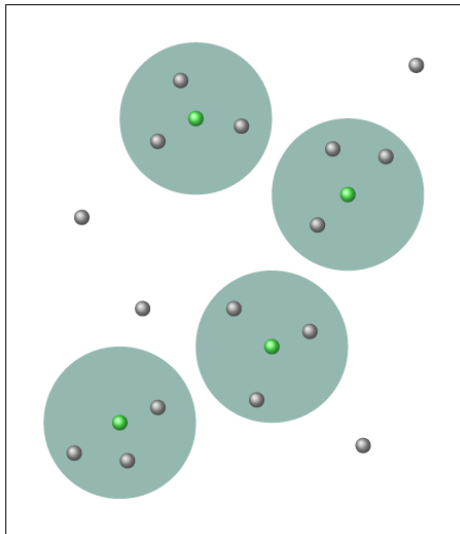
$$A_q(n, d) \leq \frac{q^n}{V_t}.$$

Определение

Код называется совершенным, если $|\mathcal{C}| = \frac{q^n}{V_t}$.

Пример

Код $\mathcal{C} = \{000, 111\} \subset \{0, 1\}^3$ является совершенным.

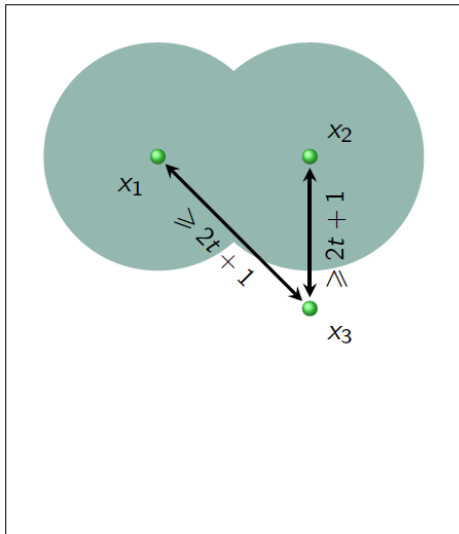


Шары радиуса t не пересекаются!

$$A_q(n, d) V_q(t) \leq q^n.$$

Теорема (Граница Гилберта)

$$A_q(n, d) \geq \frac{q^n}{V_{2t}}.$$



$$x_3 \notin B_{2t}(x_1) \cup B_{2t}(x_2)$$

$C_3 = \{x_1, x_2, x_3\}$ исправляет t ошибок.

Пусть мы построили m кодовых слов и не можем добавить новое слово

$$q^n = \left| \bigcup_{i=1}^m B_{2t}(x_i) \right| \leq mV_{2t}.$$

$$\frac{d}{n} \rightarrow \delta, \frac{\log_q M}{n} = \frac{k}{n} \rightarrow R$$

Определение

Семейство $\{C_n\}$ называется асимптотически хорошим, если $R, \delta > 0$:

- ▶ $\frac{\log_q M_n}{n} = \frac{k_n}{n} \geq R > 0$;
- ▶ $\frac{d_n}{n} \geq \delta > 0$;

Граница Хэмминга

$$R \leq 1 - h(\delta/2).$$

Граница Гилберта

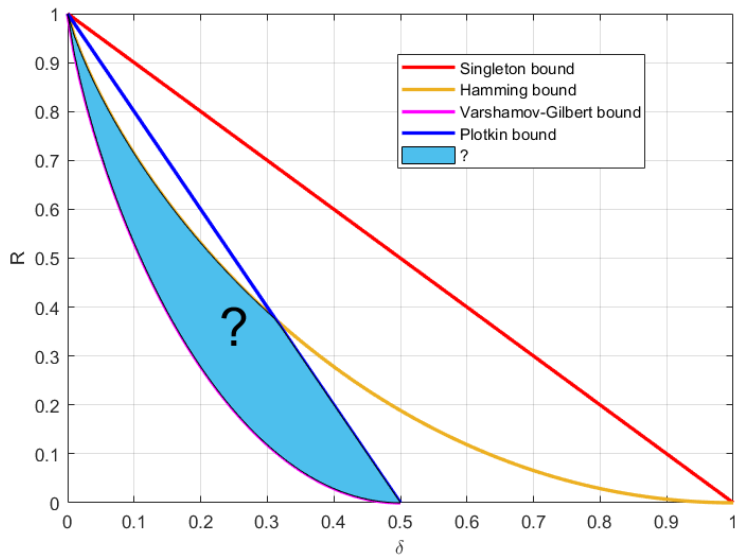
$$R \geq 1 - h(\delta).$$

Граница Синглтона

$$R \leq 1 - \delta.$$

Граница Плоткина

$$R \leq 1 - 2\delta.$$



- 1 Дискретные каналы без памяти
- 2 Теорема кодирования
- 3 Гауссовский канал
- 4 Многопользовательские каналы
- 5 Блочные коды
- 6 Границы на параметры кодов
- 7 Линейные коды

Определение

Подгруппа Абелевой группы $\mathbb{F}_q^n$ называется групповым кодом.

Определение

Подпространство $\mathcal{C}$ линейного пространства $\mathbb{F}_q^n$ называется линейным (n, k) кодом, где $k = \dim \mathcal{C}$.

Определение

G – порождающая матрица $\mathcal{C}$, если ее строки образуют базис $\mathcal{C}$.

Лемма

Пусть $\mathcal{C}$ – это линейный код, тогда

$$d(\mathcal{C}) = \min_{a \in \mathcal{C}, a \neq 0} \|a\|.$$

Определение (Дуальный код)

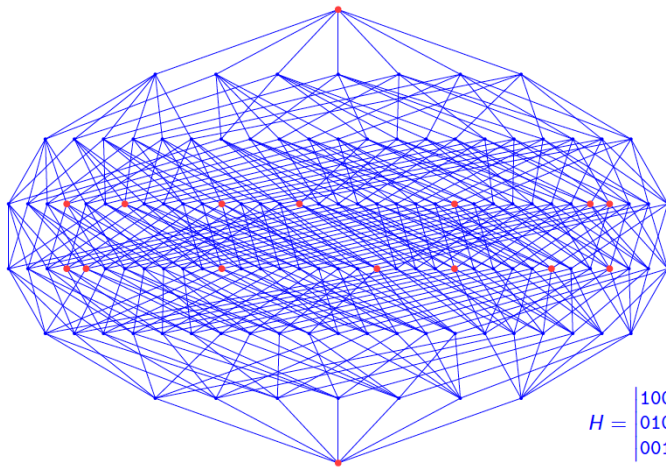
$$C^\perp = \{v \in \mathbb{F}_q^n : v \perp C\}.$$

Определение (Проверочная матрица)

H – это базис $C^\perp$.

Параметры:

$$n = 2^m - 1, k = 2^m - m - 1, d = 3.$$



$$H = \begin{vmatrix} 1000111 \\ 0101011 \\ 0011101 \end{vmatrix}$$

$$H_2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad H_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix},$$

$$H_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

- ▶ Применение нейронных сетей для декодирования линейных кодов. Подход на основе синдрома
- ▶ Применение нейронных сетей для декодирования линейных кодов. Подход на основе графа Таннера
- ▶ Построение помехоустойчивых кодов с помощью машинного обучения
- ▶ Постквантовая (кодовая) криптография
- ▶ Кодовые распределенные вычисления для задач машинного обучения.

Спасибо за внимание!