به نام خدا

- مفاهیم پایه‌ای تئوری اطلاعات
- مفهوم اطلاعات
- مفهوم آنتروپی
- آنتروپی توأم یا مشترک
- آنتروپی شرطی

* اطلاعات متقابل

منظور از اطلاعات متقابل در متغیرهای تصادفی X و Y، اطلاعاتی است که یکی از این متغیرها، در مورد دیگری به دست می‌دهد. اطلاعات متقابل X و Y را با $I(X;Y)$ نمایش می‌دهیم و به صورت زیر تعریف می‌کنیم.

$$I(X;Y) \triangleq \underset{P(x,y)}{E}\left\{\log \frac{P(X,Y)}{P(X)\,P(Y)}\right\}$$

$$\xrightarrow[\text{نتیجه گرفت}]{\text{از رابطه بالا می‌توان}} \quad I(X;Y) = I(Y;X)$$

همچنین اگر X , Y مستقل از هم باشند، داریم $P(x,y) = P(x)\,P(y)$

بنابراین در این صورت اطلاعات متقابل در متغیر تصادفی برابر صفر است.

$$X \perp\!\!\!\perp Y \implies I(X;Y) = I(Y;X) = 0$$

* همچنین می‌توان نشان داد که همواره $I(X;Y) \geqslant 0$ برقرار است.

$$\forall X, Y \; ; \quad I(X;Y) \geqslant 0$$

در ادامه می‌خواهیم رابطه بین اطلاعات متقابل $I(X;Y)$ را با مفاهیم آنتروپی بیان کنیم.
برای این منظور کافی است که رابطه اطلاعات متقابل را بازنویسی کنیم. می‌دانیم

$$I(X;Y) = E_{P(x,y)}\left\{ \log \frac{P(X,Y)}{P(X)P(Y)} \right\}$$

$$= \sum_{x}\sum_{y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

رابطه زنجیره‌ای $P(x,y) = \underbrace{P(x)P(y|x)}_{2} = \underbrace{P(y)P(x|y)}_{1}$

1

$$\Rightarrow I(X;Y) = \sum_x \sum_y P(x,y) \log \frac{\cancel{p(y)}\, p(x|y)}{P(x)\, \cancel{P(y)}}$$

$$= \sum_x \sum_y p(x,y) \left[\log p(x|y) - \log p(x) \right]$$

$$= \underbrace{\sum_x \sum_y p(x,y) \log p(x|y)}_{-H(X|Y)} - \sum_x \underbrace{\sum_y p(x,y)}_{p(x)} \log p(x)$$

$$= -H(X|Y) - \underbrace{\sum_x p(x) \log p(x)}_{H(X)}$$

① $\Longrightarrow I(X;Y) = H(X) - H(X|Y) \geq 0$

② $\xRightarrow[\text{به طریق مشابه}]{} I(X;Y) = H(Y) - H(Y|X) \geq 0$

$$\xRightarrow[H(X,Y) = H(Y) + H(X|Y)]{H(X,Y) = H(X) + H(Y|X)} I(X;Y) = H(X) + H(Y) - H(X,Y) \geq 0$$

روابط زنجیره‌ای در ادامه گفته می‌شود

$$I(X;Y) \geq 0 \Longrightarrow \quad \text{(I)} \quad H(X) \geq H(X|Y)$$

$$\text{(II)} \quad H(Y) \geq H(Y|X)$$

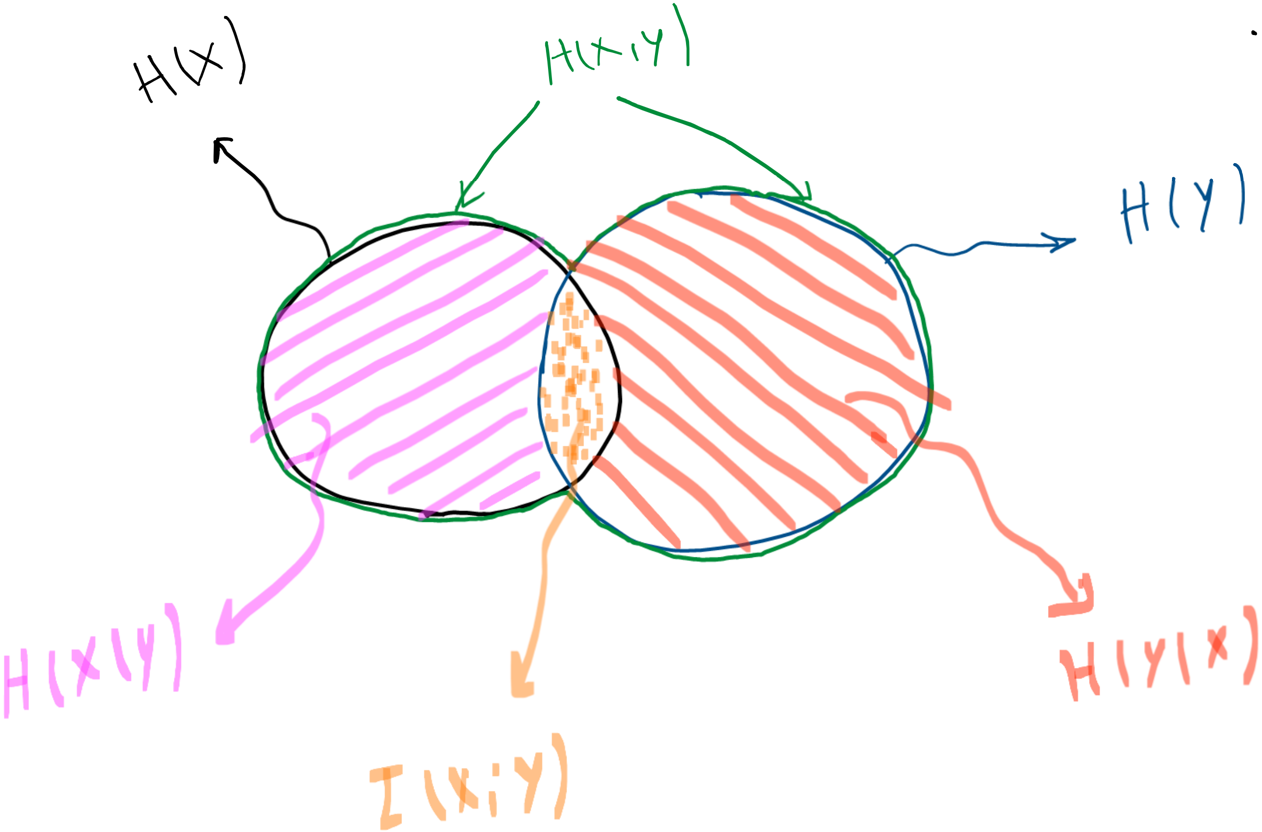
مشروط کردن از ابهام
(اطلاعات) کم می کند.

$$\text{(III)} \quad H(X) + H(Y) \geq H(X,Y)$$

$$I(X;X) = H(X)$$

$$I(X;X) = H(X) - \underbrace{H(X|X)}_{0} = H(X)$$

مطالب گفته شده را می‌توانیم با کمک دیاگرام ون (Van Diagoan) در تئوری اطلاعات، به صورت زیر خلاصه کنیم.



$H(X,Y)$

$H(X|Y)$

$H(Y|X)$

$I(X;Y)$

حدودی که تئوری اطلاعات به ما می‌دهد

دو مسأله دوگان

مسأله کدینگ منبع (فشرده‌سازی)

$$\min_{p(\hat{x}|x)} I(X;\hat{X})$$

تابع اعوجاج $R(D)$ $\leq d$

$$\underbrace{\hphantom{R(D)\leq d}}$$

X : متغیر تصادفی منبع

$\hat{X}$: بازیابی از X

مسأله ظرفیت کانال

$$\max_{p(x)} I(X;Y)$$

$$\underbrace{\hphantom{\max I(X;Y)}}$$

ظرفیت کانال

X : ورودی کانال

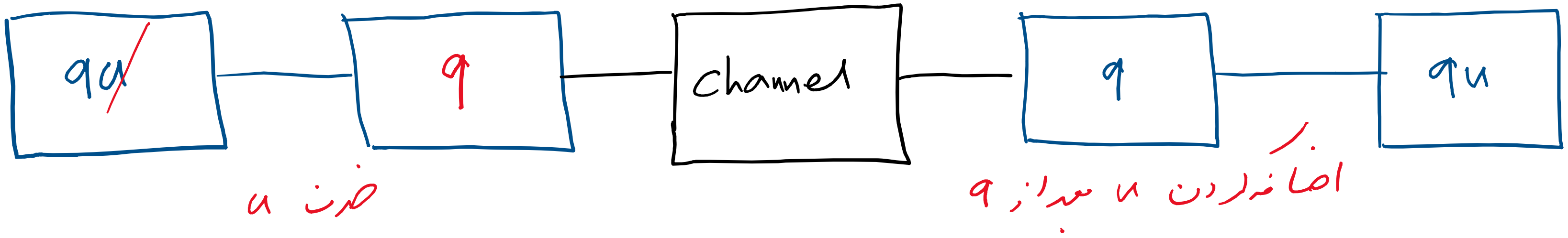
Y : خروجی کانال

در ادامه مباحث این درس، به مسأله کدینگ منبع می‌پردازیم.

* کدینگ منبع Source Coding

همانطور که قبلاً نیز اشاره شد، کدینگ منبع، عملیاتی است که بر روی خروجی منبع اطلاعات انجام می‌شود و اضافات منبع را به روشهای کنترل شده‌ای حذف می‌کند، به طوری که بازیابی اطلاعات در سمت گیرنده، با خطای به میزان دلخواه کم، امکان‌پذیر باشد. به این ترتیب از پهنای باند سیستم نیز به شکل بهتری استفاده می‌شود.

به طور مثال در یک متن انگلیسی، می‌دانیم که همواره بعد از حرف q ، حرف u می‌آید. بنابراین اگر بخواهیم یک متن انگلیسی را بفرستیم، می‌توانیم u های بعد از q را حذف کنیم (ارسال نکنیم) و گیرنده بعد از دریافت هر حرف q ، خودش یک حرف u بعد از آن اضافه کند.



به عنوان مثالی دیگر می‌توان به روش مورس در تلگراف اشاره کرد. در این روش به هر یک از الفبای خروجی منبع، یک کد اختصاص داده می‌شود و به سمت گیرنده فرستاده می‌شود. در روش‌های کدینگ منبع معمولاً از همین تکنیک استفاده می‌شود.

بحث کدینگ منبع معمولاً تحت عنوان کلی فشرده‌سازی (Compressing) مطرح می‌شود. در بحث فشرده‌سازی مسأله این است که بتوانیم یک بیان (representation) از الفبای خروجی منبع پیدا کنیم (برای ارسال روی کانال) به طوری که این بیان از الفبای منبع، کمترین میزان اعوجاج (Distortion) را نسبت به بیان اولیه منبع داشته باشد.

برای این منظور تابعی تعریف می‌شود که نشان دهنده‌ی کمترین میزان اعوجاج در کمترین نرخ ممکن است و به آن تابع Rate Distortion می‌گوییم. R(D)

به این ترتیب مسأله فشرده‌سازی به صورت زیر خواهد بود،

$$R(D) = \min_{p(x|\hat{x})} I(X;\hat{X}) \quad \text{subject to} \quad E\underbrace{d(X,\hat{X})}_{\text{که براساس خصوصیات منبع تعریف می‌شود}} \leq D$$

این مسئله به دو دسته کلی مسائل فشرده سازی lossy و مسائل فشرده سازی Lossless قابل تفکیک است.

- در روشهای فشرده سازی lossy که حداکثر اعوجاج D قابل تحمل است و بازیابی اطلاعات به صورت قابل اطمینان با این حداکثر اعوجاج امکان پذیر است.

- در روش‌های lossless ، اعوجاج قابل تحمل نیست و همواره $D = 0$ برقرار است.

بنابراین مسأله فشرده‌سازی در حالت lossless به صورت زیر قابل بیان است.

$$R(D) = \min_{\substack{p(x|\hat{x}) \\ \text{subject to } D=0}} I(X;\hat{X}) \equiv H(X)$$

می‌توان نشان داد

به عبارت دیگر می‌توان نشان داد که حد نهایی این نوع فشرده‌سازی، برابر آنتروپی منبع است.
یا با بیان معادل می‌توان گفت که کمترین میانگین طول کلمات کد منبع که برابر کمترین تعداد بیت‌های لازم
برای بیان اطلاعات منبع نیز هست، برابر آنتروپی منبع است.

⟸ در مسأله فشرده سازی lossless منبع (با کد کردن منبع) هدف این است که به حروف از الفبای خروجی منبع، یک کلمه کد نسبت بدهیم به طوری که میانگین طول کلمات کد تا حد ممکن به آنتروپی منبع نزدیک باشد (در نتیجه بتوانیم از الفبای باینری به بهترین شکل استفاده کنیم). همچنین تخصیص کلمات کد به الفبای منبع باید برگشت پذیر باشد به طوری که در گیرنده بتوانیم بدون ابهام و با خطای به میزان دلخواه کم، اطلاعات منبع را بازیابی کنیم.

* در این بحث، منابع گسسته را در نظر می گیریم.

در ادامه ابتدا تعاریف پایه‌ای در کدینگ منبع را مطرح می‌کنیم و سپس با یک مثال، خصوصیات کدینگ منبع را بیان خواهیم کرد و سپس به بیان چارچوب تئوری مسأله می‌پردازیم.

تعریف کد منبع Source Coding

فرض می‌کنیم که خروجی منبع، متغیر تصادفی X است که مقادیری از یک مجموعه شمارش‌پذیر $\mathcal{X}$ را اختیار می‌کند. یعنی

$$X = x \in \{x_1, x_2, \dots, x_m\} = \mathcal{X}$$

کد منبع C یک نگاشت از مجموعه $\mathcal{X}$ به مجموعه‌ی M^* است.

M^* مجموعه‌ی سمبل های با طول محدود M است که شامل سمبل های کلمات کد متناظر با الفبای خروجی منبع است. کلمه کد مربوط به الفبای x را با $C(x)$ و طول آن را با $l(x)$ نمایش می دهیم.

$M\text{-ary} : \{0, 1, \dots, M-1\}$

مثال: فرض کنیم الفبای خروجی منبع از مجموعه $X = \{Red, Green, Blue\}$ است و می خواهیم به هر یک از این خروجی ها، یک کد باینری اختصاص بدهیم. بنابراین داریم.

$M = 2 \longrightarrow \{0, 1\}$

$x_1 = Red \quad , \quad x_2 = Green \quad , \quad x_3 = Blue$

به طور مثال می توان کلمات کد متناظر با x_1, x_2, x_3 را به صورت زیر در نظر گرفت.

$$C(x_1) = C(\text{Red}) = 00 \longrightarrow \ell(x_1) = 2$$

$$C(x_2) = C(\text{Green}) = 01 \longrightarrow \ell(x_2) = 2$$

$$C(x_3) = C(\text{Blue}) = 11 \longrightarrow \ell(x_3) = 2$$

میانگین طول کلمات کد برای یک کد منبع C که آن را با $L(C)$ نمایش می دهیم، برابر امید ریاضی طول کلمات کد منبع است یعنی

$$L(C) = E\{\ell(X)\} = \sum_{x} p(x)\ell(x)$$

* برای مثال قبل با فرض هم احتمال بودن الفبای منبع داریم

$$L(C) = \sum_{x} p(x)\, l(x) = \sum_{i=1}^{3} p(x_i)\, l(x_i) = 3\left(\frac{1}{3} \times 2\right) = 2$$

* بر اساس قضیه های اساسی تئوری اطلاعات همواره داریم $L(C) \geq H(X)$

بنابراین در یک کدینگ منبع (فشرده سازی) lossless به دنبال اختصاص کلمات کد مناسب به الفبای خروجی منبع هستیم، به طوری که میانگین طول کلمات کد تا حد ممکن به آنتروپی منبع نزدیک باشد تا بتوانیم کمترین اضافات را روی کانال داشته باشیم و به بهترین شکل از پهنای باند استفاده کنیم.

(توجه داشته باشیم که این عملیات باید برگشت پذیر باشد و در گیرنده اطلاعات را بدون ابهام و با خطای کم بتوانیم بازیابی کنیم)

در ادامه می خواهیم با یک مثال، خصوصیات لازم برای یک کد منبع که ما را به اهداف طراحی گفته شده می رساند را بیان کنیم.

مثال: یک منبع را در نظر می گیریم که الفبای خروجی آن به صورت $X = \{\underbrace{a}_{x_1}, \underbrace{b}_{x_2}, \underbrace{c}_{x_3}, \underbrace{d}_{x_4}\}$ با احتمالات $\{1/8, 1/8, 1/4, 1/2\}$ است. با توجه به خصوصیات گفته شده در مورد کدینگ منبع، یک کد منبع مناسب پیشنهاد کنید.

$P_r\{X = x_1 = a\} = \frac{1}{2}$ $\quad C(x_1) = ?$

$P_r\{X = x_2 = b\} = \frac{1}{4}$ $\quad C(x_2) = ?$

$P_r\{X = x_3 = c\} = \frac{1}{8}$ $\quad C(x_3) = ?$

$P_r\{X = x_4 = d\} = 1/8$ $\quad C(x_4) = ?$

(کد را باینری در نظر می گیریم)

$\{0, 1\}$

برای این مثال، قبلاً $H(X)$ را محاسبه کرده‌ام.

$$H(X) = -\sum_{x} p(x) \log p(x) = -\sum_{i=1}^{4} p(x_i) \log p(x_i)$$

$$\Longrightarrow H(X) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + 2 \times \frac{1}{8}\log 8 = \frac{7}{4} \text{ bits}$$

پیشنهاد 1:

$c(x_1) = 00$

$c(x_2) = 01$

$c(x_3) = 10$

$c(x_4) = 11$

این کد، یک کد منبع با طول کلمات کد یکسان است.

$$\forall x_i \in \mathcal{X} \quad ; \quad l(x_i) = 2$$

$$\mathcal{X} = \{a, b, c, d\}$$

این کد منبع دارای پیچیدگی کم در پیاده‌سازی است و به راحتی می‌توان عملیات کدینگ را انجام داد. علاوه بر این عملیات دیکدینگ نیز به صورت ساده و بدون ابهام امکان پذیر است.

زیرا با نگاشت یک به یک از مجموعه‌ی $X=\{a,b,c,d\}$ به مجموعه‌ی کلمات کد باینری به طول 2 $\{00,01,10,11\}$ داریم. به عنوان مثال اگر رشته دریافتی در گیرنده به صورت زیر باشد،

$$\underbrace{00}_{a}\underbrace{10}_{c}\underbrace{11}_{d}\underbrace{11}_{d}\underbrace{00}_{a}\underbrace{01}_{b}\underbrace{01}_{b}\ldots$$

بازیابی کلمات بدون ابهام و به سادگی امکان پذیر است.

میانگین طول کلمات کد برای کد پیشنهادی یک (C_1) به صورت زیر است،

$$L(C_1) = E\{l(X)\} = \sum_{i=1}^{4} p(x_i)\underbrace{l(x_i)}_{2} = 2\underbrace{\sum_{i=1}^{4} p(x_i)}_{1} = 2$$

$$L(C_1) = 2 > \frac{7}{4} = H(X)$$

بنابراین $L(C_1)$ از $H(X)$ فاصله دارد. یعنی این که به کمینه میانگین طول کلمات کد، دست نمی‌یابد، زیرا در طراحی این کد به خصوصیات آماری منبع توجه نشده است و بدون توجه به احتمال وقوع الفباهای خروجی منبع، به همه‌ی آنها کدهای با طول یکسان اختصاص داده است.