نظریه اطلاعات شبکه، ترم بهار ۱۳۹۴-۱۳۹۳

مدرس: امین زاده گوهری

جلسه ۲

پوش محدب گیری تحدید شده از نمودار یک تابع*

این بخش جزو درس نیست و برای مطالعه شخصی قرار داده شده است.

دیدیم که

$$\mathbb{C}[f](p(x)) = \min_{p(y|x)} \sum_y p(Y=y) f\big(p(x|Y=y)\big).$$

حال فرض کنید که مجموعه کانال های $p(y|x)$ که روی آن مینیمم گیری میکنیم را محدود کنیم. این کار را اینگونه انجام میدهیم. اگر کانال ثابتی مانند $p(z|x)$ داشته باشیم آنوقت مجموعه کانال های $p(y|x)$ را مییابیم که رابطه مارکف $X-Z-Y$ برقرار باشد. یعنی X ابتدا از کانال $p(z|x)$ عبور کرده و سپس آن را از کانال $p(y|z)$ عبور میکند تا Y بدست آید. کانال $p(y|z)$ در اینجا دلخواه است.

تعریف ۱ پوش محدب گیری تحدید شده به کانال $p(z|x)$ از نمودار یک تابع f روی سیمپلکس احتمالی اینگونه تعریف میشود:

$$\mathbb{C}_{p(z|x)}[f](p(x)) = \min_{p(y|z)} \sum_y p(Y=y) f\big(p(x|Y=y)\big),$$

که در آن توزیع مشترک $p(x,y,z)$ به شکل $p(x,y,z)=p(x)p(z|x)p(y|z)$ است.

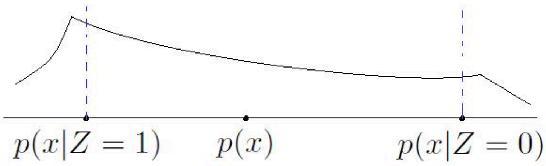
نکته ۲ از تعریف واضح است که

$$\mathbb{C}_{p(z|x)}[f](p(x)) \geq \mathbb{C}[f](p(x))$$

و بصورت مشابه

$$\mathbb{K}_{p(z|x)}[f](p(x)) \leq \mathbb{K}[f](p(x)).$$

در اینجا یک فرض ساده کننده میکنیم: فرض میکنیم که برای $p(x)$ های مورد علاقه ما نقاط پخش شدگی توسط کانال $p(z|x)$ (یعنی مجموعه نقاط $p(x|z)$ برای مقادیر مختلف z) رئوس یک چند وجهی (یا چند ضلعی) را تشکیل میدهند (یعنی هیچکدامشان قابل بیان بصورت ترکیب محدب بقیه نباشد؛ هیچکدام درون پوش بقیه قرار نگیرد). با توجه به این شرط در بخش بعدی ثابت میکنیم که معنی هندسی پوش محدب گیری تحدید شده به شرح زیر است: برای هر نقطه $p(x)$، پخش شدگی توسط کانال $p(z|x)$ را در نظر بگیرید: یعنی مجموعه نقاط $p(x|z)$ برای مقادیر مختلف z. پوش محدب تشکیل شده توسط این نقاط در سیمپلکس احتمالاتی را در نظر بگیرید. این پوش محدب یک ناحیه بسته را تشکیل میدهد که حول نقطه $p(x)$ قرار دارد. فرض کنید که این ناحیه را با $\mathcal{Q}_{p(x)}$ نشان دهیم. نمودار تابع f را به این ناحیه محدود کنید. در نمودار محدود شده پوش محدب f در نقطه $p(x)$ را حساب کنید (یعنی تنها ترکیب محدب های نقاط درون ناحیه $\mathcal{Q}_{p(x)}$ را در نظر میگیریم). مقدار حاصل همان $\mathbb{C}_{p(z|x)}[f](p(x))$ میباشد.



شکل ۱: الفبای Z دودویی بوده و تنها دو نقطه $x \mapsto p(x|Z=0)$ و $x \mapsto p(x|Z=1)$ را باید در نظر گرفت. پوش محدب آنها پاره خطی است که آنها را به هم وصل میکند. نمودار تابع f روی این پاره خط نشان داده شده است.

مثال ۳ فرض کنید که الفبای Z دودویی باشد. در این صورت همانطور که در شکل ۱ نشان داده شده است تنها دو نقطه $x \mapsto p(x|Z=0)$ و $x \mapsto p(x|Z=1)$ را داریم که $p(x)$ در پاره خطی که آنها را به هم وصل میکند قرار میگیرید. پوش محدب دو نقطه $x \mapsto p(x|Z=0)$ و $x \mapsto p(x|Z=1)$ پاره خطی است که آنها را به هم وصل میکند. نمودار تابع f روی شکل نشان داده شده است. در اینجا تابع f در پاره خط واصل دو نقطه $x \mapsto p(x|Z=0)$ و $x \mapsto p(x|Z=1)$ محدب است، در نتیجه پوش محدب آن برابر با خود تابع است:

$$\mathbb{C}_{p(z|x)}[f](p(x)) = f(p(x)).$$

اما اگر پوش محدب گیری را بصورت غیر محدود شده و در خارج از پاره خط واصل دو نقطه $x \mapsto p(x|Z=0)$ و $x \mapsto p(x|Z=1)$ انجام دهیم، میبینیم که باید دو نقطه ابتدا و انتهای نمودار را به هم دیگر وصل کنیم تا پوش محدب محاسبه شود. پس

$$\mathbb{C}[f](p(x)) < f(p(x)) = \mathbb{C}_{p(z|x)}[f](p(x)).$$

مثال ۴ در صورتی که کانال ثابت $p(a,b,z|x)$ را داشته باشیم میتوان تابع زیر را در نظر گرفت:

$$f(p(x)) = I(X;A) - I(X;B).$$

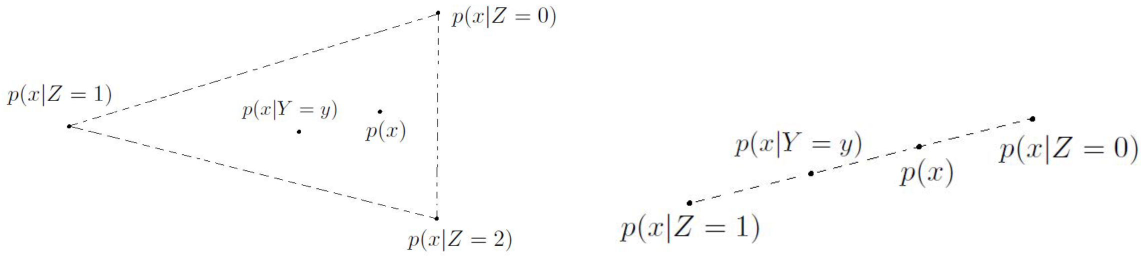
توزیع ثابت $p(x)$ را در نظر بگیرید. میتوان از بحث بالا برای محاسبه

$$\min_{Y:\, Y-Z-XAB} I(X;A|Y) - I(X;B|Y)$$

که در آن مینیمم روی تمامی متغیرهای Y که در زنجیره مارکف $A,B,X-Z-Y$ صدق میکنند استفاده کرد زیرا

$$\begin{aligned}\sum_y p(Y=y) f\big(p(x|Y=y)\big) &= \sum_y p(Y=y) I(X;A|Y=y) - I(X;B|Y=y)\\ &= I(X;A|Y) - I(X;B|Y).\end{aligned}$$

کافی است که نقاط $p(x|z)$ را مشخص کرده و فضای $\mathcal{Q}_{p(x)}$ را محاسبه کنیم. سپس تابع f را رسم کرده و پوش محدب آن را روی فضای $\mathcal{Q}_{p(x)}$ حساب کنیم. این کار از نظر محاسباتی به مراتب آسان تر از مینیمم گیری روی تمامی کانال ها میباشد. خصوصا اگر Z دودویی باشد.



شکل ۲: در شکل سمت راست الفبای Z دودویی بوده و نقطه $x \mapsto P(x|Y=y)$ برای هر مقدار y در پاره خطی که $x \mapsto p(x|Z=0)$ را به $x \mapsto p(x|Z=1)$ وصل میکند قرار میگیرد. در شکل سمت چپ الفبای Z سه تایی میباشد. نقطه $x \mapsto P(x|Y=y)$ برای هر مقدار y محدود به مثلث تشکیل شده است.

نکته ۵ اگر برای $p(a,b,x|z)$ در مثال بالا تعریف کنیم

$$g(p(z)) = I(X;A) - I(X;B).$$

آنوقت پوش محدب غیر تحدید شده g برابر با پوش محدب شده f نسبت به $p(z|x)$ است.

$$\mathbb{C}_{p(z|x)}[f](p(x)) = \mathbb{C}[g](p(z)) = \min_{Y:\ Y-Z-XAB} I(X;A|Y) - I(X;B|Y).$$

۱ اثبات روش هندسی محاسبه پوش محدب تحدید شده

حال روش هندسی محاسبه پوش محدب تحدید شده که در بالا بیان شد را ثابت میکنیم: توجه کنید که

$$p(x|y) = \sum_z p(x,z|y) = \sum_z p(z|y)p(x|y,z) = \sum_z p(z|y)p(x|z)$$

یعنی نقطه $[x \mapsto p(x|y)]$ در ترکیب خطی نقاط $[x \mapsto p(x|z)]$ قرار میگیرد:

$$[x \mapsto p(x|y)] = \sum_z p(z|y)[x \mapsto p(x|z)].$$

یعنی برای هر $p(x)$، پخش شدگی توسط کانال $p(y|x)$ به ما نقاطی را بدست میدهد که هرکدام در پوش محدب نقاط مربوط به پخش شدگی $p(z|x)$ هستند. این موضوع در شکل ۲ نشان داده شده است. قضیه بعدی وارون این موضوع را نشان میدهد.

قضیه ۶ فرض کنید که توزیع ثابت $p(x)$ و کانال ثابت $p(z|x)$ داده شده باشد. در صورتی که نقاط دلخواه

$$p_1(x), p_2(x), p_3(x), \cdots, p_k(x)$$

را در پوش محدب نقاط $x \mapsto p(x|Z=z)$ برای مقادیر مختلف z انتخاب کنیم و به آنها وزنهای نامنفی

$$\omega_1, \omega_2, \omega_3, \cdots, \omega_k$$

با مجموع یک نسبت دهیم، بطوریکه مرکز ثقل آنها $p(x)$ باشد:

$$p(x) = \sum_i \omega_i p_i(x).$$

آنوقت کانالی مانند $p(y|z)$ وجود دارد که نقاط انتخابی همان پراکندگی مربوط به کانال $p(y|x)$ باشند.

اثبات: از آنجایی که $p_i(x)$ در پوش محدب نقاط $x \mapsto p(x|Z=z)$ برای مقادیر مختلف z است، فرض کنید که

$$p_i(x) = \sum_z \lambda_{z,i} \cdot p(x|z),$$

که در آن $\lambda_{z,i} \geq 0$ و

$$\sum_z \lambda_{z,i} = 1 \qquad \forall i.$$

متغیر تصادفی Y روی مجموعه $\{1, 2, 3, \cdots, k\}$ را اینگونه تعریف میکنیم

$$p(Y=i) = \omega_i.$$

سپس متغیر Z را با استفاده از کانال $p(z|y)$ زیر تعریف میکنیم:

$$p_{Z|Y}(z|i) = \lambda_{z,i}.$$

و سپس متغیر X را با گذراندن Z از کانال $p_{X|Z}$ تعریف میکنیم. ادعا میکنیم که توزیع القایی روی X همان $p(x)$ داده شده است و همچنین

$$p(x|Y=i) = p_i(x).$$

ابتدا توجه کنید که

$$\begin{aligned} Prob(X=x) &= \sum_{i,z} \omega_i \cdot \lambda_{z,i} \cdot p_{X|Z}(x|z) \\ &= \sum_i \omega_i \sum_z \lambda_{z,i} \cdot p_{X|Z}(x|z) \\ &= \sum_i \omega_i p_i(x) \\ &= p(x). \end{aligned}$$

همچنین داریم:

$$\begin{aligned} Prob(X=x|Y=i) &= \sum_z \lambda_{z,i} \cdot p_{X|Z}(x|z) \\ &= p_i(x). \end{aligned}$$

تا اینجا متغیر های X, Y, Z را ساختیم که $X \sim p(x)$ و همچنین $p(x|z)$ همان توزیع $p_{X|Z}$ داده شده است. آیا از روی اینها میتوان نتیجه گرفت که $p(z|x)$ همان توزیع $p_{X|Z}$ داده شده است؟ تحت چه شرایطی $p(x)$ و $p(x|z)$ بصورت یکتا توزیع $p(z|x)$ را مشخص میکند؟ در اینجا نیاز داریم که از فرض ساده کننده استفاده کنیم: برای $p(x)$ های مورد علاقه ما نقاط پخش شدگی توسط کانال $p(z|x)$ (یعنی مجموعه نقاط $p(x|z)$ برای مقادیر مختلف z) رئوس یک چند وجهی (یا چند ضلعی) را تشکیل میدهند (یعنی هیچکدامشان قابل بیان بصورت ترکیب محدب بقیه نباشد؛ هیچکدام درون پوش بقیه قرار نگیرد). توجه کنید که $p(x) = \sum_z p(z)p(x|z)$ در نتیجه $p(x)$ مرکز ثقل نقاط پخش شدگی با وزن های $p(z)$ است. اگر نقاط پخش شدگی رئوس یک چند وجهی باشند، نحوه وزن گذاری روی $p(x|z)$ برای رسیدن به $p(x)$ یکتا است. در نتیجه $p(z)$ و $p(x, z) = p(z)p(x|z)$ بصورت یکتا مشخص میشوند.

$\square$