

فصل ۳

استنباط شبکه تنظیم ژنی از داده‌های زیستی

۱.۳ فیزیک آماری وارون

هدف اصلی فیزیک آماری به دست آوردن کمیت‌های مشاهده‌پذیر از قوانین میکروسکوپی حاکم بر اجزا سیستم است. به عنوان مثال، مدل آیزینگ (یک مدل ریاضی از فرومغناطیس) توسط هامیلتونی زیر توصیف می‌شود:

$$H(\{s_i\}) = - \sum_i h_i s_i - \sum_{i < j} J_{ij} s_i s_j \quad (1.3)$$

منظور از s_i یک آهن‌ربای کوچک یا اسپین است که مقدار ۱ یا -۱ به خود می‌گیرد. h_i میدان خارجی اعمال شده بر s_i و J_{ij} قدرت برهمکنش یا ضریب جفت‌شدگی بین دو اسپین i و j است. در مدل آیزینگ معمولی فرض بر این است که اسپین‌ها روی شبکه منظمی قرار دارند و هر اسپین فقط با همسایه‌های مرتبه اول خود دارای برهمکنش است. برای مثال برای مدل آیزینگ دو بعدی روی شبکه مربعی، برای هر اسپین i ، مقدار J_{ij} بین آن اسپین و اسپین بالایی و پایینی، چپ و راستی برابر با مقدار ثابت J است و برای سایر همسایه‌ها مقدار آن صفر است.

هدف ما در فیزیک آماری این است که با استفاده از مدل فوق که توصیف کننده برهمکنش بین اجزا است، کمیت‌های مشاهده‌پذیری چون مغناطش یا طول هم‌بستگی را بیابیم. از آن‌جا که میدان خارجی اعمال شده (جفت‌شدگی میدان و اسپین) و مشخصات فیزیکی بلور (جفت‌شدگی بین اسپین‌ها) را می‌دانیم، پارامترهای مدل که بردار h و ماتریس J هستند از قبل مشخص هستند.

در مسئله آیزینگ وارون، نقطه شروع داشتن مشاهده‌پذیرهای سیستم مانند مغناطش و همبستگی‌ها است و هدف ما یادگرفتن یا استنباط میدان خارجی و برهمکنش بین اجزا است. به عنوان مثال اگر بخواهیم بلوری با مشخصات خاصی

بسازیم، هدف ما پیدا کردن پتانسیل‌های مشخص بین اتم‌هاست که منجر به شکل‌گیری یک شبکه خاص شود [۲۶] یا مثلاً در فیزیک ماده چگال به دنبال پیدا کردن پارامترهای یک هامیلتونی هستیم که منجر به یک ماتریس چگالی خاص شود [۲۷].

در دو دهه گذشته، مسائل فیزیک آماری وارون در زمینه‌های مختلفی توجه متخصصان را به خود جلب کرده است. دلیل این جلب توجه این است که اخیراً از یک سو مقیاس‌های میکروسکوپی از لحاظ تجربی در دسترس قرار گرفته‌اند و از سوی دیگر، توانایی ذخیره‌سازی حجم زیادی از داده‌ها فراهم گشته است. به‌طور خاص در علوم زیستی، مسائل وارون زیادی در حوزه بازسازی شبکه‌های عصبی و تنظیم ژنی و مشخص کردن ساختار ۳ بعدی پروتئین مطرح شده است. در حوزه ماده چگال نرم، مسئله‌ای وجود دارد که چگونه یک سیستم بس‌ذره‌ای طراحی کنیم که به‌طور خودسامانده در یک حالت فضایی خاص قرار گیرد یا این‌که خواص حجمی خاصی داشته باشد [۲۸]، [۲۹].

در تمام مثال‌های فوق، متغیر «اسپین» توصیف‌کننده درجه آزادی میکروسکوپی مخصوص آن سیستم است. برای مثال، در یک شبکه عصبی، اسپین وضعیت یک نورون (سلول عصبی) را مشخص می‌کند. در ساده‌ترین حالت، این درجات آزادی، متغیرهای دوتایی تصادفی هستند که به مدل آیزینگ معمولی منجر می‌شوند. با این وجود این متغیرها می‌توانند مقادیر پیوسته‌ای به خود بگیرند و محدود به شبکه‌های منظم هم نباشند که در این صورت مسئله تبدیل به مدل‌های آیزینگ تعمیم یافته یا مدل شیشه-اسپینی می‌شود.

به‌طور کلی، انگیزه اصلی مسائل در فیزیک آماری وارون، به خاطر چالش کلان داده^۱ در حوزه‌های مختلف مخصوصاً وجود اندازه‌گیری‌هایی با توان بالا در زیست‌شناسی است. در مسائل وارون، به‌جای محاسبه مشاهده‌پذیرها براساس پارامترهای مدل، به دنبال استنباط پارامترهای مدل براساس مشاهده‌پذیرها هستیم.

۲.۳ روش‌های داده‌محور

پیش از آن که سراغ چگونگی استنباط پارامترهای یک مدل براساس مشاهده‌پذیرها رویم، مروری بر روش‌های دم‌دستی داده‌محور بر این مسائل می‌کنیم.

^۱ big data

۱.۲.۳ شبکه‌های همبستگی

یکی از ساده‌ترین معیارهایی که می‌توانیم برای دو اندازه‌گیری که به صورت برداری از مقادیر گزارش شده است داشته باشیم، همبستگی بین آن دو است. برای دو بردار یا آرایه X و Y همبستگی پیرسون بین آن دو برابر خواهد بود با:

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (۲.۳)$$

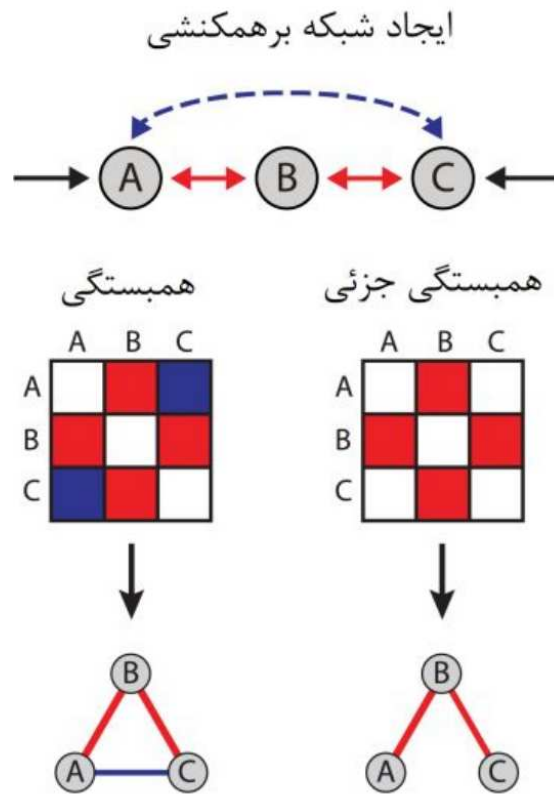
که در آن E عملگر امید ریاضی (مقدار متوسط)، cov به معنای کوواریانس، corr نماد معمول برای همبستگی (کورلیشن) پیرسون، و سیگما نماد انحراف معیار است.

در عمل هنگامی که N اندازه‌گیری برای بیان G ژن در شرایط مختلف داشته باشیم، آن‌گاه داده‌ها را می‌توانیم در ماتریس $D \in \mathbb{R}^{N \times G}$ قرار دهیم. با محاسبه همبستگی بین ستون‌های ماتریس D به ماتریسی $G \times G$ از همبستگی‌های دوتایی از ژن‌ها می‌رسیم. درایه‌های این ماتریس، وزن یال‌های شبکه همبستگی است که با در نظر گرفتن حد آستانه معقولی می‌توانیم همبستگی‌های ضعیف را دور بریزیم و به شبکه‌ای که توصیف‌کننده ارتباطات بین ژن‌ها است برسیم.

پیچیدگی یک روش یا الگوریتم، تعداد گام‌های مورد نیاز برای حل یک نمونه از یک مسئله به عنوان تابعی از اندازه ورودی به وسیله کارآمدترین الگوریتم می‌باشد. شبکه‌های همبستگی، در عمل بسیار ساده به دست می‌آیند. پیچیدگی آن‌ها به صورت خطی نسبت به تعداد آزمایش‌ها و به صورت مربعی با تعداد ژن‌ها رشد می‌کند. بنابراین برای مطالعاتی که تعداد آزمایش‌ها در آن زیاد است، گزینه بسیار ساده‌ای برای به کارگیری می‌باشند. این فرض که ژن‌های در حال برهمکنش باید مقدار بیان همبسته‌ای داشته باشند از لحاظ زیستی معقول است. نشان داده شده که روش‌هایی مانند «تحلیل شبکه هم‌بیانی وزن‌دار» [۳۰] قابل اعتماد هستند و به طور گسترده‌ای نیز به کار گرفته شده‌اند.

با این وجود شبکه‌های همبستگی محدودیت‌هایی نیز دارند. نخست این که دو ژن می‌توانند به ظاهر همبسته باشند ولی نه به این خاطر که به طور خالص بر یکدیگر اثر می‌گذارند بلکه ممکن است وجود همبستگی بین آنان به خاطر اثر ژن یا ژن‌های واسط بین آن دو باشد. برای حذف اثرات وابسته باید همبستگی جزئی بین دو ژن را حساب کرد. در شکل ۱.۳ ژن A و C به طور مستقیم بر یکدیگر اثر نمی‌گذارند با این وجود، همبستگی این دو ژن به خاطر وجود ژن واسط B مقداری غیرصفر است.

از سوی دیگر از آن‌جا که همبستگی متقارن است یعنی $\text{corr}(X, Y) = \text{corr}(Y, X)$ ، شبکه همبستگی شبکه جهت‌داری نیست. مشکل دیگر شبکه‌های همبستگی در این است که در آزمایش‌هایی که در مقیاس‌های نه‌چندان بزرگ انجام می‌شوند، همبستگی به دست آمده از داده‌ها، مقادیر بزرگی را نشان می‌دهند که از لحاظ آماری ارزش چندانی ندارند. از آن‌جا که همبستگی پیرسون، معیار خطی از همبستگی بین دو متغیر تصادفی است، اثرات تنظیمی غیرخطی ژن‌ها بر یکدیگر ممکن است توسط شبکه‌های همبستگی نمایش داده نشوند. بنابراین می‌بینیم که شبکه همبستگی می‌تواند برای



شکل ۱.۳: ماتریس همبستگی می‌تواند اطلاعات نادرستی در مورد اثرگذاری خالص اجزا سیستم بر روی یکدیگر را نشان دهد. برای حذف اثرات کاذب، باید به جای ماتریس همبستگی، ماتریس همبستگی جزئی را محاسبه کنیم. [۳۱]

هر یال، وزنی را پیشنهاد کند که یا از مقدار واقعی بیشتر است و یا کمتر. بنابراین این روش، اطلاعات گمراه‌کننده‌ای برای پیدا کردن درک درستی از سیستم پیشنهاد می‌کند.

۲.۲.۳ روش‌های مبتنی بر نظریه اطلاعات

اگر X و Y دو متغیر تصادفی باشند و $P(X, Y)$ توزیع احتمال مشترک آن‌ها باشد، آنگاه اطلاعات متقابل بین آن دو به صورت زیر تعریف می‌شود:

$$MI[X, Y] = \sum_{x_i, y_j} P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)} = \sum_{x_i, y_j} P(x_i, y_j) \log \frac{P(x_i|y_j)}{P(x_i)} \quad (۳.۳)$$

که x_i و y_j مقادیری هستند که دو متغیر تصادفی می‌توانند اختیار کنند. $P(X)$ و $P(Y)$ احتمال‌های حاشیه‌ای^۲ X و Y هستند که به ترتیب با جمع زدن توزیع مشترک روی Y و X به دست می‌آیند. به طور شهودی، اطلاعات متقابل، درجه وابستگی دو متغیر تصادفی را به طور کمی نمایش می‌دهد؛ اطلاعات متقابل صفر است اگر دو متغیر مستقل از هم باشند و برای زمانی که دو متغیر به طور تعینی به یکدیگر مرتبط باشند، اطلاعات متقابل، برابر با آنتروپی توزیع حاشیه‌ای می‌باشد. اطلاعات متقابل مانند همبستگی پیرسون معیاری متقارن است، برای همین شبکه‌های مبتنی بر اطلاعات متقابل نیز، بدون جهت هستند. با این وجود، این معیار توانایی بیشتری در نمایش وابستگی‌های غیرخطی دارد و به همین خاطر کاربردهای بیشتری نسبت به شبکه‌های همبستگی دارد. ضعف این روش نسبت به شبکه‌های همبستگی در این است که پیچیدگی محاسباتی این روش نسبت به تعداد آزمایش‌ها و ژن‌ها به صورت مربعی رشد می‌کند. از طرف دیگر، هنگامی که اندازه نمونه کوچک باشد، به دست آوردن توزیع مشترک که در رابطه فوق وجود دارد می‌تواند نسبت به نوفه (نویز) بسیار حساس باشد.

۳.۲.۳ روش‌های مبتنی بر رگرسیون

یکی از راه‌های به دست آوردن وابستگی دو متغیر، پیش‌بینی کردن یکی از روی دیگری است. در ساده‌ترین حالت، می‌توان از یک رهیافت رگرسیون خطی استفاده کرد به این شکل که شیب خط رگرسیون می‌تواند برای مشخص کردن قدرت وابستگی استفاده شود. در شبکه تنظیم ژنی، می‌خواهیم مقدار بیان ژنی را نسبت به بقیه چنان تغییر دهیم که وزن‌های شبکه به دست آید. بنابراین، برای هر ژن g که مقدار بیان آن در نمونه i ام را با x_{gi} نمایش می‌دهیم، این مسئله رگرسیون را با

^۲Marginal Probabilities

E_i که یک جمله نوفه‌ای (نویز) است را می‌خواهیم حل کنیم:

$$x_{gi} = \sum_{j \neq g} w_j x_{ji} + \varepsilon_i \quad (۴.۳)$$

و وزن w_j به‌دست آمده را به عنوان وزن یال متصل کننده ژن j و g استفاده کنیم. این رهیافت، به ما شبکه‌ای جهت‌دار می‌دهد. این ایده هسته مرکزی روشی به اسم TIGRESS است [۳۲]. راه دیگر این است که مدل رگرسیون خطی در معادله فوق را با مدلی منعطف‌تر و ناپارامتری جایگزین کرد. روش‌های مبتنی بر رگرسیون، برای کار با داده‌های سری زمانی نیز بسیار مشهور هستند. به این صورت که بیان ژن g در زمان t توسط سایر ژن‌ها در زمان $t-1$ تغییر می‌کند (مدل خودهمبسته) [۳۳].

روش‌های رگرسیون محور رهیافت‌های بسیار مشهوری برای بازسازی شبکه‌های جهت‌دار در مقیاس‌های مختلف هستند. این روش‌ها در مقایسه با سایر روش‌های داده‌محور، پیچیدگی محاسباتی بالاتری دارند، با این وجود دارای قابلیت پیش‌بینی هستند. به این معنا که اگر زیرمجموعه‌ای از بیان ژن‌ها را داشته باشیم، در اصول می‌توانیم مقدار بیان سایر ژن‌ها را پیش‌بینی کنیم. افزون بر این، روش‌های رگرسیون محور دارای این توانایی هستند که وابستگی‌های شرطی مرتبه بالا را میان الگوی بیان ژن‌ها پیدا کنند. در صورتی که روش‌های مبتنی بر همبستگی یا اطلاعات متقابل تنها به وابستگی‌های دوتایی بسنده می‌کنند. در عمل، شناسایی‌پذیر بودن مدل‌های رگرسیون از داده‌های محدود مشکل‌ساز است! شناسایی‌پذیری، ویژگی است که یک مدل باید ارضا کند تا استنباط دقیق ممکن باشد. به عبارت دیگر، مدلی شناسایی‌پذیر است که به‌طور نظری بتوان پس از به دست آوردن بی‌شمار مشاهده از آن، مقادیر درست پارامترهای آن را یادگرفت. ژن‌های مختلف معمولاً الگوهای بیان به‌شدت همبسته‌ای دارند و رگرسیون با متغیرهای کمکی^۳ همبسته، به‌شدت متمایل به نتایج جعلی است!

۳.۳ شبکه اسپینی و مدل گاوسی چند متغیره

تمام روش‌های داده‌محوری که تاکنون مورد بحث قرار گرفت، همگی با معیاری از وابستگی شروع می‌شدند ولی هیچ‌کدام به طور خاص مدلی از داده‌ها را براساس احتمالات معرفی نمی‌کردند. قصد ما در این است که شبکه تنظیمی ژن‌ها را به صورت یک شبکه اسپینی ببینیم و با در نظر گرفتن یک مدل فیزیکی برای برهمکنش‌ها، به دنبال استنباط پارامترهای آن مدل برآییم.

اگر فرض کنیم که هر ژن، یک اسپین و مقدار هر اسپین برابر با مقدار بیان یک ژن است و مدل حاکم بر ژن‌ها یک

^۳covariate

مدل دوتایی است، خواهیم داشت:

$$H(\{s_i\}) = - \sum_{i < j} J_{ij} s_i s_j \quad (۵.۳)$$

که این همیلتونی مدل شیشه-اسپینی در صورت نبود میدان خارجی است. $\{s_i, i = 1, \dots, n\}$ داده‌های تجربی موجود از مقدار بیان ژن‌ها و J_{ij} ضریب جفت‌شدگی ژن i به ژن j که همان برهمکنش تنظمی بین این دو ژن می‌باشد. هدف ما استنباط ضرایب J_{ij} از داده‌های تجربی است. با توجه به همیلتونی فوق، توزیع تعادلی بولترمن برابر خواهد بود با:

$$P(\{s_i\}) = \frac{1}{Z} \exp[\sum_{i < j} J_{ij} s_i s_j] \quad (۶.۳)$$

که در آن Z تابع پارش است و دما را در دل جفت‌شدگی‌ها جا داده‌ایم. توزیع فوق همان توزیعی است که انتروپی گیس $S[P]$ را تحت شرایط بهنجار بودن $P(\{s_i\})$ و دارا بودن گشتاور اول و دوم، بیشینه می‌کند.

$$S[P] = - \sum_{(s_i)} P(\{s_i\}) \ln(P(\{s_i\})) \quad (۷.۳)$$

در حقیقت، فرض ما در این مدل‌سازی این است که سیستم ما دارای برهمکنش‌های دوتایی بدون جهت است و خود را محدود به مدلی با دو پارامتر و با متغیرهایی با مقادیر پیوسته و حقیقی کرده‌ایم. بیشینه کردن آنتروپی به این خاطر است که می‌خواهیم کمترین اطلاعات اضافی ممکن را در مدل‌سازی خود داشته باشیم و کلی‌ترین حالت ممکن را استنباط کنیم. از آن‌جا که از داده‌های تجربی مقادیر s_i و حاصل ضرب $s_i s_j$ قابل محاسبه است، گشتاور اول و دوم این داده‌ها را می‌توانیم محاسبه کنیم. یادآوری می‌شود که ابتدا یک مدل دوتایی (شیشه-اسپینی) را انتخاب کردیم و اکنون به دنبال استنباط ضرایب J_{ij} [۳۴] از طریق بیشینه کردن آنتروپی تحت قیود زیر هستیم:

۱. بهنجارش:

$$\int_S P(S) dS = 1 \quad (۸.۳)$$

۲. برابر بودن گشتاور اول مدل و داده‌های تجربی (می‌خواهیم متوسطی که مدل تولید می‌کند با مقدار متوسط محاسبه شده از داده‌های تجربی باید یکسان باشد):

$$\langle s_i \rangle = \int_S s_i P(S) dS = \frac{1}{N} \sum_{m=1}^N s_i^m = \overline{s_i} \quad (9.3)$$

۳. برابر بودن گشتاور دوم مدل و داده‌های تجربی:

$$\langle s_i s_j \rangle = \int_S s_i s_j P(S) dS = \frac{1}{N} \sum_{m=1}^N s_i^m s_j^m = \overline{s_i s_j} \quad (10.3)$$

اکنون با استفاده از روش ضرایب لاگرائز نشان داده می‌شود که [۳۱]، [۳۵]:

$$P(S, \langle S \rangle, C) = \frac{1}{Z} \exp[-(S - \langle S \rangle)^T C^{-1} (S - \langle S \rangle)] \quad (11.3)$$

که در آن:

$$Z = \frac{(2\pi)^{-L}}{2 \det(2C)^{-\frac{1}{2}}}, J = -C^{-1} \quad (12.3)$$

و N تعداد اسپین‌ها و L تعداد نمونه‌ها است. به وارون ماتریس کواریانس، ماتریس دقت گفته می‌شود. هر درایه i, j از ماتریس دقت، همبستگی جزئی بین ژن i و ژن j را نشان می‌دهد. همبستگی جزئی نشان دهنده همبستگی باقی‌مانده بین دو متغیر است هنگامی که اثرات متغیرهای واسطه حذف شده باشد. بنابراین برای به دست آوردن ضرایب J_{ij} که معادل با برهمکنش بین ژن i و ژن j است، کافی است که منفی وارون ماتریس کواریانس که از داده‌های تجربی به دست آورده‌ایم را محاسبه کنیم. روش‌های این‌چنینی عضو کلاس بزرگی از روش‌های معروف به مدل‌های گاوسی گرافیکی هستند [۳۶]. ایده این‌گونه روش‌ها این است که اندازه‌گیری‌های بیان ژن را به عنوان یک بردار تصادفی گاوسی چندمتغیره در نظر بگیریم که هر عضو از این بردار شامل مقدار بیان یک ژن می‌باشد. سپس با توجه به شرایط مسئله با استفاده از برآورد درست‌نمایی بیشینه، ماتریس دقت را تخمین می‌زنیم. برآورد حداکثر درست‌نمایی (که به‌طور خلاصه به آن MLE نیز گفته می‌شود) روشی است برای برآورد کردن پارامترهای یک مدل آماری. وقتی بر مجموعه‌ای از داده‌ها عملیات

انجام می‌شود یک مدل آماری به دست می‌آید آنگاه حداکثر درست‌نمایی می‌تواند تخمینی از پارامترهای مدل ارائه دهد. از آنجا که تعداد پارامترهایی که باید تخمین بزنم متناسب با مربع تعداد ژن‌هاست نیازمند روش‌های منظم‌سازی^۵ هستیم. برای این‌کار از الگوریتم لاسو گرافیکی^۶ استفاده می‌کنیم. لاسو گرافیکی الگوریتمی است که وارون ماتریس کواریانس یک توزیع چند متغیره گاوسی را حساب می‌کند. [۳۷] اگر از یک توزیع چند متغیره گاوسی با متوسط صفر و انحراف معیار سیگما $(X \sim N(0, \Sigma))$ ، n مشاهده به صورت X_1, X_2, \dots, X_n داشته باشیم و بخواهیم وارون ماتریس سیگما را تخمین بزنیم، آنگاه تخمین‌زن لاسو گرافیکی نتایی است که:

$$\hat{\Theta} = *argmin_{\Theta \geq 0} \left(tr(S\Theta) - \log[det(\Theta)] + \lambda \sum_{j \neq k} |\Theta_{jk}| \right) \quad (13.3)$$

که S کواریانس داده‌های تجربی و λ ضریب جریمه است [۳۷]. لاسو گرافیکی با تک‌کردن ماتریس به این صورت که درایه‌های نزدیک به صفر را به صفر تبدیل می‌کند (جریمه کردن) محاسبه ماتریس کواریانس و وارون آن را از روی نمونه‌های تجربی انجام می‌دهد. مشکل مدل‌های گاوسی گرافیکی در این است که داده‌ها باید بهنجار باشند که برای این مطالعه تقریب معقولی می‌باشد.

وجود همبستگی، شرط لازم و نه کافی برای ظهور یک حرکت جمعی است. شرط کافی برای حرکت جمعی در چنین شبکه‌ای مربوط می‌شود به چگونگی قرار گرفتن دوتایی‌ها در کنار یکدیگر. نگاه تقلیل‌گرایانه به یک سیستم پیچیده، تحلیل درستی به ما نمی‌دهد و برای درک سیستم نیاز داریم که ساختارهای بزرگ‌تر از اجزا سازنده سیستم، مانند دوتایی‌ها را در نظر بگیریم! بنابراین، شیوه‌ی قرارگیری عناصر ماتریس J که ایده‌ای از چگونگی کنار هم نشستن هر دو ژن در این سیستم است بسیار مهم می‌باشد. به طور خلاصه، حرکت جمعی حاصل از این سیستم بس‌ذره در الگوی قرار گرفتن درایه‌های ماتریس برهمکنش J نشسته است.

ماتریس C^{-1} اطلاعات مربوط به همبستگی جزئی متغیرها را نشان می‌دهد. در حقیقت این ماتریس معیاری است از اینکه متغیرها چقدر نزدیک مقدار متوسط (عناصر قطری) به‌طور متراکم، قرار دارند و تا چه مقدار با سایر متغیرها (عناصر غیرقطری) هم‌زمان تغییر نمی‌کنند. از سوی دیگر، از آنجا که C ماتریس کواریانس یک توزیع چندمتغیره گاوسی است، $C_{ij}^{-1} = 0$ اگر و تنها اگر در صورت داشتن همه داده‌ها، ژن i و ژن j به طور مشروط مستقل باشند.

اما معنی ماتریس $J = -C^{-1}$ چیست؟ در حقیقت، J ماتریس همسایگی شبکه است. به این معنی که اگر هر ژن را یک راس در شبکه در نظر بگیریم، یال بین دو ژن i و j وزنی معادل J_{ij} خواهد داشت. بدیهی است که اگر دو ژن برهمکنشی با یکدیگر نداشته باشند، در شبکه برهمکنشی پیوندی بین آن دو نیست و $J_{ij} = 0$. اکنون می‌توانیم از روی

داده تجربی، از روی نمونه‌های سالم، شبکه سالم و از روی نمونه‌های سرطانی شبکه سرطانی را بازسازی می‌کنیم. در هر دو شبکه، رئوس ژن‌ها هستند و یال‌های بین رئوس نشان‌دهنده وجود برهمکنش بین آن‌هاست. اندازه برهمکنش برابر با وزن یال است. اگر یالی وزنش مثبت باشد یعنی دو راس متصل به آن یال، توسط اثرات تنظیمی، سبب بیشتر بیان شدن یکدیگر می‌شوند و اگر یالی وزنش منفی بود، دو راس متصل به آن سبب کاهش مقدار بیان یکدیگر می‌شوند.

۴.۳ سایر کاربردها

از روش‌های فیزیک آماری وارون، همانند تکنیکی که برای استنباط شبکه‌های تنظیم ژن معرفی شد در مسائل مختلف در حوزه‌های گوناگون استفاده می‌شود. به عنوان مثال، برای مدل‌سازی الگوی آتش‌کردن^۷ سلول‌های عصبی و بازسازی ارتباطات شبکه عصبی [۳۸]-[۴۲] تعیین ساختار پروتئین‌ها، استنباط برهمکنش بین گونه‌ها و افراد [۴۳]-[۴۶] و استنباط شبکه‌های مالی [۴۷]-[۴۹] از ایده‌های فیزیک آماری وارون استفاده می‌شود.