شبکه‌های عصبی: جلسه هشتم و آخر

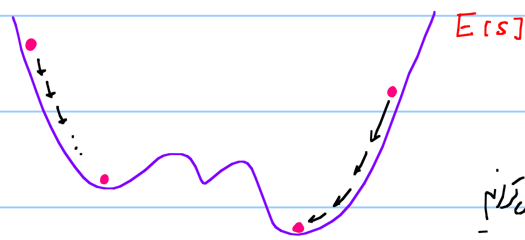
۱_ مقدمه: در بخش‌های گذشته دیدیم که دینامیک معین یک شبکه عصبی یعنی رابطه

$$S_i(t+1) = \mathrm{sgn}(h_i(t)) \qquad ①$$

همواره یک تابع هزینه یا انرژی یعنی

$$E[S] := \frac{-1}{2} \sum_{i,j} \omega_{ij} S_i S_j \qquad ②$$

را کاهش می‌دهد، یعنی هرگاه از یک نقطه دلخواه شروع کنیم، به یک کمینه موضعی از تابع انرژی می‌رسیم، مطابق شکل زیر:



اما برای نورون‌های استوکاستیک رابطه ① جای خود را به یک رابطه احتمالاتی می‌دهد. به این معنا که برای چنین نورون‌هایی تنها احتمالات گذار را می‌توانیم تعیین کنیم. یعنی خواهیم داشت:

$$W(S_i = -1 \longrightarrow S_i = 1) = \frac{1}{1 + e^{-2\beta h_i}} \qquad ③$$

و

$$W(S_i = 1 \longrightarrow S_i = -1) = \frac{1}{1 + e^{2\beta h_i}} \qquad ④$$

یا به عبارت فشرده‌تر خواهیم داشت:

$$W(S_i \longrightarrow -S_i) = \frac{1}{1 + e^{2\beta h_i S_i}} \qquad ⑤$$

حال از خود سوال می‌کنیم که با چنین احتمالات گذاری، احتمال این که شبکه عصبی در یک حالت اسپینی معین باشد چقدر است؟ برای پاسخ به چنین سوالی مثل همه فرایندهای تصادفی دیگر، به معادله مادر توجه می‌کنیم: حالت‌های نورونی را با C، C'، ... نشان می‌دهیم.

$$P(C,t+1) = \sum_{C'\neq C} P(C',t)\,W(C'\to C) + P(C,t)\,W(C\to C) \qquad (۶)$$

منظور از $W(C\to C)$ احتمال ماندن در حالت C است. یعنی

$$W(C\to C) = 1 - \sum_{C'\neq C} W(C\to C') \qquad (۷)$$

از ترکیب ۶ و ۷ به دست می‌آید:

$$P(C,t+1) - P(C,t) = \sum_{C'\neq C} \{P(C',t)\,W(C'\to C) - P(C,t)\,W(C\to C')\}. \qquad (۸)$$

در حالت تعادل که شرط توازن تفصیلی برقرار است داریم:

$$P(C',t)\,W(C'\to C) = P(C,t)\,W(C\to C') \quad \forall\, C, C'. \qquad (۹)$$

تحت این شرایط طرف چپ نیز صفر شده و تعادل برقرار شده یعنی هیچ چیز با زمان تغییر نمی‌کند. از رابطه ۹ به دست می‌آوریم:

$$\frac{P(s_1, s_2, \ldots, -s_1, \ldots s_N)}{P(s_1, s_2, \ldots, s_1, \ldots s_N)} = \frac{W(s_1\to -s_1)}{W(-s_1\to s_1)} \overset{(۵)}{=} e^{-2\beta h_1 s_1} \qquad (۱۰)$$

با تکرار رابطه ۱۰ به دست می‌آید که

$$P(s_1, s_2, s_3, \ldots s_N) = \frac{1}{Z}\, e^{\frac{1}{2}\beta[h_1 s_1 + h_2 s_2 + \ldots h_N s_N]} \qquad (۱۱)$$

که در آن Z یک ضریب ثابت است. بنابراین با توجه به این که

$$h_i = \sum_j \omega_{ij}\, s_j \qquad (۱۲)$$

$$\to P(s_1, s_2, s_3, \ldots s_N) = \frac{1}{Z}\, e^{\beta \frac{1}{2}\sum_{ij} \omega_{ij} s_i s_j} = \frac{1}{Z}\, e^{-\beta E[s_1, \ldots s_N]} \qquad (۱۳)$$

که در آن عبارت

$$E[s_1 \dots s_N] = \frac{-1}{2} \sum_{i,j} \omega_{ij}\, s_i s_j \qquad (۱۴)$$

را می‌توان به عنوان انرژی یک حالت نورونی در نظر گرفت.

حال که عبارت انرژی را برای شبکه عصبی پایستگی به دست آوردیم می‌توانیم از خود بپرسیم که در حالت تعادل این شبکه در کدام حالت قرار می‌گیرد. پاسخ این است که مثل هر دستگاه مکانیک آماری دیگری این سیستم با احتمال $P(E) = \frac{1}{Z} e^{-\beta E}$ در حالتی با انرژی E قرار می‌گیرد و در حالت تعادل می‌توان مقدار متوسط اسپین‌ها را حساب کرد. بعد برای محاسبه مقدار متوسط نیز می‌بایست تابع پارش را حساب کنیم. تابع پارش برابر است با:

$$Z = \sum_{s_1, s_2, \dots s_N} e^{\beta/2 \sum_{i,j} \omega_{ij} s_i s_j} \qquad (۱۵)$$

برای آنکه مقدار متوسط $\langle s_i \rangle$ را حساب کنیم، می‌بایست تابع پارش در حضور میدان را حساب کنیم. این میدان لزومی ندارد که یک میدان واقعی باشد. کافی است که به آن به عنوان یک عامل کمکی برای محاسبه نگاه کنیم: این تابع پارش عبارت است از:

$$Z[\beta, h] = \sum_{s_1 \dots s_N} e^{\beta/2 \sum_{i,j} \omega_{ij} s_i s_j + \sum_i h_i s_i} \qquad (۱۶)$$

۲- مدل آیزینگ قبل از آنکه تابع پارش (۱۵) و (۱۶) را حساب کنیم، یک تابع پارش ساده‌تر را حساب می‌کنیم. می‌دانیم که در یک شبکه عصبی که با قانون Hebb، p تا طرح را حفظ کرده است، w_{ij} ها برابر است با:

$$\omega_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \sigma_i^{(\mu)} \sigma_j^{(\mu)} \qquad (۱۷)$$

ما این مدل را بعداً بررسی می‌کنیم و حال مدل ساده‌تری را در نظر می‌گیریم که در آن w_{ij} ها همه با هم برابرند و مقدار آنها برابر است با: $\omega_{ij} = \frac{\omega}{N}$.

این مدل چیزی نیست جز مدل آیزینگ، برهمکنش بلندبرد، چرا که در آن همه اسپین‌ها با شدت مساوی با هم برهمکنش دارند.

انرژی هر حالت از اسپین‌ها در این مدل برابر است با:

$$E[s_1 \dots s_N] = \frac{-\omega}{2N} \sum_{i,j} s_i s_j \qquad (۱۸)$$

و تابع پارش برابر است با:

$$Z(\beta,h) = \sum_{s_1\cdots s_N} e^{\frac{\beta\omega}{2N}\sum_{i,j} s_i s_j + h\beta\sum_i s_i} \quad (19)$$

با تعریف $S = \sum_i s_i$ خواهیم داشت:

$$Z(\beta,h) = \sum_{s_1\cdots s_N} e^{\frac{\beta\omega}{2N} S^2 + h\beta S} \quad (۲۰)$$

حال می‌توانیم از ترفند ریاضی که خواص شوید از رابطهٔ انتگرال گاوسی استفاده کنیم: می‌دانیم که:

$$\int e^{-\frac{a}{2}x^2 + bx}\, dx = \sqrt{\frac{2\pi}{a}}\; e^{\frac{1}{2}\frac{b^2}{a}} \quad (۲۱)$$

قرار می‌دهیم: $b \longrightarrow S$, $a = \frac{N}{\beta\omega}$, $x \to m$

بنابراین

$$e^{\frac{\beta\omega}{2N}S^2} = \sqrt{\frac{N}{2\pi\beta\omega}} \int dm\; e^{-\frac{N}{2\beta\omega}m^2 + mS} \quad (۲۲)$$

در نتیجه ←

$$Z(\beta,h) = \left(\frac{N}{2\pi\beta\omega}\right)^{1/2} \sum_{s_1\cdots s_N} \int dm\; e^{-\frac{N}{2\beta\omega}m^2 + mS + \beta h S} \quad ۲۳$$

حال می‌توان جمع روی s_i ها را انجام داد:

$$\sum_{s_i} e^{(\beta h + m)s_i} = 2\cosh(\beta h + m).$$

$$\rightarrow \sum_{s_1 s_2 \cdots s_N} e^{(\beta h+m)S} = \left[2\cosh(\beta h+m)\right]^N \quad ۲۴$$

$$\rightarrow Z(\beta h) = \left(\frac{N}{2\pi\beta\omega}\right)^{1/2}\int dm\; e^{-\frac{N}{2\beta\omega}m^2 + N\ln[2\cosh(\beta h+m)]}$$ ۲۵

حال برای انتگرال فوق ما، بهترین تقریب، تقریب زینی است که. یادآوری می‌کنیم که این روش در حد $N\rightarrow\infty$، مسئله دقیق انتگرال را بدست می‌دهد. داریم:

$$I = \int dm\; e^{-N\beta f(m)} \simeq \int dm\; e^{-N\beta\{f(m_0)+\frac{1}{2}f''(m_0)(m-m_0)^2\}}$$

که در آن

$$\boxed{f(m) = \frac{m^2}{2\beta^2\omega} - \frac{1}{\beta}\ln[2\cosh(\beta h+m)].}$$

و m_0 نقطه‌ای است که در آن $f(m)$ مینیمم می‌شود: بنابراین

$$I = e^{-N\beta f(m_0)}\sqrt{\frac{2\pi}{N\beta f''(m_0)}}$$

$$\rightarrow Z(\beta,h) = \sqrt{\frac{1}{\beta\omega f''(m_0)}}\; e^{-N\beta f(m_0)}$$ ۲۶

سه نکته را باید بررسی کنیم: ① این که تعبیر فیزیکی $f(m)$ چیست؟ ② این که مسئله m_0 چقدر است؟

③ تعبیر فیزیکی m_0 چیست؟

گفت به نکته ① می‌پردازیم: می‌دانیم که رابطه انرژی آزاد هلمهولتز با تابع پارش به شکل زیر است:

$$F = -\frac{1}{\beta}\ln Z.$$

و اگر تابع انرژی آزاد بر واحد اسپین‌ها لازم باشد، که آن را با $\tilde{f}$ نشان می‌دهیم داریم:

$$\tilde{f} = \frac{F}{N} = \frac{-1}{N\beta}\ln Z$$

بنابراین ←
$$\tilde{f} = f(m_0).$$

دقت کنید که محاسبات بالا را در حد ترمودینامیکی یعنی $N\to\infty$ حساب کردیم. بنابراین تابع انرژی آزاد چیزی نیست جز مقدار تابع $f(m_0)$. یعنی حداقل. تابع $f(m,h,T)$ را پیدا کنیم، می‌توانیم آن را به عنوان تابعی از m پیدا کنیم، سپس انرژی آزاد به عنوان تابعی از h, T پیدا می‌شود.

حال به نکتهٔ دوم می‌پردازیم: یعنی بفهمیم مقدار m_0 چقدر است!

می‌دانیم که m_0 از رابطهٔ $\frac{d}{dm}f(m)\Big|_{m=m_0}=0$ به دست می‌آید. می‌دانیم

$$f(m) = \frac{1}{2\beta^2 w}m^2 - \frac{1}{\beta}\ln\left[2\cosh(\beta h+m)\right] \qquad ۲۷$$

$$\rightarrow\quad f'(m) = \frac{m}{\beta^2 w} - \frac{1}{\beta}\tanh(\beta h+m). \quad\rightarrow$$

$$\boxed{m_0 = \beta w\tanh(\beta h+m_0).} \qquad ۲۸$$

رابطهٔ بالا را تنها می‌توان به روش ترسیمی حل کرد. اما قبل از حل این معادله بهتر است که معنای فیزیکی و شهودی m_0 را بفهمیم. با استفاده از تابع پارش می‌توان مقدار متوسط $\langle s_i\rangle$ یعنی مغناطش در هر اسپین را حساب کنید. رابطهٔ ۱۹، روش این کار را به ما یاد می‌دهد:

$$(19)\rightarrow\quad \sum_i\langle s_i\rangle = \frac{1}{Z}\frac{\partial}{\partial h\beta}Z = \frac{\partial}{\partial h\beta}\ln Z \quad\rightarrow\quad \langle s_i\rangle = \frac{1}{N}\frac{\partial}{\partial h\beta}\ln Z.$$

از طرفی می‌دانیم که در حد $N\to\infty$،

$$\frac{1}{N}\ln Z = -g(m_0) \quad\text{، بنابراین}\leftarrow \qquad ۲۹$$

$$\langle s_i\rangle = -\frac{\partial}{\partial h}g(m_0). \qquad ۳۰$$

اما باید دقت کنیم که g را باید به شکل دقیق‌تر نوشت:

$$g = g(m_0(h),h).$$

یعنی اینکه g یک بستگی صریح به h دارد و یک بستگی ضمنی از طریق m_0. دقت کنید m_0 از طریق رابطه ۳۱، خودش تابعی از h است.

بنابراین →
$$\langle s_i \rangle = \left(\frac{\partial g}{\partial m}\right)_{m_0} \frac{\partial m_0}{\partial h} + \frac{\partial g(m,h)}{\partial h}\Big|_{m=m_0} \qquad ۳۱$$

اما $\left(\frac{\partial g}{\partial m}\right)_{m_0}$ برابر با صفر است. بنابراین

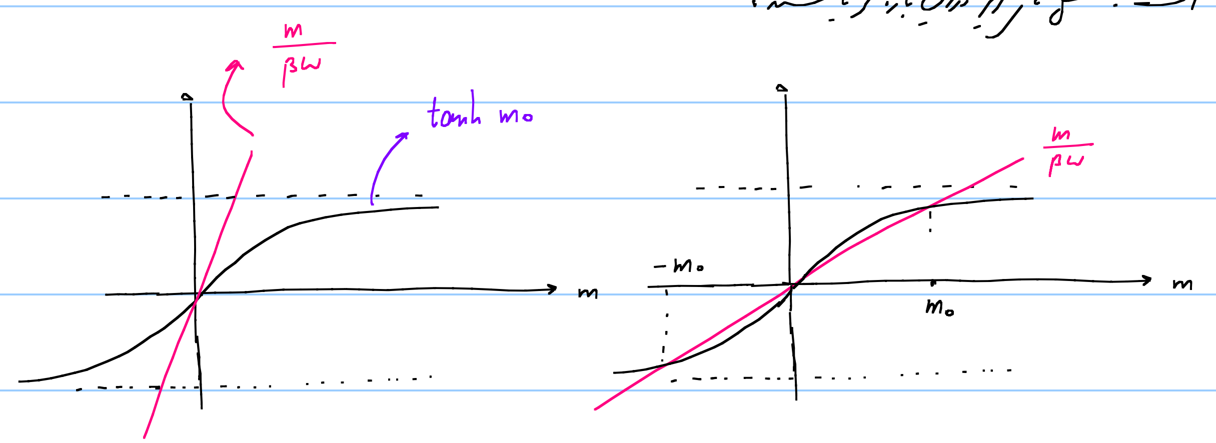
$$\langle s_i \rangle = \frac{\partial g}{\partial h}(m,h)\Big|_{m=m_0} = \tanh(\beta h + m_0) = \frac{m_0}{\beta w}. \qquad ۳۲$$

پس در واقع m_0 متناسب با مغناطش ترمودینامیکی است.

حال می‌توانیم مقدار m_0 را به روش ترسیمی از رابطه ۲۸ پیدا کنیم. معمولاً علاقه‌مندیم به یافتن میزان مغناطش خودبخودی یعنی مغناطش در نبود میدان مغناطیسی هستیم. بنابراین در رابطه ۲۸ می‌بایست قرار دهیم $h=0$. پس هدف ما حل ترسیمی معادله

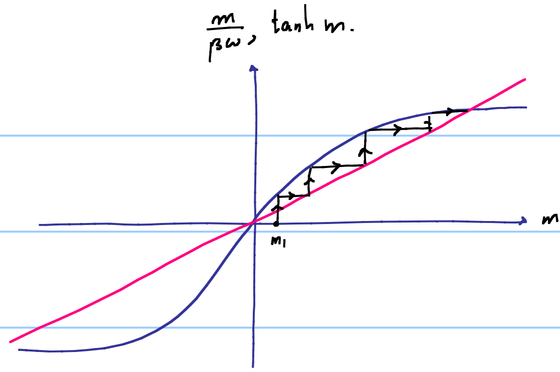
$$m_0 = \beta w \tanh m_0 \qquad ۳۳$$

است. شکل ۴ زیر در این باره راهگشا هستند.



$\beta w < 1$ $\qquad$ $\beta w > 1$

بنابراین دمای بحرانی از رابطه $\beta w = 1$ بدست می‌آید. یعنی دمای بحرانی عبارت است از $T_c = w$. برای $T < T_c$، سه جواب $+m_0$، $-m_0$، 0 برای معادله ۳۳ وجود دارد ولی برای $T > T_c$ تنها یک جواب 0 داریم. اما باید دقت کنیم که برای $T < T_c$ جواب 0 ناپایدار است به این معنا که با کمی انحراف از 0، جواب معادله $m = \beta w \tanh m$ به یکی از دو مقدار m_0 یا $-m_0$ سوق پیدا خواهد کرد. این



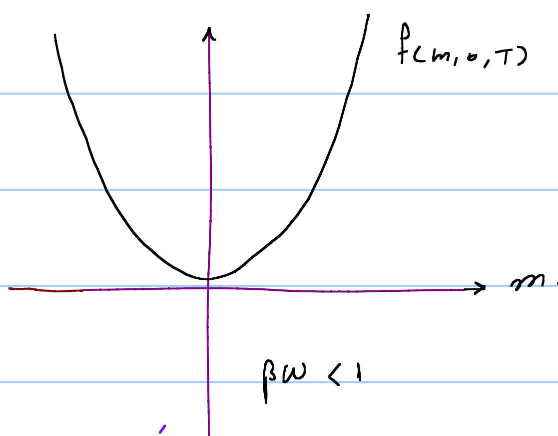
این موضوع را با دقت به شکل روبرو می‌توانیم بفهمیم.

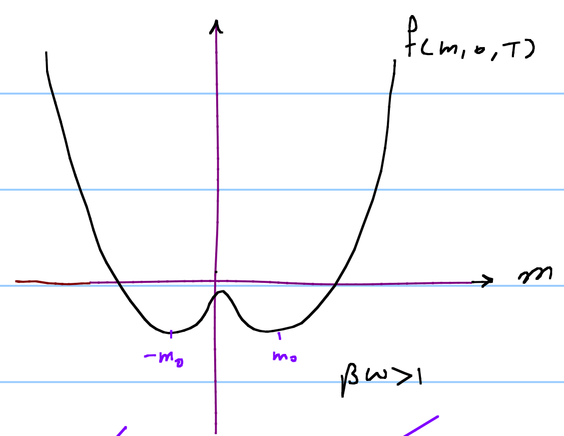
برای فهم بهتر این موضوع می‌توانیم به تابع انرژی آزاد هلمهولتز نیز رجوع کنیم. می‌دانیم که این تابع برابر است با: $\tilde{f}(h,T) = f(m_0,h,T)$

در این جا می‌توانیم به شرایطی که میدان مغناطیسی حضور ندارد. یعنی شرایط $h=0$. بنابراین انرژی آزاد را تابعی از دما خواهیم بود. یعنی $\tilde{f}(T) = f(m_0, 0, T)$. هم چنین می‌دانیم که یک سیستم بسته در تعادل گرمایی در حالتی قرار خواهد گرفت که انرژی آزاد آن به کمترین مقدار خود برسد. بنابراین تابع $f(m,0,T)$ را در نظر بگیریم. این تابع برابر است با:

$$f(m,0,T) = \frac{m^2}{2\beta^2 w} - \frac{1}{\beta}\ln[2\cosh m]. \qquad ۳٤$$

خواهیم دید بر اساس تغییرات دما، شکل این تابع بسته به مقدار βw یکی از دو شکل زیر است:





بنابراین معلوم می‌شود که در $T > T_c$، کمترین مقدار انرژی آزاد مربوط به $m=0$ و برای $T < T_0$ کمترین مقدار آن مربوط به $+m_0$ و $-m_0$ است. به دلیل تقارن نیز هر کدام از مقادیر m_0 یا $-m_0$ می‌توانند انتخاب شوند و مغناطش خودبخودی سیستم می‌تواند هر کدام از این دو مقدار باشد.

۳- مدل Hopfield در حد $\frac{p}{N}\to 0$

مطالعه‌ی مدل از زاویه‌ی دیگر. در بخش گذشته ما سعی بر مطالعه‌ی مدل Hopfield در این بخش آغاز کرده است. در ادامه این مدل را دقیق‌تر مثل یک شبکه‌ی عصبی با ثابت‌های جفت‌شدگی $w_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\sigma_i^{(\mu)}\sigma_j^{(\mu)}$ در نظر می‌گیریم و بسیاری از نتایج بخش قبلی را تکرار کنیم. هدف ما آن است که می‌توانیم تابع آزاد را برای چنین شبکه‌ای پیدا کنیم. در واقع تابع انرژی برای هر حالت این شبکه از رابطه‌ی $E[S]=-\frac{1}{2}\sum_{i,j}w_{ij}S_iS_j$ تعیین می‌شود و می‌دانیم که در دینامیک تعیینی (deterministic) رفتار این شبکه‌ی عصبی، تابع انرژی نقش مهمی را بازی می‌کند. این امر به این معناست که هر گاه از یک طرح اولیه شروع کنیم، دینامیک شبکه‌ی عصبی ما را به جایی می‌برد که در آن یک مینیمم موضعی تابع E قرار دارد. ولی وقتی که دینامیک شبکه تعیینی نیست بلکه یک دینامیک تصادفی در دمای T است، آنگاه طرح‌های تصادفی طرح‌هایی هستند که مینیمم تابع انرژی آزاد $F=-kT\ln Z$ هستند و نه مینیمم‌های انرژی. در دمای صفر و با توجه به رابطه‌ی ترمودینامیکی $F=E-TS$، انرژی آزاد با انرژی E یکی می‌شود ولی تحت دمای غیر صفر نیز پیدا کردن مینیمم‌های تابع F با استفاده از روش‌های مکانیک آماری بسیار آسان‌تر از پیدا کردن مینیمم‌های E است. بعد از این مقدمات به سراغ تابع پارش می‌رویم.

$$Z=\sum_{S_1\cdots S_N}e^{-\beta E[S_1\cdots S_N]}=\sum_{S_1\cdots S_N}e^{\frac{\beta}{2}\sum_{i,j}w_{ij}S_iS_j}$$

$$=\sum_{S_1S_2\cdots S_N}e^{\frac{\beta}{2N}\sum_{i,j,\mu}\sigma_i^{(\mu)}\sigma_j^{(\mu)}S_iS_j} \qquad ۳۵$$

حال ضرب داخلی زیر را تعریف می‌کنیم:

$$\boxed{\langle\sigma^{(\mu)},S\rangle=\sum_i\sigma_i^{(\mu)}S_i} \qquad ۳۶$$

این عبارت در واقع میزان هم‌پوشانی طرح موجود در شبکه یعنی $S=(S_1,S_2,\dots S_N)$ را با طرح ذخیره شده $\sigma^{(\mu)}=(\sigma_1^{(\mu)},\dots,\sigma_N^{(\mu)})$ به دست می‌دهد. مقدار این هم‌پوشانی می‌تواند از 0 (عدم هم‌پوشانی) تا N (هم‌پوشانی کامل) تغییر کند.

در نتیجه تابع پارش به شکل زیر می‌آید:

$$Z(\beta) = \sum_{s_1 \dots s_N} e^{\frac{\beta}{2N}\sum_{\mu=1}^{p} \langle \sigma^{\mu}, s\rangle^2} \qquad ۳۷$$

اگر بخواهیم تابع پارش را در حضور میدان مغناطیسی حساب کنیم، باید به کمیت زیر توجه کنیم که در آن $\vec{h} = (h_1, h_2, \dots \dots h_N)$ بردار
میدان مغناطیسی خارجی است:

$$Z(\beta, \vec{h}) = \sum_{s_1, s_2 \dots s_N} e^{\frac{\beta}{2N}\sum_{\mu=1}^{p} \langle \sigma^{\mu}, s\rangle^2 + \beta \langle \vec{h}, s\rangle} \qquad ۳۸$$

باز هم از رابطه انتگرال گاوسی استفاده می‌کنیم.

$$e^{\frac{\beta}{2N}\langle \sigma^{\mu}, s\rangle^2} = Const \int dm_\mu \, e^{-\frac{N}{2\beta} m_\mu^2 + m_\mu \langle \sigma^{\mu}, s\rangle} \qquad ۳۹$$

که در آن m_μ متغیر تکی انتگرال‌گیری است. ولی در این جا به جای یک متغیر تکی، تعداد p تا متغیر تکی داریم. تابع پارش
برابر می‌شود:

$$Z(\beta, \vec{h}) = Const \int d\vec{m} \sum_{s_1 \dots s_N} e^{-\frac{N}{2\beta}\sum_{\mu=1}^{p} m_\mu^2 + \sum_{\mu=1}^{p} m_\mu \langle \sigma^{\mu}, s\rangle + \beta \langle h, s\rangle} \qquad ۴۰$$

حال در این تابع پارش فوق جمع روی s ها را انجام می‌دهیم: داریم:

$$\sum_{s_1 \dots s_N} e^{\sum_{i=1}^{N}\sum_{\mu=1}^{p} m_\mu \sigma_i^\mu s_i + \beta \sum_{i=1}^{N} h_i s_i} = \prod_{i=1}^{N} \sum_{s_i = 1, -1} e^{[(\sum_{\mu=1}^{p} m_\mu \sigma_i^\mu) + \beta h_i] s_i}$$

$$= \prod_{i=1}^{N} 2\cosh\left(\beta h_i + \sum_{\mu=1}^{p} m_\mu \sigma_i^\mu\right) \qquad ۴۱$$

حال فرض کنید که یک نوع دیگر ضرب داخلی به شکل زیر تعریف کنیم :

$$m\cdot\sigma_i := \sum_{\mu=1}^{p} m_\mu \sigma_i^{(\mu)}$$

در این نوع ضرب داخلی ، هر یک اسپین متغیری است و در این ضرب از $\mu=1$ تا $\mu=p$ جمع زده‌ایم. به همین ترتیب داریم

$$m\cdot m = \sum_{\mu=1}^{p} m_\mu^2 \quad .$$

بنابراین تابع پارش به شکل زیر در می‌آید :

$$Z(\beta,\vec{h}) = \text{Const}\int d\vec{m}\; e^{-\frac{N}{2\beta}\, m\cdot m \;+\; \sum_{i=1}^{N} \ln[2\cosh(\beta h_i + m\cdot\sigma_i)]}$$

$$= \text{Const}\int d\vec{m}\; e^{-N\beta f(\vec{m})} \qquad ٤٢$$

که در آن $f(\vec{m})$ عبارت زیر است :

$$\boxed{f(\vec{m}) = \frac{1}{2\beta^2}\, m\cdot m - \frac{1}{N\beta}\sum_{i=1}^{N} \ln[2\cosh(\beta h_i + m\cdot\sigma_i)]} \qquad ٤٣$$

باید دقت کنیم که f در واقع تابع $\vec{m}$ ، $\vec{h}$ ، و β است، و تنها برای اختصار آن‌ها را صراحتاً در $f(\vec{m})$ ننوشته‌ایم.
اگر این رابطه را با عبارت $f(m)$ برای مدل ساده آیزینگ ببینیم متوجه شباهت و تفاوت آن می‌شویم : در مورد مدل آیزینگ داشتیم ،

$$\boxed{f(m) = \frac{1}{2\beta^2}\, m^2 - \frac{1}{\beta}\ln[2\cosh(\beta h + m)]\, .} \qquad ٤٤$$

به غیر از اینکه کمیت عددی m به کمیت برداری $\vec{m}$ تبدیل شده است ، تفاوت مهم این دو عبارت آن است که عبارت ٤٤ ، هنوز به N

تنها دلیلی که وابستگی به Σ ، نیز N است که در حد بزرگ به بی‌نهایت میل می‌کند. بنابراین عبارت دقیق تابع ۴۲ معادل نیست مگر آن که بتوانیم تقریب مناسبی برای آن قائل شویم. تقریبی که قائل می‌شویم آن است که می‌نویسیم:

$$\frac{1}{N}\sum_{i=1}^{N}\sigma_i^{\mu} \Rightarrow \langle\langle \sigma^{\mu}\rangle\rangle .$$

معنای این حرف آن است که متغیرهای تصادفی σ_i^μ را که روی i (به ازای هر μ) جمع زده‌ایم، با مقدار $\langle\langle \sigma^\mu\rangle\rangle$ عوض کرده‌ایم و فرض کرده‌ایم که درباره‌ی آن چیزی می‌دانیم. مثلاً فرض کرده‌ایم $\langle\langle \sigma^\mu\rangle\rangle=0$. هم‌چنین می‌نویسیم:

$$\frac{1}{N}\sum_{i=1}^{N}\sigma_i^{\mu}\sigma_i^{\nu} \Rightarrow \langle\langle \sigma^{\mu}\sigma^{\nu}\rangle\rangle = \delta^{\mu\nu} . \qquad ۴۵$$

در واقع به متغیرهای σ^μ به عنوان متغیرهای مستقل با توزیع احتمال یکنواخت نگاه می‌کنیم.

$$\boxed{f(\vec m) = \frac{1}{2\beta^2}\, m\cdot m - \frac{1}{\beta}\langle\langle \ln[2\cosh(\beta h + m\cdot\sigma)]\rangle\rangle} \qquad ۴۶$$

بنابراین

$$Z(\beta,h) = \text{Const}\; e^{-\beta N f(\vec m_0)} \qquad ۴۷$$

که در آن $\vec m_0$ جایی است که f مقدار مینیمم خود را اختیار می‌کند. نکته‌ی جالب آن است که به صورت مدل تصادفی

مثل آنتروپی تابع $f(\vec m)$ تعداد بسیار زیادی مینیمم دارد و پیدا کردن آن‌ها نیز ساده نیست.

حالا مقدار $f(\vec m)$ را در غیاب $\vec h$ محاسبه می‌کنیم. پس از این به بعد قرار می‌دهیم $\vec h=0$. داریم

$$\boxed{f(\vec m) = \frac{1}{2\beta^2}\, m\cdot m - \frac{1}{\beta}\langle\langle \ln[2\cosh(m\cdot\sigma)]\rangle\rangle} \qquad ۴۸$$

مقدار $\vec m$ با این شرط به دست می‌آید که مشتقات $f(\vec m)$ در آن صفر باشند یعنی $\frac{\partial f}{\partial m^\mu}=0\ \ \forall\mu$.

$$m_0^{\mu} = \beta \langle\langle \sigma^{\mu} \tanh (\vec{m}_0 \cdot \vec{\sigma}) \rangle\rangle \qquad ۵۰$$

این رابطه جایگزین رابطه ساده $m_0 = \beta w \tanh \beta m_0$ شده است. در حالی که حل این معادله خیلی آسان به روش ترسیمی به دست آمد، به دست آوردن حل‌های معادله ۵۰ بسیار مشکل‌تر است.

دقت کنید که معادله ۵۰، یک دسته معادلات غیرخطی جفت شده بهم برای یک بردار p مؤلفه‌ای $\vec{m}^0=(m_1^0, m_2^0, \ldots m_j^0 \ldots m_p^0)$ است. با فرض اینکه در مورد متوسط‌گیری‌ها، می‌توانیم جا فاصله‌ها را به طور خلاصه بر حسب m_μ بنویسیم (یعنی توزیع‌های σ^μ و $\sigma^\mu \sigma^\nu$ و ... را حساب کنیم) ولی باز هم حل کردن این معادلات بسیار مشکل خواهد بود و ناچاریم که باز هم فرض‌های ساده‌کننده‌ای به کار ببریم تا موفق به حل این معادلات شویم. اما قبل از آنکه شروع به حل این معادلات کنیم، بهتر است معنای فیزیکی بردار $\vec{m}^0=(m_1^0, m_2^0, \ldots m_j^0 \ldots m_p^0)$ را که تحول اصلی ماست ببینیم.

در درس سوم که فصل مربوط به شبکه‌های عصبی بود متوجه شدیم که وقتی یک طرح را به شبکه عرضه می‌کنیم، آنگاه با الگوریتم دینامیک دوران شبکه، طرح ξ^μ، به وضوح به یاد آورده می‌شود که مقدار متوسط $\langle S_i \rangle$ برای هر نورون برابر است با:

$$\langle S_i \rangle = m_0 \sigma_i$$

به عبارت دیگر m درجه وضوح طرح به یاد آورده شده است. در آن درس برای هر طرح یک ضریب بیان وضوح قائل شدیم (تنها برای اینکه فرض ساده‌کننده‌ای برای حل معادلات به کار برده باشیم). بردار $\vec{m}^0$ در این جا معنای مشابهی دارد به این معنا که این بار $\vec{m}$ دارای p تا مؤلفه است، $\vec{m}_0=(m_0^1, m_0^2, \ldots m_0^p)$ و هر مؤلفه آن در واقع نشان دهنده این است که طرح مربوطه با چه وضوحی به یاد آورده می‌شود.

برای اینکه این موضوع را به طور دقیق‌تر ببینیم، متوسط $\langle S_i \rangle$ را حساب می‌کنیم.

از معادله برای متوسط اسپین i ام پیدا کنیم داریم:

$$\langle S_i \rangle = \frac{1}{Z} \frac{\partial}{\partial \beta h_i} Z \Big|_{h=0} = \frac{\partial}{\partial \beta h_i} \ln Z \Big|_{h=0} \qquad ۵۱$$

منظور از $h=0$ آن است که همه h_i ها را پس از مشتق‌گیری برابر با صفر قرار دهیم. اما ما به ترتیب متوسط اسپین‌ها نورون مخصوصی علاقه نداریم. بلکه می‌خواهیم ببینیم که این طرح چقدر شبیه به یک طرح ذخیره شده است. یعنی علاقه‌مند به کمیت زیر هستیم:

$$q^2 := \frac{1}{N} \sum_{i=1}^{N} \sigma_i^{(2)} \langle S_i \rangle . \qquad ۵۲$$

این کمیت مقدار همپوشانی طرح $\langle S_i \rangle$ ها با طرح σ^2 را به دست می‌دهد و مقدار آن بین ۱- و ۱ تغییر می‌کند.

$$q^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \frac{\partial}{\partial \beta h_i} \ln Z \qquad ۵۳$$

با توجه به این که $Z = C e^{-N\beta f(\vec{m}_0, \vec{h})}$ و رابطه ۴۳ به دست می‌آید:

$$q^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \langle S_i \rangle = - \sum_{i=1}^{N} \sigma_i^2 \frac{\partial}{\partial h_i} f(\vec{m}_0, \vec{h}) . \qquad ۵۴$$

دقت کنید که $\vec{m}_0$ از طریق معادله خود تابع h است. یعنی باید بنویسیم:

$$q^2 = - \sum_{i=1}^{N} \sigma_i^2 \left\{ \frac{\partial f(\vec{m}_0(h), h)}{\partial \vec{m}_0} \frac{\partial \vec{m}_0}{\partial h_i} + \frac{\partial f(m_0(h), h)}{\partial h_i} \right\} \qquad ۵۵$$

جمله اول بنا بر تعریف $\vec{m}_0$ مساوی صفر است و نتیجه:

$$q^2 = - \sum_{i=1}^{N} \sigma_i^2 \frac{\partial f(m_0(h), h)}{\partial h_i} \qquad ۵۶$$

با توجه به رابطه ۴۳ یعنی

$$\boxed{f(\vec{m},h) = \frac{1}{2\beta^2}\, m\cdot m - \frac{1}{N\beta}\sum_{i=1}^{N} \ln\left[2\cosh(\beta h_i + m\cdot\sigma_i)\right]}$$ ٤٣

بدست می‌آوریم:

$$\rightarrow\quad q^2 = -\sum_{i=1}^{N} \sigma_i^2 \left(\frac{1}{N\beta}\right)\beta \tanh(\beta h_i + m_0\cdot\sigma_i)\Big|_{h=0}$$

بدست می‌آوریم:

$$\rightarrow\quad q^2 = -\sum_{i=1}^{N} \sigma_i^2 \left(\frac{1}{N\beta}\right)\beta \tanh(\beta h_i + m_0\cdot\sigma_i)\Big|_{h=0}$$

$$\rightarrow\quad q^2 = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 \tanh(m_0\cdot\sigma_i).$$

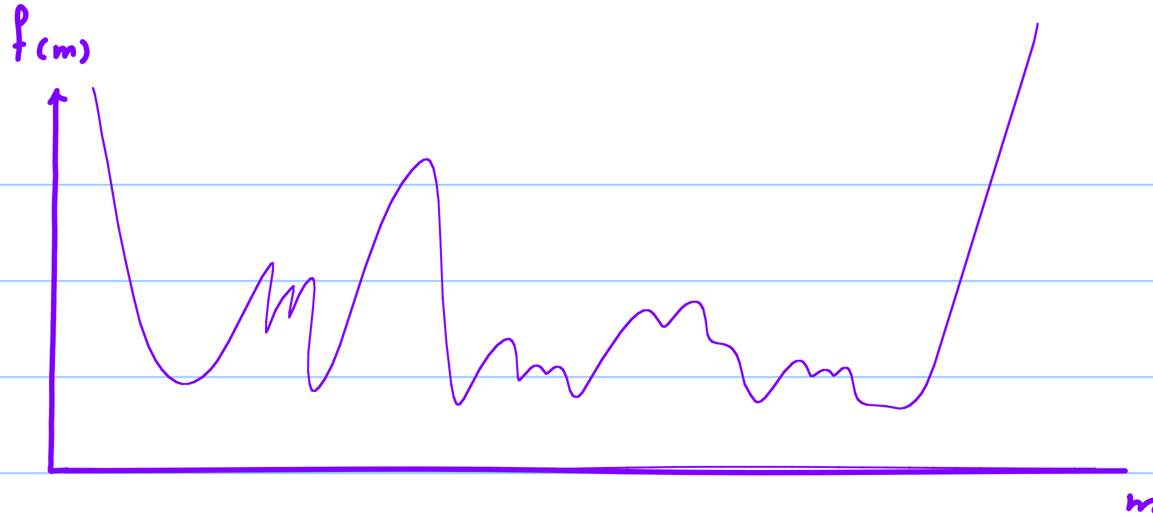
$$\rightarrow\quad q^2 = \langle\langle \sigma^2 \tanh(m_0\cdot\sigma)\rangle\rangle \overset{(50)}{=} \frac{1}{\beta} m_0^2.$$ ٥٧

بنابراین m^2 متناسب است با همپوشانی طرح $\langle S_i \rangle$ با طرح حفظ شده σ^2.

حال می‌پردازیم به حل معادلات

$$\boxed{m_0^2 = \beta \langle\langle \sigma^{(2)} \tanh(m_0\cdot\sigma)\rangle\rangle.}$$ ٥٨

همانطور که گفتیم، این معادلات، مجموعه‌ای p تایی از معادلات غیرخطی کوپل شده برای متغیرهای m^2 هستند. این معادلات می‌توانند در تابع انرژی آزاد $f(\vec{m}(T),T)$ ها نشان داده شوند. دقت کنید که $\vec{h}$ را برابر با صفر قرار داده‌ایم.
تعداد این جواب‌ها بسیار زیاد است (شکل زیر).



ساختار تابع f در رژیم ۲ در آن

قبل از آنکه به بررسی جواب‌های پیدا کردن این مینیمم‌ها بپردازیم، به یک خاصیت مهم آنها بدون اثبات اشاره می‌کنیم. نشان داده شده است که در رژیم ۲ این تابع دارای خاصیتی موسوم به ultrametricity هست که به ساختار درختی و شاخه شاخهٔ تابع f اشاره دارد. به عبارت دیگر اگر یکی از دره‌های تابع f را بزرگ کنیم، درون آن تعداد زیادی درهٔ دیگر پیدا می‌کنیم و این خاصیت در هر بزرگنمایی حفظ و تا آخر تکرار می‌شود:

حال می‌پردازیم به حل معادلات

$$\boxed{m^\mu = \beta \langle\langle \sigma^\mu \tanh(m\cdot\sigma)\rangle\rangle.} \qquad ۵۸$$

همانطور که گفتیم، این معادلات، مجموعه‌ای p تایی از معادلات غیرخطی کوپل شده برای متغیرهای m^μ هستند. واضح است که هیچ راه آسانی برای حل دقیق و تحلیلی این معادلات بدون فرض‌های ساده‌کننده وجود ندارد. در نخستین تقریب بیایید فرض کنیم که همه طرح‌ها با وضوح یکسان به خاطر آورده می‌شوند. یعنی $\vec{m} = (m, m, m, \ldots m).$

در این صورت معادلهٔ ۵۸ تبدیل می‌شود به:

$$m = \beta \langle\langle \sigma^\mu \tanh m\left(\sum_{\nu=1}^{p} \sigma^\nu\right)\rangle\rangle \qquad ۵۹$$

با فرض $\langle\langle \sigma^\mu \sigma^\nu \rangle\rangle = \delta^{\mu\nu}$ $\leftarrow$

$$m = \beta \tanh m.$$

که قبلاً دیده‌ایم اگر $\beta < 1$ باشد، جواب $m=0$ دارد. اگر $\beta > 1$ باشد، دو جواب $\pm m$ خواهد داشت.

حال فرض کنید که k تا طرح با وضوح یکسان به خاطر آورده می‌شوند و $p-k$ تا طرح اصلاً به یاد آورده نمی‌شوند. بدون نقص کلیت، بردار $\vec{m}$ را به صورت زیر در نظر گرفت:

$$\vec{m} = (\underbrace{m, m, m, \dots m}_{k}, \underbrace{0, 0, 0, \dots 0}_{p-k}) \qquad ۶۰$$

در این صورت داریم:

$$\mu = 1 \rightarrow \quad m = \beta \langle\langle \sigma^1 \tanh m(\sigma^1 + \sigma^2 + \dots \sigma^k) \rangle\rangle$$

$$\mu = 2 \rightarrow \quad m = \beta \langle\langle \sigma^2 \tanh m(\sigma^1 + \sigma^2 + \dots \sigma^k) \rangle\rangle \qquad ۶۱$$

$$\vdots$$

$$\mu = k \rightarrow \quad m = \beta \langle\langle \sigma^k \tanh m(\sigma^1 + \sigma^2 + \dots \sigma^k) \rangle\rangle$$

$$\rightarrow \boxed{k m = \beta \langle\langle (\sigma^1 + \sigma^2 + \dots \sigma^k) \tanh m(\sigma^1 + \sigma^2 + \dots \sigma^k) \rangle\rangle} \qquad ۶۲$$

مثال ۱) $k = 2 \leftarrow$

$$2m = \beta \langle\langle (\sigma^1 + \sigma^2) \tanh m(\sigma^1 + \sigma^2) \rangle\rangle \qquad ۶۳$$

با توجه به این که فرض کرده‌ایم که متغیرهای $\sigma^1, \sigma^2, \dots$ متغیرهای مستقل با توزیع یکنواخت هستند که مقادیر ۱ و ۱- را اختیار می‌کنند، معادلات بالا تبدیل می‌شود به:

$$2m = \beta \frac{1}{4} \{ 2 \tanh 2m + 0 + 0 + (-2) \tanh(-2m) \}$$

یا

$$2m = \beta \tanh 2m.$$

که حل آن در مسئله قبلی برای m بحث کردیم (نسبت به مقدار β). بنابراین نتیجه می‌گیریم $\binom{n}{2}$ تا جواب برای معادلات ۵۸ پیدا کرده‌ایم که در آن $\vec{m}$ دو تا مولفه غیرصفر دارد. عدد $\binom{n}{2}$ از این جا آمده که مولفه‌های غیرصفر را می‌توانیم

در سیستم‌های مختلف باشند.

مثال ۲) $k=3.$ ← از معادله ۴۳ بدست می‌آوریم:

$$3m = \beta \langle\langle (\sigma^1+\sigma^2+\sigma^3)\tanh m(\sigma^1+\sigma^2+\sigma^3)\rangle\rangle$$

$$\to 3m = \frac{\beta}{8}\left\{ 6\tanh 3m + 6\tanh m\right\}$$

$$\to \boxed{m = \frac{\beta}{4}\left(\tanh m + \tanh 3m\right)}$$

با حل کردن این معادله (بصورت ترسیمی) می‌توان مقدار m را بدست آورد. به این ترتیب تعداد $\binom{n}{3}$ جواب از معادله ۵۸ بدست می‌آوریم که در آنها $\vec{m}$ دارای ۳ مولفه غیرصفر است. یعنی در این حالت‌ها، سه تا از طرح‌ها با ضریب m به یاد آورده می‌شوند و p-3 تا از طرح‌ها به یاد آورده نمی‌شوند. به همین ترتیب می‌توان حالاتی با تعداد بیشتری از جواب‌ها را بدست آورد. ولی واضح است که این جواب‌ها که از معادله ۴۳ بدست می‌آیند همه جواب‌های معادله ۵۸ هستند.

در اینجا باید تاکید کنیم که فرض ما درباره اینکه $\langle\sigma^\mu\rangle = 0$ و $\langle\sigma^\mu\sigma^\nu\rangle = \delta^{\mu\nu}$ یک فرض بسیار ساده‌کننده است. در این فرض هیچ نوع اطلاعاتی در مورد همبستگی طرح‌ها با یکدیگر وجود ندارد. در واقع فرض کرده‌ایم که هیچ نوع همبستگی بین طرح‌ها وجود ندارد. در یک مطالعه دقیق‌تر باید همبستگی طرح‌ها را هم در نظر گرفت. این موضوعی است که ما به دلیل کمبود وقت نتوانستیم به آن بپردازیم. شاید در فرصت دیگری این کمبود را کامل کنیم.

در اینجا درسنامه درس «فیزیک سیستم‌های پیچیده» به شکل خام و ابتدایی‌اش به پایان می‌رسد.

۲۷ دی ماه ۱۳۸۷ پایان.