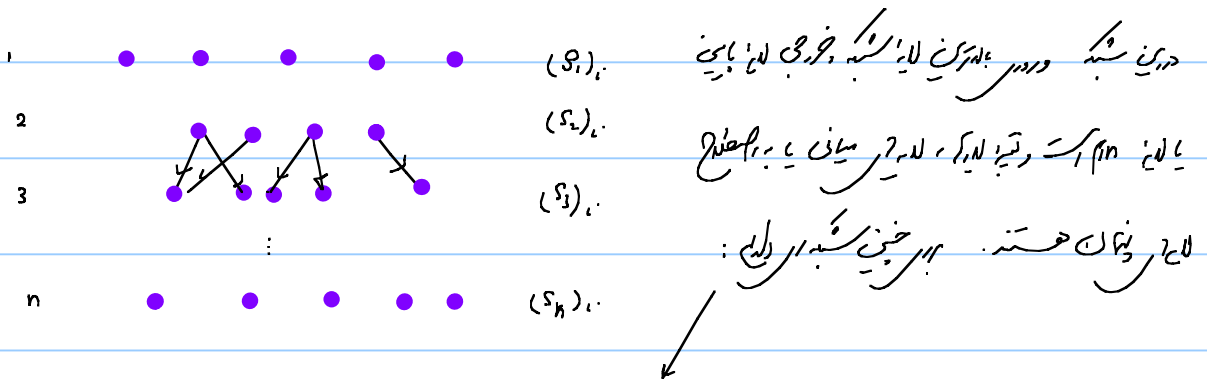


آنچه در تلاطم کتبی به یاد آورده، در یادآوری طرح؟ (به مناسبتی آن) تاکیدی از یادگیری مغز است. به ما هم در مغز در کما، عصبی کنترل نمی تواند که خودش نشان دادن به حرکت است و به احتمال زیاد در طی فرآیند تکامل به روشی غنیتی به مغز با هم چینی هم در دست آورده است. از این جهت که گویا شبکه عصبی از یک طرف آن کنترل در این رخ داده، حرکت در هر یونی است ساختاری اما استناد به شبکه عصبی مربوط به حافظه دارد. چنین شبکه های عصبی در جهت دیگر داده به آن به هر دریافت در ردی کدنگ حرف و نشانه های خارجی در کما نیز مناسب است. ساد مغز این شبکه ها، شبکه های عصبی به

Feed Forward layered Network هستند به شکل زیر هستند:



$$(s_i)_i = f(h_i)$$

که در آن h_i مجموع سینه های در هر یونی در لایه k ام است یعنی:

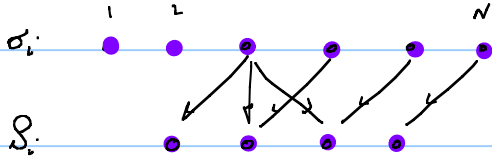
$$h_i = \sum_j w_{ij} (s_{i-1})_j$$

برای تابع f معمولی که از دو شکل زیر در نظر گرفته می شود که شکل اول به شبکه های عصبی در شکل دوم به شبکه های عصبی است:

$$f(h_i) = \begin{cases} \text{sgn}(h_i - \theta_i) & s_i = 1 \text{ یا } -1 & \text{برای شبکه های معنی} \\ \tanh \mu(h_i - \theta_i) & -1 < s_i < 1 & \text{برای شبکه های تقاضی} \end{cases}$$

دقت کنید، برای شبکه های تقاضی متغیر μ بسته به ظاهر گرفته شد است که متغیر آن می باشد.

ساده ترین شبکه عصبی تک لایه آن perceptron نیز گفته می شود. این شبکه از یک لایه عصبی پنهانی تشکیل شده است. در هر دور یک بار شبکه را با داده های خروجی آن با S_i نشان می دهیم. (شکل زیر)



لایه پنهان شبکه در حالت این مدل اشغال نمی شود.

محور S_i ، S_j و S_k به ترتیب این شبکه این مدل است. S_i و S_j و S_k

$$S_i^{(3)} = f(h_i) = f\left(\sum_j \omega_{ij} S_j^{(2)}\right) \quad \text{Actual output}$$

درین $S_i^{(1)}$ Desired output.

هدف از این شبکه یاد دادن به این شبکه است که «دهن» آن یاد بگیرد. این کار را می توانیم با تغییر وزن ها انجام دهیم:

Learning rule:
$$\delta \omega_{ij} = \epsilon (S_i^{(3)} - S_i^{(1)}) S_j^{(2)}$$

این تغییر در شبکه ها در نظر منحصراً به منای یادگیری است. در این مورد S_i است. دقت یادگیری را می توانیم به این طریق با یک مقدار مثبت و منفی در فرآیند یادگیری کم کنیم. مقدار زیر را به این شبکه ها اختصاص می دهیم:

$$\omega_{ij} = \frac{1}{N} \sum_j S_j^{(2)} S_i^{(1)}$$

درین صورت، هر دوری که S_i فراموش است:

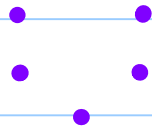
$$S_i = \text{Sgn}\left(\frac{1}{N} \sum_j S_j^{(2)} S_i^{(1)}\right)$$

$$= \text{Sgn}\left(\frac{1}{N} S_i^{(1)} \sum_j S_j^{(2)}\right) = S_i^{(1)} (S_i^{(1)}) = S_i^{(1)}$$

مکان است که خواننده سوال کند این درین صورت شبکه عصبی چگونه از مغرب با حرکت کردن توانی در هر دو جهت است. احتمالی رخ نمی دهد. اما باید توجه داشت که تعیین هر S_i به شکل بالا عددی نیست، زیرا اصولاً می توانیم که در هر موردی که S_i (محرک به S_i) در (Training Set) به شبکه عصبی آموزش دهیم.

که با هر یک مجرباً از بهترین حالت که در این لحظه بدست می آید، به عنوان بهترین معلوم نیست، زیرا آن هنوز در حال
 آتیمی است. محاسبه این به روشی است که در این بخش به آن خواهیم پرداخت. در واقع در این روش، ما با استفاده از XOR
 این دو نشان دهنده یک شباهت است. در اینجا هر چه XOR بیشتر باشد، شباهت کمتر است. در اینجا ما داریم
 آزمون: نشان دهنده یک شباهت است. در اینجا هر چه XOR بیشتر باشد، شباهت کمتر است. در اینجا ما داریم

تایخ XOR با نایخ دهنده.



باید متوجه شویم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم
 این محاسبه را می توانیم به روشی دیگر نیز انجام دهیم. در اینجا ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم
 یادگیری را می توانیم به روشی دیگر نیز انجام دهیم. در اینجا ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم
 در فرآیند یادگیری، ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم
 که در اینجا ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم

Learning by gradient method.

هدف آموزش آن است که ما داشته باشیم یک روشی که در اینجا ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم
 ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم که هر چه شباهت بیشتر باشد، شباهت بیشتر است. در اینجا ما داریم

$$D := \frac{1}{2} \sum_{i,j} (s_i^t - s_j^t)^2 = \sum_{i,j} (1 - s_i^t s_j^t)$$

$$= \sum_{i,j} (1 - s_i^t f(h_{ij}^t))$$

حال می خواهیم این حاصل به مشتق نسبت به هر یک از سیناپس ها، $\frac{\partial D}{\partial w_{jk}}$ را حساب کنیم.

$$\frac{\partial D}{\partial w_{jk}} = \sum_{i,j} -s_i^t f'(h_{ij}^t) \frac{\partial h_{ij}^t}{\partial w_{jk}}$$

اداره:

$$h_i^t = \sum_j \omega_{ij} \sigma_j^{t-1}$$

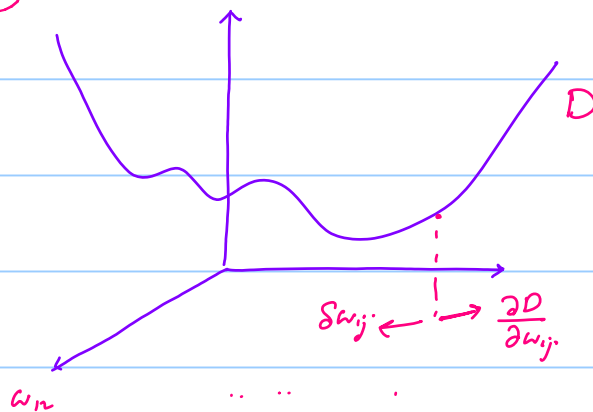
در تکرار یک است:

$$\frac{\partial h_i^t}{\partial \omega_{jk}} = \delta_{ij} \sigma_k^t$$

$$\begin{aligned} \rightarrow \frac{\partial D}{\partial \omega_{jk}} &= - \sum_{i,j} s_i^t f'(h_i^t) \delta_{ij} \sigma_k^t \\ &= - \sum_j s_j^t f'(h_j^t) \sigma_k^t = - \sum_j \Delta_j^t \sigma_k^t \end{aligned}$$

حال فرض کنیم که ما در یک نقطه از سطح D ایستاده باشیم و می‌خواهیم بدانیم که در چه جهت باید حرکت کنیم تا D کم شود. بنابراین

$$\rightarrow \delta \omega_{jk} = \epsilon \sum_{\mu} \Delta_j^{\mu} \sigma_k^{\mu} \quad \leftarrow \text{فرمول جدید}$$



حالا اگر ما در یک نقطه از سطح D ایستاده باشیم و می‌خواهیم بدانیم که در چه جهت باید حرکت کنیم تا D کم شود. یعنی:

$$\begin{aligned} \frac{\partial D}{\partial \theta_j} &= - \sum_{i,j} s_i^t f'(h_i^t) \frac{\partial h_i^t}{\partial \theta_j} \\ &= \sum_{i,j} s_i^t f'(h_i^t) \delta_{ij} = \sum_j s_j^t f'(h_j^t) \end{aligned}$$

$$\delta \theta_j = - \epsilon \sum_j s_j^t f'(h_j^t)$$

که در آن نتیجه می‌گیریم:

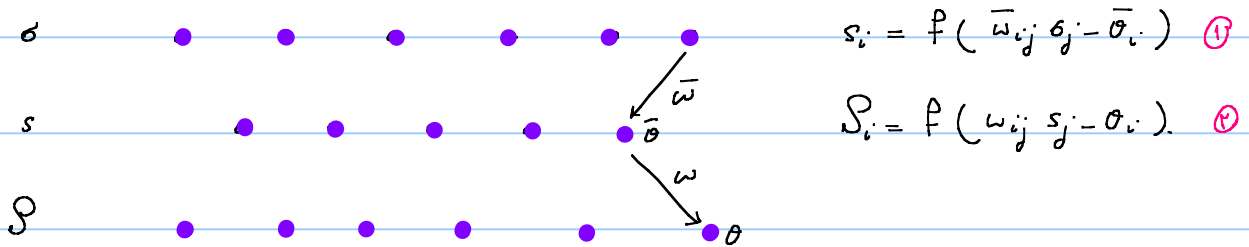
برای ترتیب کردن سینی هموزون و آستانه‌ها، ما زمانی که در این سینی ایستاده باشیم و می‌خواهیم بدانیم که در چه جهت باید حرکت کنیم تا D کم شود. در این ترتیب سینی هموزون، perceptron بکار می‌رود. بنابراین، جهت‌های سینی هموزون، در این ترتیب سینی هموزون

آخرین لایه تصحیح می‌کند، لایه‌های میانی هم تصحیح می‌کنند تا اینکه در لایه اول تصحیح شوند. همین دلیل این است که اگر یک تصحیح خطای از ابتدا شروع کنیم:

نامیوس: ←

Learning by error-back propagation

مشکل زیر یک شبکه عصبی با یک لایه ورودی نشان می‌دهد. متغیرهایی که حالت نورون i را به لایه اول نشان می‌دهد و در همدوم چنین سینی i و j در شکل زیر نشان داده شده اند.



$$D = \sum_{i,j} (1 - \sigma_i^{(j)} p_i^{(j)}) \quad \text{دلیل:}$$

رابطه ①، نشان می‌دهد که این فاصله تابعی از w_{ij} ، θ_i ، σ_j ، p_i است. بر این اساس، با آنچه در مورد شبکه در فصل‌های انجام داده استات D نسبت به این متغیرها حساب کرده و پس به هر شبکه با شکل زیر تغییر داد:

$$\delta w_{ij} = -\epsilon \frac{\partial D}{\partial w_{ij}}, \quad \delta \bar{w}_{ij} = -\epsilon \frac{\partial D}{\partial \bar{w}_{ij}}$$

$$\delta \theta_i = -\epsilon \frac{\partial D}{\partial \theta_i}, \quad \delta \bar{\theta}_i = -\epsilon \frac{\partial D}{\partial \bar{\theta}_i}$$

Representation of Boolean

این شبکه یک شبکه با n نورون ورودی و یک نورون خروجی دارد. این شبکه ممکن است لایه‌های ورودی نیز داشته باشد. حال از همه سوال می‌کنیم: آیا ممکن است این شبکه را چنان تغییر دهیم که به اندازه یک تابع می‌تواند

$D, 0 \rightarrow \{0, 1\}^n$ به نایک دصرینی به ایزر حروری $x \in \{0, 1\}^n$ درزون خوری معادله $f(x)$ پوره کوه؟

دیه که ارسبده لایه مینی نمانده بشه پانچ این کوال منی است. مثال نقض کنند: آن تابع XOR است. درین جا

می خوام نشان دهم که ارسبده معنی دلبریک لایه مینی بشه، آنجا هر از آن طریق دکوای را انتخاب سیندین آرسبده

نایک دله. در واقع نشان خوام دهم که مبر نایک تابع $f: \{0, 1\}^n \rightarrow \{0, 1\}$ یک لایه مینی با 2^n درزون

نمانده است. دقت کنید که مکن است این شبکه بهترین شبکه ممکن نباشه. هم چنین دقت کنید که تعداد درزون هر لایه

مینی به جهت نایک با n بیشتر کند درین امر نشان میده که چرا نمی توان یک شبکه معنی به جهت سازنده Constructive

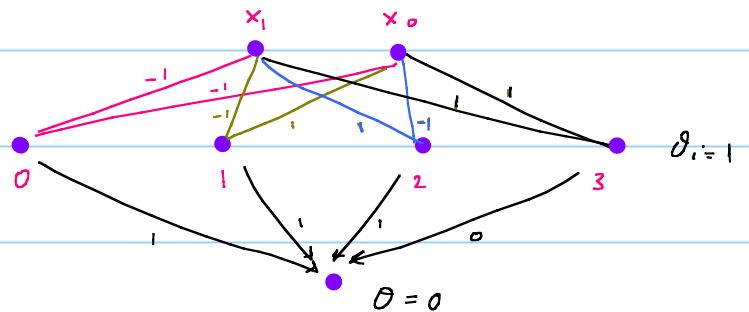
بر نایک حریتی ساخت در جهت مبر نایک دهم بر آرسبده آن شبکه. این یکی اثبات با دکر در حال نشان

دج. خواننده با مطالعه این امثال و خواندن ترنجات بعد از آن که مر واند قضایای مینبر.

x_1	x_0	$f(x)$
0	0	1
0	1	1
1	0	1
1	1	0

مثال (۱) جدول مبریک تابع $f: \{0, 1\}^2 \rightarrow \{0, 1\}$

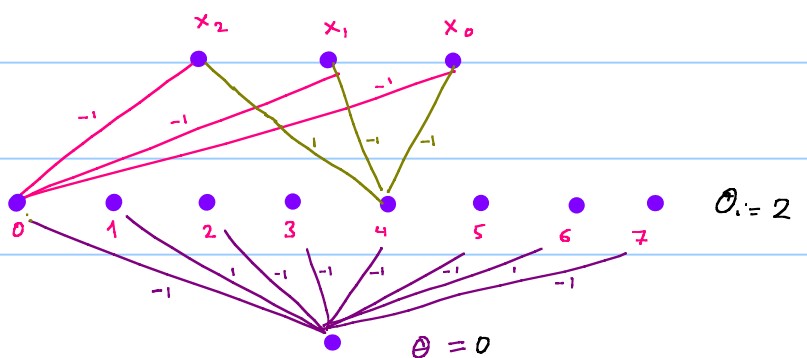
ساختار شبکه دینایک درین تابع با نایک مبرده به جهت زیر است:



x_2	x_1	x_0	$f(x)$
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	1
1	1	1	0

مثال (۲) جدول مبریک تابع نشان مبرده معنی که مبریت این

تابع نشان مده مطابق زیر است:



خواننده بدقت در مثال در یاد می‌گیرد که هر زمان که s و m ضریب منطقی باشد، اثبات کلی منجر به n نموده. لایه سیانی دایره 2^n نورون است که هر کدام از سطوح یک عدد $1 \leq x \leq 2^{n-1}$ هستند. در مثال ۱ نورون در لایه اول را متناظر با عدد ۵، ۲، ۱ و ۳ هستند. سینی در لایه اول هر انتخاب شده اند که هرگاه n باشد در لایه اول قرار گرفت تا نورون برده با آن عدد در لایه سیانی درش شده و تعبیر نورون که خارج می‌باشند

(دست‌نویس از کتاب اینی $s = \pm 1$ بهر وضعیت نورون را دست‌نویس می‌کنیم یعنی $(s=1) \leftrightarrow n=0$ و $(s=-1) \leftrightarrow n=1$).

بنابراین مثال هرگاه در لایه اول قرار می‌گیرد $x_0=1, x_1=1$ ، تا نورون $n=3$ درش خواهد شد، هرگاه قرار می‌گیرد $x_0=-1, x_1=1$ ، تا نورون $n=2$ درش خواهد شد. شکل ۵ بهر تا نورون n در لایه سیانی بهر $n-1$ است (۱ در مثال اول و ۲ در مثال دوم).

حال به سینی در لایه سیانی می‌پردازیم: سینی که نورون x به نورون لایه آخری که وصل می‌باشد w_x نشان میدهیم. در نتیجه خواهیم داشت:

$$S = \text{sgn} \left(\sum_{x=0}^{n-1} w_x x - \theta \right)$$

حال اگر تا نورون x در لایه سیانی درش کند خواهیم داشت:

$$S = \text{sgn} (w_x - \theta)$$

برخوب که در این قسمت داشته باشیم $S = P(x_0)$ یعنی $P(x_0) = \text{sgn}(w_x - \theta)$

$$\begin{cases} w_x > \theta & \text{if } P(x_0) = 1 \\ w_x < \theta & \text{if } P(x_0) = -1 \end{cases}$$

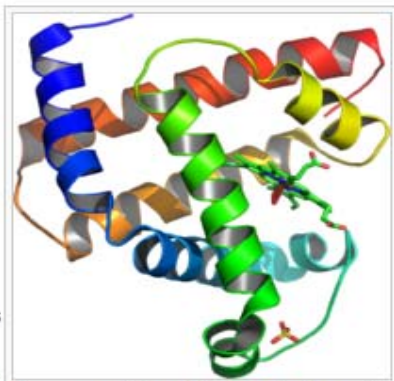
بنابراین کافی است که در خروجی:

Applications

بعد از این مدت است که در کتاب مثال مشخص از بار در یک شیب معینی قرار دارد. باید به یاد داشته باشیم که شیب معینی بار در منحنی درونی یک مشتق است که تانگنسی به آن در نقطه α و β است. این مجموعه از بارها در یک شیب معینی قرار دارد، آنکه α و β که در مجموعه از بارها در یک شیب معینی قرار دارد. این مجموعه از بارها در یک شیب معینی قرار دارد، آنکه α و β که در مجموعه از بارها در یک شیب معینی قرار دارد.

مثال: پیش بینی ساختار جانور کاتین؟

هر چه پیشتر یک مارکد کول رشته ای را به یاد داشتیم که از توانایی مجموعه از α آمینو اسید تشکیل شده است. این در ساختار یک پروتئین چه نوع آمینو اسید ایی را به چه ترتیبی به کار رفته است ساختار اولیه یا *primary structure* پروتئین نامیده می شود. به محض آنکه ساختار اولیه پروتئین تعیین شود، بلافاصله پروتئین به دو نوع دیگر ساختار ثانویه یا *secondary structure* می یابد که شامل α -helix و β -sheet و *random coil* تقسیم می شوند. در هر دو آنرا مارکد کول پروتئین همان دو نوع فشرده شده و شکلی پیدا کند که به نام *random coil* می نامند.

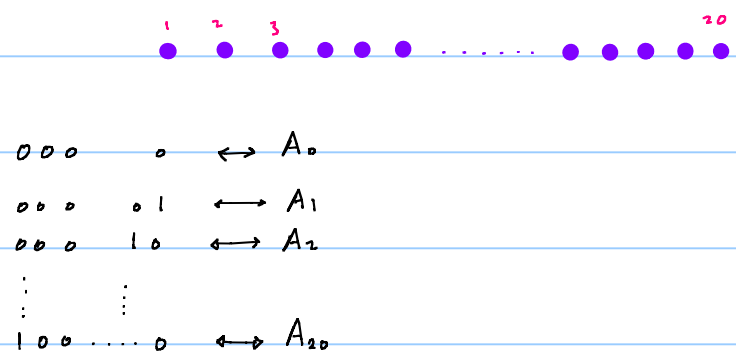


A representation of the 3D structure of the myoglobin protein. Alpha helices are shown in colour, and random coil in white, there are no beta sheets shown. This protein was the first to have its structure resolved by X-ray crystallography by Max Perutz and Sir John Cowdery Kendrew in 1958, which led to them receiving a Nobel Prize in Chemistry in 1962.

این که در سطح لوله نیز در شیب معینی بین آن دو آمینو اسید، حتی برای دلیلی که در یک شیب معینی قرار دارد، آنکه α و β که در مجموعه از بارها در یک شیب معینی قرار دارد. این مجموعه از بارها در یک شیب معینی قرار دارد، آنکه α و β که در مجموعه از بارها در یک شیب معینی قرار دارد.

پرسش **protein folding problem** مشهور است. با ظهور روش‌های شبیه‌سازی محاسباتی، پیشرفت‌های فزاینده در

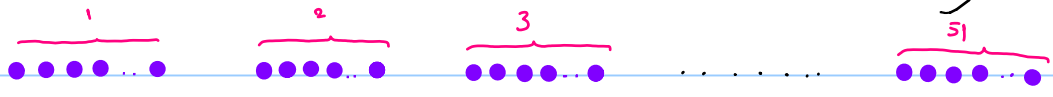
درک این مسئله منطقی است. یک مدل ساده از یک مولکول پلی‌پپتید، راه‌حلی در فضای بسیار بزرگ است. برای یافتن یک ساختار خاص، باید از میان تمام حالت‌های ممکن، آن را شناسایی کرد. این مسئله را می‌توان به عنوان یک مسئله بهینه‌سازی در فضای بسیار بزرگ در نظر گرفت. یکی از روش‌های رایج برای حل این مسئله، استفاده از روش‌های مبتنی بر انرژی است. در این روش، انرژی هر یک از حالت‌های ممکن را محاسبه می‌کنند و آن حالتی که کمترین انرژی را دارد، به عنوان ساختار پایدار در نظر می‌گیرند. این روش‌ها معمولاً برای سیستم‌های کوچک و متوسط استفاده می‌شوند. برای سیستم‌های بزرگ، روش‌های مبتنی بر هم‌بستگی (coupling) استفاده می‌شود. این روش‌ها با استفاده از داده‌های تجربی، به یافتن ساختار پایدار کمک می‌کنند. روش‌های مبتنی بر هم‌بستگی معمولاً برای سیستم‌های بزرگ و پیچیده استفاده می‌شوند.



ترتیب که در شکل زیر است:

این نحوه‌ی نمایش، ساختار پروتئین را نشان می‌دهد (مثلاً به دلیل اهمیت). بهترین روش برای یافتن این ساختار، استفاده از روش‌های مبتنی بر انرژی است.

در ترجمه به آنگاه، می‌توانیم به مطالعه‌ی ساختارهای دیگر، مانند آنزیم‌ها، پرداختیم. برای مطالعه‌ی ساختارهای دیگر، می‌توانیم از روش‌های مبتنی بر انرژی استفاده کنیم. روش‌های مبتنی بر انرژی معمولاً برای سیستم‌های کوچک و متوسط استفاده می‌شوند. برای سیستم‌های بزرگ، روش‌های مبتنی بر هم‌بستگی استفاده می‌شود.



لازم به ذکر است که در روش‌های مبتنی بر انرژی، معمولاً از مدل‌های ساده‌تری استفاده می‌شود. این مدل‌ها معمولاً شامل α -helix و β -sheet می‌باشند. این مدل‌ها معمولاً برای سیستم‌های کوچک و متوسط استفاده می‌شوند. برای سیستم‌های بزرگ، روش‌های مبتنی بر هم‌بستگی استفاده می‌شود.

همچنین، روش‌های مبتنی بر هم‌بستگی معمولاً برای سیستم‌های بزرگ و پیچیده استفاده می‌شوند. این روش‌ها با استفاده از داده‌های تجربی، به یافتن ساختار پایدار کمک می‌کنند. روش‌های مبتنی بر هم‌بستگی معمولاً برای سیستم‌های بزرگ و پیچیده استفاده می‌شوند.