

Bir Coklu Ikisel Dagilim (Multivar. Binary Distribution) ve Boltzmann Dagilimi

$$P(x; W) = \frac{1}{Z(W)} \exp \left[\frac{1}{2} x^T W x \right]$$

Olurluk (likelihood)

$$\prod_{n=1}^N P(x^{(n)}; W) = \frac{1}{Z(W)} \exp \left[\frac{1}{2} x^{(n)T} W x^{(n)} \right]$$

Log olurluk

$$\mathcal{L} = \ln \left(\prod_{n=1}^N P(x^{(n)}; W) \right) = \sum_{n=1}^N \left[\frac{1}{2} x^{(n)T} W x^{(n)} - \ln Z(W) \right] \quad (1)$$

Birazdan $\frac{\partial \mathcal{L}}{\partial w_{ij}}$ turevini alacagiz, o sirada $\ln Z(W)$ 'nin turevi lazim, daha dogrusu $Z(W)$ 'yi nasil turevi alinir hale getiririz?

$Z(W)$ normalizasyon sabiti olduguna gore, dagilimin geri kalaninin sonsuzlar uzerinden entegrali (ya da toplami) normalizasyon sabitine esittir,

$$Z(W) = \sum_x \exp \left[\frac{1}{2} x^T W x \right]$$

$$\ln Z(W) = \ln \left[\sum_x \exp \left(\frac{1}{2} x^T W x \right) \right]$$

Log bazli turev alinca log icindeki hersey oldugu gibi bolume gider, ve log icindeki turevi alinirak bolume koyulur. Fakat log icine dikkatli bakarsak bu zaten $Z(W)$ 'nin tanimidir, boylece denklemi temizleme sansi dogdu, bolume hemen $Z(W)$ deriz, ve turevi log'un icine uygulariz,

$$\begin{aligned} \frac{\partial}{\partial w_{ij}} \ln Z(W) &= \frac{1}{Z(W)} \left[\sum_x \frac{\partial}{\partial w_{ij}} \exp \left(\frac{1}{2} x^T W x \right) \right] \\ \frac{\partial}{\partial w_{ij}} \exp \left(\frac{1}{2} x^T W x \right) &= \frac{1}{2} \exp \left(\frac{1}{2} x^T W x \right) \frac{\partial}{\partial w_{ij}} x^T W x \end{aligned} \quad (2)$$

(2)'in icindeki bolumu acalim,

$$\frac{\partial}{\partial w_{ij}} x^T W x = x_i x_j$$

Simdi (2)'ye geri koyalim,

$$\begin{aligned}
&= \frac{1}{2} \exp \left(\frac{1}{2} \mathbf{x}^T \mathbf{W} \mathbf{x} \right) x_i x_j \\
\frac{\partial}{\partial w_{ij}} \ln Z(\mathbf{W}) &= \frac{1}{Z(\mathbf{W})} \left[\sum_{\mathbf{x}} \frac{1}{2} \exp \left(\frac{1}{2} \mathbf{x}^T \mathbf{W} \mathbf{x} \right) x_i x_j \right] \\
&= \frac{1}{2} \sum_{\mathbf{x}} \frac{1}{Z(\mathbf{W})} \exp \left(\frac{1}{2} \mathbf{x}^T \mathbf{W} \mathbf{x} \right) x_i x_j \\
&= \frac{1}{2} \sum_{\mathbf{x}} P(\mathbf{x}; \mathbf{W}) x_i x_j
\end{aligned}$$

Ustteki son ifadede bir kisaltma kullanalim,

$$\sum_{\mathbf{x}} P(\mathbf{x}; \mathbf{W}) x_i x_j = \langle x_i, x_j \rangle_{P(\mathbf{x}; \mathbf{W})}$$

Artik $\ln Z(\mathbf{W})$ 'nin turevini biliyoruz. O zaman tum log olurlugun turevine (1) donebiliriz,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial w_{ij}} &= \sum_{n=1}^N \left[\frac{\partial}{\partial w_{ij}} \frac{1}{2} \mathbf{x}^{(n)T} \mathbf{W} \mathbf{x}^{(n)} - \frac{\partial}{\partial w_{ij}} \ln Z(\mathbf{W}) \right] \\
&= \sum_{n=1}^N \left[\frac{1}{2} x_i^{(n)T} x_j^{(n)} - \frac{\partial}{\partial w_{ij}} \ln Z(\mathbf{W}) \right] \\
&= \sum_{n=1}^N \left[\frac{1}{2} x_i^{(n)T} x_j^{(n)} - \frac{1}{2} \langle x_i x_j \rangle_{P(\mathbf{x}; \mathbf{W})} \right]
\end{aligned}$$

1/2 sabitlerini atalim,

$$= \sum_{n=1}^N \left[x_i^{(n)T} x_j^{(n)} - \langle x_i x_j \rangle_{P(\mathbf{x}; \mathbf{W})} \right]$$

Eger

$$\langle x_i x_j \rangle_{\text{Data}} = \frac{1}{N} \sum_{n=1}^N x_i^{(n)T} x_j^{(n)}$$

olarak alırsak, esitligin sag tarafi verisel kovaryansi (empirical covariance) temsil eder. Duzenleyince,

$$N \cdot \langle x_i x_j \rangle_{\text{Data}} = \sum_{n=1}^N x_i^{(n)T} x_j^{(n)}$$

simdi esitligin sag tarafi uc ustteki formule geri koyulabilir,

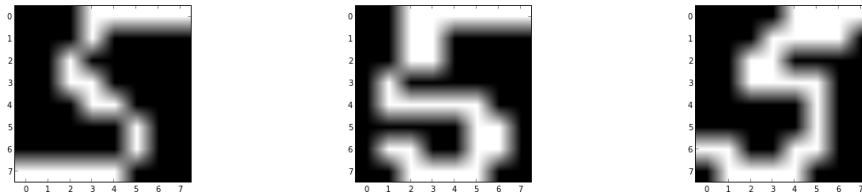
$$\frac{\partial \mathcal{L}}{\partial w_{ij}} = N \left[\langle x_i x_j \rangle_{\text{Data}} - \langle x_i x_j \rangle_{P(x;W)} \right]$$

Bu bir gradyan guncelleme formulu olarak gorulebilir.

```
Y = np.loadtxt('../stat/stat_mixbern/binarydigits.txt')
label = np.ravel(np.loadtxt('../stat/stat_mixbern/binarydigitlabels.txt'))
Y5 = Y[label==5]
plt.imshow(Y5[0,:].reshape((8,8),order='C'), cmap=plt.cm.gray)
plt.savefig('boltzmann_01.png')

plt.imshow(Y5[1,:].reshape((8,8),order='C'), cmap=plt.cm.gray)
plt.savefig('boltzmann_02.png')

plt.imshow(Y5[2,:].reshape((8,8),order='C'), cmap=plt.cm.gray)
plt.savefig('boltzmann_03.png')
```



Information Theory, Inference and Learning Algorithms

<http://nbviewer.ipython.org/gist/aflaxman/7d946762ee99daf739f1>