DEPARTAMENTO DE
MATEMÁTICA

The Principles of Deepn Learning Theory
Capítulo 4: RG-flow de pré-ativações

Felipe Kaminsky Riffel
Universidade Federal de Santa Catarina

9 de maio de 2025

Sumário

MTM

Introdução

Dataset com inputs n_0-dimensionais e com N_D amostras:

$$D = \{x_{i,\alpha}\}_{i=1,\dots,n_0;\alpha=1,\dots,N_D}$$

Pré-ativações da 1ª camada são dadas por:

$$z_{i,\alpha}^{(1)} \equiv z_i^{(1)}(x_\alpha) = b_i^{(1)} + \sum_{j=1}^{n_0} W_{ij}^{(1)}\, x_{j,\alpha}, \quad \text{para } i = 1, \dots, n_1.$$

Na inicialização, estabelecemos $b^{(1)}, W^{(1)}$ independentemente distribuídas e tais que:

$$\mathbb{E}[b_i^{(1)}] = \mathbb{E}[W_{i_1 j_1}^{(1)}] = 0$$

e

$$\mathbb{E}[b_i^{(1)} b_j^{(1)}] = \delta_i j C_b^{(1)}$$

$$\mathbb{E}[W_{i_1 j_1}^{(1)} W_{i_2 j_2}^{(1)}] = \delta_{i_1 i_2} \delta_{j_1 j_2} \frac{C_W^{(1)}}{n_0}$$

Estamos interessados em:

$$p(z^{(1)}|D) = p(z^{(1)}(x_1), \ldots, z^{(1)}(x_{N_D}))$$

Vamos fazer a demonstração via 2 caminhos:

- combinatório, via contrações de Wick e correlatores;
- algébrico, via transformada de Hubbard-Stratonovich.

Vemos primeiramente que:

$$\mathbb{E}[z_{i,\alpha}^{(1)}] = \mathbb{E}[b_i^{(1)} + \sum_{j=1}^{n_0} W_{ij}^{(1)} x_{j,\alpha_1}]$$

$$= \mathbb{E}[b_i^{(1)}] + \sum_{j=1}^{n_0} \mathbb{E}[W_{ij}] x_{j,\alpha} = 0$$

dado que as distribuições tem média 0.
É "fácil ver que"todos os correlatores de ordem ímpar de $p(z^{(1)}|D)$ são zerados, porque sempre sobra um $b^{(1)}$ ou $W^{(1)}$ em cada termo que zera a conta.

Para um correlator de dois pontos:

$$\mathbb{E}\left[z^{(1)}_{i_1;\alpha_1}\, z^{(1)}_{i_2;\alpha_2}\right] = \mathbb{E}\left[\left(b^{(1)}_{i_1} + \sum_{j_1=1}^{n_0} W^{(1)}_{i_1 j_1} x_{j_1;\alpha_1}\right)\left(b^{(1)}_{i_2} + \sum_{j_2=1}^{n_0} W^{(1)}_{i_2 j_2} x_{j_2;\alpha_2}\right)\right]$$

Para um correlator de dois pontos:

$$\mathbb{E}\left[z^{(1)}_{i_1;\alpha_1}\, z^{(1)}_{i_2;\alpha_2}\right] = \mathbb{E}\left[\left(b^{(1)}_{i_1} + \sum_{j_1=1}^{n_0} W^{(1)}_{i_1 j_1} x_{j_1;\alpha_1}\right)\left(b^{(1)}_{i_2} + \sum_{j_2=1}^{n_0} W^{(1)}_{i_2 j_2} x_{j_2;\alpha_2}\right)\right]$$

$$= \mathbb{E}[b_{i_1} b_{i_2}] + \sum_{j_1} \mathbb{E}[b_{i_2} W_{i_1,j_1}] x_{j_1,\alpha_1} + \sum_{j_2} \mathbb{E}[b_{i_1} W_{i_2,j_2}] x_{j_2,\alpha_2}$$

$$+ \mathbb{E}[(\sum_{j_1=1}^{n_0} W^{(1)}_{i_1 j_1} x_{j_1;\alpha_1})(\sum_{j_2=1}^{n_0} W^{(1)}_{i_2 j_2} x_{j_2;\alpha_2})]$$

Para um correlator de dois pontos:

$$\mathbb{E}\left[z^{(1)}_{i_1;\alpha_1}\, z^{(1)}_{i_2;\alpha_2}\right] = \mathbb{E}\left[\left(b^{(1)}_{i_1} + \sum_{j_1=1}^{n_0} W^{(1)}_{i_1 j_1} x_{j_1;\alpha_1}\right)\left(b^{(1)}_{i_2} + \sum_{j_2=1}^{n_0} W^{(1)}_{i_2 j_2} x_{j_2;\alpha_2}\right)\right]$$

$$= \mathbb{E}[b_{i_1} b_{i_2}] + \sum_{j_1} \mathbb{E}[b_{i_2} W_{i_1,j_1}] x_{j_1,\alpha_1} + \sum_{j_2} \mathbb{E}[b_{i_1} W_{i_2,j_2}] x_{j_2,\alpha_2}$$

$$+ \mathbb{E}[(\sum_{j_1=1}^{n_0} W^{(1)}_{i_1 j_1} x_{j_1;\alpha_1})(\sum_{j_2=1}^{n_0} W^{(1)}_{i_2 j_2} x_{j_2;\alpha_2})]$$

$$= \delta_{i_1 i_2} C^{(1)}_b + \sum_{j_1,j_2=1}^{n_0} \mathbb{E}[W_{i_1 j_1} W_{i_2 j_2}] x_{j,\alpha_1} x_{j,\alpha_2}$$

Para um correlator de dois pontos:

$$
\begin{aligned}
\mathbb{E}\left[z^{(1)}_{i_1;\alpha_1}\, z^{(1)}_{i_2;\alpha_2}\right] &= \mathbb{E}\left[\left(b^{(1)}_{i_1} + \sum_{j_1=1}^{n_0} W^{(1)}_{i_1 j_1} x_{j_1;\alpha_1}\right)\left(b^{(1)}_{i_2} + \sum_{j_2=1}^{n_0} W^{(1)}_{i_2 j_2} x_{j_2;\alpha_2}\right)\right] \\
&= \mathbb{E}[b_{i_1} b_{i_2}] + \sum_{j_1} \mathbb{E}[b_{i_2} W_{i_1,j_1}] x_{j_1,\alpha_1} + \sum_{j_2} \mathbb{E}[b_{i_1} W_{i_2,j_2}] x_{j_2,\alpha_2} \\
&\quad + \mathbb{E}[(\sum_{j_1=1}^{n_0} W^{(1)}_{i_1 j_1} x_{j_1;\alpha_1})(\sum_{j_2=1}^{n_0} W^{(1)}_{i_2 j_2} x_{j_2;\alpha_2})] \\
&= \delta_{i_1 i_2} C^{(1)}_b + \sum_{j_1,j_2=1}^{n_0} \mathbb{E}[W_{i_1 j_1} W_{i_2 j_2}] x_{j,\alpha_1} x_{j,\alpha_2} \\
&= \delta_{i_1 i_2}\left(C^{(1)}_b + C^{(1)}_W \frac{1}{n_0} \sum_{j=1}^{n_0} x_{j;\alpha_1} x_{j;\alpha_2}\right)
\end{aligned}
$$

Definimos a seguinte expressão como a **métrica de primeira camada**:

$$G^{(1)}_{\alpha_1\alpha_2} := C^{(1)}_b + C^{(1)}_W \frac{1}{n_0} \sum_{j=1}^{n_0} x_{j;\alpha_1} x_{j;\alpha_2}$$

Definimos a seguinte expressão como a **métrica de primeira camada**:

$$G^{(1)}_{\alpha_1\alpha_2} := C^{(1)}_b + C^{(1)}_W \frac{1}{n_0} \sum_{j=1}^{n_0} x_{j;\alpha_1} x_{j;\alpha_2}$$

Note que é uma função de $x_{\alpha_1}, x_{\alpha_2}$. Assim, nossa igualdade acima se traduz como:

Definimos a seguinte expressão como a **métrica de primeira camada**:

$$G^{(1)}_{\alpha_1\alpha_2} := C_b^{(1)} + C_W^{(1)} \frac{1}{n_0} \sum_{j=1}^{n_0} x_{j;\alpha_1} x_{j;\alpha_2}$$

Note que é uma função de $x_{\alpha_1}, x_{\alpha_2}$. Assim, nossa igualdade acima se traduz como:

$$\mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} \, z^{(1)}_{i_2;\alpha_2} \right] = \delta_{i_1 i_2} G^{(1)}_{\alpha_1\alpha_2}$$

Seguindo a mesma ideia:

$$\mathbb{E}[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_3;\alpha_3} z^{(1)}_{i_4;\alpha_4}] = \delta_{i_1 i_2}\delta_{i_3 i_4} G^{(1)}_{\alpha_1\alpha_2} G^{(1)}_{\alpha_3\alpha_4} + \delta_{i_1 i_3}\delta_{i_2 i_4} G^{(1)}_{\alpha_1\alpha_3} G^{(1)}_{\alpha_2\alpha_4} + \delta_{i_1 i_4}\delta_{i_2 i_3} G^{(1)}_{\alpha_1\alpha_4} G^{(1)}_{\alpha_2\alpha_3}$$

Seguindo a mesma ideia:

$$
\begin{aligned}
\mathbb{E}[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_3;\alpha_3} z^{(1)}_{i_4;\alpha_4}] &= \delta_{i_1 i_2}\delta_{i_3 i_4} G^{(1)}_{\alpha_1\alpha_2} G^{(1)}_{\alpha_3\alpha_4} + \delta_{i_1 i_3}\delta_{i_2 i_4} G^{(1)}_{\alpha_1\alpha_3} G^{(1)}_{\alpha_2\alpha_4} \\
&\quad + \delta_{i_1 i_4}\delta_{i_2 i_3} G^{(1)}_{\alpha_1\alpha_4} G^{(1)}_{\alpha_2\alpha_3} \\
&= \mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_2;\alpha_2}\right] \mathbb{E}\left[z^{(1)}_{i_3;\alpha_3} z^{(1)}_{i_4;\alpha_4}\right] + \mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_3;\alpha_3}\right] \mathbb{E}\left[z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_4;\alpha_4}\right] \\
&\quad + \mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_4;\alpha_4}\right] \mathbb{E}\left[z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_3;\alpha_3}\right]
\end{aligned}
$$

Seguindo a mesma ideia:

$$
\begin{aligned}
\mathbb{E}[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_3;\alpha_3} z^{(1)}_{i_4;\alpha_4}] &= \delta_{i_1 i_2}\delta_{i_3 i_4} G^{(1)}_{\alpha_1\alpha_2} G^{(1)}_{\alpha_3\alpha_4} + \delta_{i_1 i_3}\delta_{i_2 i_4} G^{(1)}_{\alpha_1\alpha_3} G^{(1)}_{\alpha_2\alpha_4} \\
&\quad + \delta_{i_1 i_4}\delta_{i_2 i_3} G^{(1)}_{\alpha_1\alpha_4} G^{(1)}_{\alpha_2\alpha_3} \\
&= \mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_2;\alpha_2}\right]\mathbb{E}\left[z^{(1)}_{i_3;\alpha_3} z^{(1)}_{i_4;\alpha_4}\right] + \mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_3;\alpha_3}\right]\mathbb{E}\left[z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_4;\alpha_4}\right] \\
&\quad + \mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_4;\alpha_4}\right]\mathbb{E}\left[z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_3;\alpha_3}\right]
\end{aligned}
$$

Lembrando de discussão anterior, isso significa que

$$
\mathbb{E}[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_3;\alpha_3} z^{(1)}_{i_4;\alpha_4}]\big|_{\text{connected}} = 0
$$

de modo que a distribuição é Gaussiana.

Seguindo a mesma ideia:

$$\begin{aligned}\mathbb{E}[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_3;\alpha_3} z^{(1)}_{i_4;\alpha_4}] &= \delta_{i_1 i_2}\delta_{i_3 i_4} G^{(1)}_{\alpha_1\alpha_2} G^{(1)}_{\alpha_3\alpha_4} + \delta_{i_1 i_3}\delta_{i_2 i_4} G^{(1)}_{\alpha_1\alpha_3} G^{(1)}_{\alpha_2\alpha_4} \\ &\quad + \delta_{i_1 i_4}\delta_{i_2 i_3} G^{(1)}_{\alpha_1\alpha_4} G^{(1)}_{\alpha_2\alpha_3} \\ &= \mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_2;\alpha_2}\right] \mathbb{E}\left[z^{(1)}_{i_3;\alpha_3} z^{(1)}_{i_4;\alpha_4}\right] + \mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_3;\alpha_3}\right] \mathbb{E}\left[z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_4;\alpha_4}\right] \\ &\quad + \mathbb{E}\left[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_4;\alpha_4}\right] \mathbb{E}\left[z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_3;\alpha_3}\right]\end{aligned}$$

Lembrando de discussão anterior, isso significa que

$$\mathbb{E}[z^{(1)}_{i_1;\alpha_1} z^{(1)}_{i_2;\alpha_2} z^{(1)}_{i_3;\alpha_3} z^{(1)}_{i_4;\alpha_4}]\big|_{\text{connected}} = 0$$

de modo que a distribuição é Gaussiana.
De forma similar, podemos verificar que todos os correlatores conexos de ordem mais alta zeram, e portanto, podem ser gerados por uma Gaussiana.

Lembrando, uma distribuição Gaussiana multivariada z de média s e matriz de covariância K pode ser escrita como

$$p(z) = \frac{1}{\sqrt{|2\pi K|}} \exp\left[-\frac{1}{2}\sum_{\mu,\nu=1}^{N}(z-s)_\mu K^{\mu\nu}(z-s)_\nu\right]$$
$$= \frac{1}{\sqrt{|2\pi K|}} \exp\left[-\frac{1}{2}(z-s)^T K^{-1}(z-s)\right]$$

onde $K^{\mu\nu} = (K^{-1})_{\mu\nu}$.

Para escrever a distribuição de primeira camada, precisamos da inversa de $\delta_{i_1 i_2} G^{\alpha_1 \alpha_2}_{(1)}$, para cada par de índices i_1, i_2, que satisfaça

Para escrever a distribuição de primeira camada, precisamos da inversa de $\delta_{i_1 i_2} G_{(1)}^{\alpha_1 \alpha_2}$, para cada par de índices i_1, i_2, que satisfaça

$$\sum_{j=1}^{n_1} \sum_{\beta \in D} (\delta_{i_1 j} G_{(1)}^{\alpha_1 \beta})(\delta_{i_2 j} G^{(1)}_{\beta \alpha_2}) = \delta_{i_1 i_2} \delta^{\alpha_1}_{\alpha_2}$$

Para escrever a distribuição de primeira camada, precisamos da inversa de $\delta_{i_1 i_2} G_{(1)}^{\alpha_1 \alpha_2}$, para cada par de índices i_1, i_2, que satisfaça

$$\sum_{j=1}^{n_1} \sum_{\beta \in D} (\delta_{i_1 j} G_{(1)}^{\alpha_1 \beta})(\delta_{i_2 j} G^{(1)}_{\beta \alpha_2}) = \delta_{i_1 i_2} \delta^{\alpha_1}_{\alpha_2}$$

$G_{(1)}^{\alpha\beta}$ denota as entradas da matriz $(G^{(1)})^{-1}_{\alpha\beta} \in \mathbb{R}^{N_D}$, com as métricas $G^{(1)}_{\alpha\beta}$ para duas entradas x_α, x_β, satisfazendo

$$\sum_{\beta \in D} G_{(1)}^{\alpha_1 \beta} G^{(1)}_{\beta \alpha_2} = \delta^{\alpha_1}_{\alpha_2}$$

Podemos, então, escrever a distribuição $p(z^{(1)}|D)$ na forma:

$$p(z^{(1)}|D) = \frac{1}{Z} e^{-S(z^{(1)})}$$

onde

Podemos, então, escrever a distribuição $p(z^{(1)}|D)$ na forma:

$$p(z^{(1)}|D) = \frac{1}{Z} e^{-S(z^{(1)})}$$

onde S é a *ação quadrática*

$$S(z^{(1)}) = \frac{1}{2} \sum_{i=1}^{n_1} \sum_{\alpha_1, \alpha_2 \in D} z_{i;\alpha_1}^{(1)} G_{(1)}^{\alpha_1 \alpha_2} z_{i;\alpha_2}^{(1)}$$

Podemos, então, escrever a distribuição $p(z^{(1)}|D)$ na forma:

$$p(z^{(1)}|D) = \frac{1}{Z} e^{-S(z^{(1)})}$$

onde S é a *ação quadrática*

$$S(z^{(1)}) = \frac{1}{2} \sum_{i=1}^{n_1} \sum_{\alpha_1,\alpha_2 \in D} z^{(1)}_{i;\alpha_1} G^{\alpha_1\alpha_2}_{(1)} z^{(1)}_{i;\alpha_2}$$

$$\left(= \sum_{i,j=1}^{n_1} \sum_{\alpha_1,\alpha_2 \in D} z^{(1)}_{i;\alpha_1} \delta_{ij} G^{\alpha_1\alpha_2}_{(1)} z^{(1)}_{j;\alpha_2} \right)$$

Podemos, então, escrever a distribuição $p(z^{(1)}|D)$ na forma:

$$p(z^{(1)}|D) = \frac{1}{Z} e^{-S(z^{(1)})}$$

onde S é a *ação quadrática*

$$S(z^{(1)}) = \frac{1}{2} \sum_{i=1}^{n_1} \sum_{\alpha_1,\alpha_2 \in D} z_{i;\alpha_1}^{(1)} G_{(1)}^{\alpha_1\alpha_2} z_{i;\alpha_2}^{(1)}$$

$$\left(= \sum_{i,j=1}^{n_1} \sum_{\alpha_1,\alpha_2 \in D} z_{i;\alpha_1}^{(1)} \delta_{ij} G_{(1)}^{\alpha_1\alpha_2} z_{j;\alpha_2}^{(1)} \right)$$

e Z é a *função de partição*:

$$Z = \int \left[\prod_{i,\alpha} dz_{i,\alpha}^{(1)} \right] e^{-S(z^{(1)})}$$

Como visto no capítulo anterior, podemos expressar a distribuição na forma

$$p(z \mid D) = \int \left[\prod_j db_j \, p(b_j)\right] \left[\prod_{j,k} dW_{jk} \, p(W_{jk})\right] \prod_{j,\alpha} \delta\left(z_{j;\alpha} - b_i - \sum_j W_{jk} x_{j;\alpha}\right)$$

suprimindo por hora os sobrescritos (1).

Por hipótese:

$$p(b_j) = \frac{1}{\sqrt{2\pi C_b}} \exp(-\frac{b_j^2}{2C_b})$$

$$p(W_{jk}) = \frac{1}{\sqrt{2\pi C_W/n_0}} \exp(-\frac{n_0 W_{jk}^2}{2C_W})$$

Vamos usar a **transformação de Hubbard-Stratonovich**.
I.e., a representação integral do delta de Dirac:

$$\delta(z-a) = \int \frac{d\Lambda}{2\pi} e^{i\Lambda(z-a)}$$

Obtemos:

$$p(z \mid D) = \int \left[\prod_j \frac{db_j}{\sqrt{2\pi C_b}} \exp(-\frac{b_j^2}{2C_b})\right] \left[\prod_{j,k} \frac{dW_{jk}}{\sqrt{2\pi C_W/n_0}} \exp(-\frac{n_0 W_{jk}^2}{2C_W})\right] \left[\prod_{j,\alpha} \int \frac{d\Lambda}{2\pi} e^{i\Lambda(z-a)}\right]$$

Obtemos:

$$p(z \mid D) = \int \left[\prod_j \frac{db_j}{\sqrt{2\pi C_b}} \exp(-\frac{b_j^2}{2C_b})\right] \left[\prod_{j,k} \frac{dW_{jk}}{\sqrt{2\pi C_W/n_0}} \exp(-\frac{n_0 W_{jk}^2}{2C_W})\right] \left[\prod_{j,\alpha} \int \frac{d\Lambda}{2\pi} e^{i\Lambda(z-a)}\right]$$

$$= \int \left[\prod_j \frac{db_i}{\sqrt{2\pi C_b}}\right] \left[\prod_{i,j} \frac{dW_{ij}}{\sqrt{2\pi C_W/n_0}}\right] \left[\prod_{i,\alpha} \frac{d\Lambda_i^\alpha}{2\pi}\right]$$

$$\times \exp\left[-\sum_j \frac{b_j^2}{2C_b} - n_0 \sum_{jk} \frac{W_{jk}^2}{2C_W} + i \sum_{j,\alpha} \Lambda_j^\alpha \left(z_{j;\alpha} - b_j - \sum_j W_{jk} x_{j;\alpha}\right)\right]$$

Completando quadrados:

$$\sum_j \frac{b_j^2}{2C_b} - n_0 \sum_{i,j} \frac{W_{jk}^2}{2C_W} + i \sum_{j,\alpha} \Lambda_j^\alpha \left(z_{j;\alpha} - b_j - \sum_j W_{jk} x_{k;\alpha} \right)$$

$$= -\frac{1}{2C_b} \sum_j \left(b_j + iC_b \sum_\alpha \Lambda_j^\alpha \right)^2 - \frac{n_0}{2C_W} \sum_{j,k} \left(W_{jk} + i \frac{C_W}{n_0} \sum_\alpha \Lambda_j^\alpha x_{k;\alpha} \right)^2$$

$$- \frac{C_W}{2n_0} \sum_{j,k} \left(\sum_\alpha \Lambda_j^\alpha x_{k;\alpha} \right)^2 - \frac{C_b}{2} \sum_j \left(\sum_\alpha \Lambda_j^\alpha \right)^2 + i \sum_{j,\alpha} \Lambda_j^\alpha z_{j;\alpha}$$

$$\int \left[\prod_j \frac{db_j}{\sqrt{2\pi C_b}}\right] \left[\prod_{i,j} \frac{dW_{ij}}{\sqrt{2\pi C_W/n_0}}\right] \left[\prod_{i,\alpha} \frac{d\Lambda_i^\alpha}{2\pi}\right]$$
$$\times \exp\left[-\sum_j \frac{b_j^2}{2C_b} - n_0 \sum_{jk} \frac{W_{jk}^2}{2C_W} + i \sum_{j,\alpha} \Lambda_j^\alpha \left(z_{j;\alpha} - b_j - \sum_j W_{jk} x_{j;\alpha}\right)\right]$$

$$
\begin{aligned}
&\int \left[\prod_j \frac{db_j}{\sqrt{2\pi C_b}}\right] \left[\prod_{i,j} \frac{dW_{ij}}{\sqrt{2\pi C_W/n_0}}\right] \left[\prod_{i,\alpha} \frac{d\Lambda_i^\alpha}{2\pi}\right] \\
&\times \exp\left[-\sum_j \frac{b_j^2}{2C_b} - n_0 \sum_{jk} \frac{W_{jk}^2}{2C_W} + i \sum_{j,\alpha} \Lambda_j^\alpha \left(z_{j;\alpha} - b_j - \sum_j W_{jk} x_{j;\alpha}\right)\right] \\
&= \int \left[\prod_j \frac{db_i}{\sqrt{2\pi C_b}}\right] \left[\prod_{i,j} \frac{dW_{ij}}{\sqrt{2\pi C_W/n_0}}\right] \left[\prod_{i,\alpha} \frac{d\Lambda_i^\alpha}{2\pi}\right] \\
&\exp\left[-\frac{1}{2C_b} \sum_j \left(b_j + iC_b \sum_\alpha \Lambda_j^\alpha\right)^2\right] \exp\left[-\frac{n_0}{2C_W} \sum_{j,k} \left(W_{jk} + i\frac{C_W}{n_0} \sum_\alpha \Lambda_j^\alpha x_{k;\alpha}\right)^2\right] \\
&\exp\left[-\frac{C_W}{2n_0} \sum_{j,k} \left(\sum_\alpha \Lambda_j^\alpha x_{k;\alpha}\right)^2 - \frac{C_b}{2} \sum_j \left(\sum_\alpha \Lambda_j^\alpha\right)^2 + i \sum_{j,\alpha} \Lambda_j^\alpha z_{j;\alpha}\right]
\end{aligned}
$$

$$
\begin{aligned}
= \int \left[\prod_j \frac{db_i}{\sqrt{2\pi C_b}}\right] \left[\prod_{i,j} \frac{dW_{ij}}{\sqrt{2\pi C_W/n_0}}\right] \left[\prod_{i,\alpha} \frac{d\Lambda_i^\alpha}{2\pi}\right] \\
\exp\left[-\frac{1}{2C_b}\sum_j \left(b_j + iC_b \sum_\alpha \Lambda_j^\alpha\right)^2\right] \exp\left[-\frac{n_0}{2C_W}\sum_{j,k}\left(W_{jk} + i\frac{C_W}{n_0}\sum_\alpha \Lambda_j^\alpha x_{k;\alpha}\right)^2\right] \\
\exp\left[-\frac{C_W}{2n_0}\sum_{j,k}\left(\sum_\alpha \Lambda_j^\alpha x_{k;\alpha}\right)^2 - \frac{C_b}{2}\sum_j\left(\sum_\alpha \Lambda_j^\alpha\right)^2 + i\sum_{j,\alpha}\Lambda_j^\alpha z_{j;\alpha}\right]
\end{aligned}
$$

$$= \int \left[\prod_j \frac{db_i}{\sqrt{2\pi C_b}}\right] \left[\prod_{i,j} \frac{dW_{ij}}{\sqrt{2\pi C_W/n_0}}\right] \left[\prod_{i,\alpha} \frac{d\Lambda_i^\alpha}{2\pi}\right]$$

$$\exp\left[-\frac{1}{2C_b}\sum_j \left(b_j + iC_b \sum_\alpha \Lambda_j^\alpha\right)^2\right] \exp\left[-\frac{n_0}{2C_W}\sum_{j,k}\left(W_{jk} + i\frac{C_W}{n_0}\sum_\alpha \Lambda_j^\alpha x_{k;\alpha}\right)^2\right]$$

$$\exp\left[-\frac{C_W}{2n_0}\sum_{j,k}\left(\sum_\alpha \Lambda_j^\alpha x_{k;\alpha}\right)^2 - \frac{C_b}{2}\sum_j\left(\sum_\alpha \Lambda_j^\alpha\right)^2 + i\sum_{j,\alpha}\Lambda_j^\alpha z_{j;\alpha}\right]$$

$$\times = \int \left[\prod_j \frac{db_i}{\sqrt{2\pi C_b}}\right] \exp\left[-\frac{1}{2C_b}\sum_j \left(b_j + iC_b \sum_\alpha \Lambda_j^\alpha\right)^2\right]$$

$$\times \int \left[\prod_{i,j} \frac{dW_{ij}}{\sqrt{2\pi C_W/n_0}}\right] \exp\left[-\frac{n_0}{2C_W}\sum_{j,k}\left(W_{jk} + i\frac{C_W}{n_0}\sum_\alpha \Lambda_j^\alpha x_{k;\alpha}\right)^2\right]$$

$$\int \left[\prod_{i,\alpha} \frac{d\Lambda_i^\alpha}{2\pi}\right] \exp\left[-\frac{C_W}{2n_0}\sum_{j,k}\left(\sum_\alpha \Lambda_j^\alpha x_{k;\alpha}\right)^2 - \frac{C_b}{2}\sum_j\left(\sum_\alpha \Lambda_j^\alpha\right)^2 + i\sum_{j,\alpha}\Lambda_j^\alpha z_{j;\alpha}\right]$$

Olhando bem, são integrais Gaussianas deslocadas por uma média:

$$\int \left[\prod_j \frac{db_j}{\sqrt{2\pi C_b}}\right] \exp\left[-\frac{1}{2C_b}\sum_j \left(b_j + iC_b \sum_\alpha \Lambda_j^\alpha\right)^2\right]$$
$$= \prod_j \int \left[\frac{db_j}{\sqrt{2\pi C_b}} \exp\left(-\frac{(b_j - \mu_b)^2}{2C_b}\right)\right]$$

com a média $\mu_b = -iC_b \sum_\alpha \Lambda_j^\alpha$. Assim, cada integral dessa é igual a 1, e some da representação.

Olhando bem, são integrais Gaussianas deslocadas por uma média:

$$\int \left[\prod_j \frac{db_j}{\sqrt{2\pi C_b}}\right] \exp\left[-\frac{1}{2C_b}\sum_j \left(b_j + iC_b \sum_\alpha \Lambda_j^\alpha\right)^2\right]$$
$$= \prod_j \int \left[\frac{db_j}{\sqrt{2\pi C_b}} \exp\left(-\frac{(b_j - \mu_b)^2}{2C_b}\right)\right]$$

com a média $\mu_b = -iC_b \sum_\alpha \Lambda_j^\alpha$. Assim, cada integral dessa é igual a 1, e some da representação.

De forma análoga, a integral envolvendo W_{ij} vira 1.

Então...

$$p(z \mid D) = \int \left[\prod_{i,\alpha} \frac{d\Lambda_i^\alpha}{2\pi}\right] \exp\left[-\frac{C_W}{2n_0}\sum_{j,k}\left(\sum_\alpha \Lambda_j^\alpha x_{k;\alpha}\right)^2 - \frac{C_b}{2}\sum_j\left(\sum_\alpha \Lambda_j^\alpha\right)^2 + i\sum_{j,\alpha}\Lambda_j^\alpha z_{j;\alpha}\right]$$

$$= \int \left[\prod_{i,\alpha} \frac{d\Lambda_i^\alpha}{2\pi}\right] \exp\left[-\frac{C_W}{2n_0}\sum_{j,k}\sum_{\alpha_1\alpha_2}\Lambda_j^{\alpha_1}\Lambda_j^{\alpha_2}x_{k;\alpha_1}x_{k;\alpha_2} - \frac{C_b}{2}\sum_j\sum_{\alpha_1\alpha_2}\Lambda_j^{\alpha_1}\Lambda_j^{\alpha_2} + i\sum_{j,\alpha}\Lambda_j^\alpha z_{j;\alpha}\right]$$

$$= \int \left[\prod_{i,\alpha} \frac{d\Lambda_{i\alpha}}{2\pi}\right] \exp\left[-\frac{1}{2}\sum_{j,\alpha_1\alpha_2}\Lambda_j^{\alpha_1}\Lambda_j^{\alpha_2}\left(C_b + C_W\frac{1}{n_0}\sum_k x_{k;\alpha_1}x_{k;\alpha_2}\right) + i\sum_{j,\alpha}\Lambda_j^\alpha z_{j;\alpha}\right]$$

$$= \int \left[\prod_{i,\alpha} \frac{d\Lambda_{i\alpha}}{2\pi}\right] \exp\left[-\frac{1}{2}\sum_{j,\alpha_1\alpha_2}\Lambda_j^{\alpha_1}\Lambda_j^{\alpha_2}G_{\alpha_1\alpha_2} + i\sum_{j,\alpha}\Lambda_j^\alpha z_{j;\alpha}\right]$$

$$\mathbb{E}\left[\sigma\left(z^{(1)}_{i_1;\alpha_1}\right)\sigma\left(z^{(1)}_{i_1;\alpha_2}\right)\right]$$

$$= \int \left[\prod_{i=1}^{n_1} \prod_{\alpha \in D} \frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]$$

$$\mathbb{E}\left[\sigma\left(z^{(1)}_{i_1;\alpha_1}\right)\sigma\left(z^{(1)}_{i_1;\alpha_2}\right)\right]$$

$$= \int \left[\prod_{i=1}^{n_1}\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right] \times \exp\left(-\frac{1}{2}\sum_{j=1}^{n_1}\sum_{\beta_1,\beta_2\in D} G^{(1)}_{\beta_1\beta_2} z_{j;\beta_1} z_{j;\beta_2}\right)\sigma(z_{i_1;\alpha_1})\sigma(z_{i_1;\alpha_2})$$

$$\mathbb{E}\left[\sigma\left(z^{(1)}_{i_1;\alpha_1}\right)\sigma\left(z^{(1)}_{i_1;\alpha_2}\right)\right]$$

$$=\int\left[\prod_{i=1}^{n_1}\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\times\exp\left(-\frac{1}{2}\sum_{j=1}^{n_1}\sum_{\beta_1,\beta_2\in D}G^{(1)}_{\beta_1\beta_2}z_{j;\beta_1}z_{j;\beta_2}\right)\sigma(z_{i_1;\alpha_1})\sigma(z_{i_1;\alpha_2})$$

$$=\left\{\prod_{i\neq i_1}\int\left[\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left[-\frac{1}{2}\sum_{\beta_1,\beta_2\in D}G^{(1)}_{\beta_1\beta_2}z_{i;\beta_1}z_{i;\beta_2}\right]\right\}$$

$$\mathbb{E}\left[\sigma\left(z^{(1)}_{i_1;\alpha_1}\right)\sigma\left(z^{(1)}_{i_1;\alpha_2}\right)\right]$$

$$= \int \left[\prod_{i=1}^{n_1}\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right] \times \exp\left(-\frac{1}{2}\sum_{j=1}^{n_1}\sum_{\beta_1,\beta_2\in D} G^{(1)}_{\beta_1\beta_2} z_{j;\beta_1} z_{j;\beta_2}\right)\sigma(z_{i_1;\alpha_1})\sigma(z_{i_1;\alpha_2})$$

$$= \left\{\prod_{i\neq i_1}\int\left[\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left[-\frac{1}{2}\sum_{\beta_1,\beta_2\in D} G^{(1)}_{\beta_1\beta_2} z_{i;\beta_1} z_{i;\beta_2}\right]\right\}$$

$$\times \int\left[\prod_{\alpha\in D}\frac{dz_{i_1;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left(-\frac{1}{2}\sum_{\beta_1,\beta_2\in D} G^{\beta_1\beta_2}_{(1)} z_{i_1;\beta_1} z_{i_1;\beta_2}\right)\sigma(z_{i_1;\alpha_1})\sigma(z_{i_1;\alpha_2})$$

$$
\begin{aligned}
&\mathbb{E}\left[\sigma\left(z^{(1)}_{i_1;\alpha_1}\right)\sigma\left(z^{(1)}_{i_1;\alpha_2}\right)\right] \\
&= \int \left[\prod_{i=1}^{n_1}\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right] \times \exp\left(-\frac{1}{2}\sum_{j=1}^{n_1}\sum_{\beta_1,\beta_2\in D} G^{(1)}_{\beta_1\beta_2} z_{j;\beta_1} z_{j;\beta_2}\right)\sigma(z_{i_1;\alpha_1})\sigma(z_{i_1;\alpha_2}) \\
&= \left\{\prod_{i\neq i_1}\int\left[\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left[-\frac{1}{2}\sum_{\beta_1,\beta_2\in D} G^{(1)}_{\beta_1\beta_2} z_{i;\beta_1} z_{i;\beta_2}\right]\right\} \\
&\times \int\left[\prod_{\alpha\in D}\frac{dz_{i_1;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left(-\frac{1}{2}\sum_{\beta_1,\beta_2\in D} G^{\beta_1\beta_2}_{(1)} z_{i_1;\beta_1} z_{i_1;\beta_2}\right)\sigma(z_{i_1;\alpha_1})\sigma(z_{i_1;\alpha_2}) \\
&= \{1\} \times \left[\int\prod_{\alpha\in D}\frac{dz_{i_1\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left(-\frac{1}{2}\sum_{\beta_1,\beta_2\in D} G^{\beta_1\beta_2}_{(1)} z_{i_1\beta_1} z_{i_1\beta_2}\right)\sigma(z_{i_1\alpha_1})\sigma(z_{i_1\alpha_2})
\end{aligned}
$$

$$\mathbb{E}\left[\sigma\left(z^{(1)}_{i_1;\alpha_1}\right)\sigma\left(z^{(1)}_{i_1;\alpha_2}\right)\right]$$

$$= \int \left[\prod_{i=1}^{n_1}\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right] \times \exp\left(-\frac{1}{2}\sum_{j=1}^{n_1}\sum_{\beta_1,\beta_2\in D} G^{(1)}_{\beta_1\beta_2} z_{j;\beta_1} z_{j;\beta_2}\right)\sigma(z_{i_1;\alpha_1})\sigma(z_{i_1;\alpha_2})$$

$$= \left\{\prod_{i\neq i_1}\int\left[\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left[-\frac{1}{2}\sum_{\beta_1,\beta_2\in D} G^{(1)}_{\beta_1\beta_2} z_{i;\beta_1} z_{i;\beta_2}\right]\right\}$$

$$\times \int\left[\prod_{\alpha\in D}\frac{dz_{i_1;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left(-\frac{1}{2}\sum_{\beta_1,\beta_2\in D} G^{\beta_1\beta_2}_{(1)} z_{i_1;\beta_1} z_{i_1;\beta_2}\right)\sigma(z_{i_1;\alpha_1})\sigma(z_{i_1;\alpha_2})$$

$$= \{1\} \times \left[\int\prod_{\alpha\in D}\frac{dz_{i_1\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left(-\frac{1}{2}\sum_{\beta_1,\beta_2\in D} G^{\beta_1\beta_2}_{(1)} z_{i_1\beta_1} z_{i_1\beta_2}\right)\sigma(z_{i_1\alpha_1})\sigma(z_{i_1\alpha_2})$$

$$:= \langle\sigma(z_{i_1\alpha_1})\sigma(z_{i_1\alpha_2})\rangle_{G^{(1)}}.$$

Com a expressão do final, reintroduzimos a notação:

$$\left\langle F(z_{\alpha_1}, \ldots, z_{\alpha_m}) \right\rangle_g \equiv \int \left[\frac{\prod_{\alpha \in D} dz_\alpha}{\sqrt{|2\pi g|}} \right] \exp\left(-\frac{1}{2} \sum_{\beta_1, \beta_2 \in D} g^{\beta_1 \beta_2} z_{\beta_1} z_{\beta_2} \right) F(z_{\alpha_1}, \ldots, z_{\alpha_m})$$

Com a expressão do final, reintroduzimos a notação:

$$\Big\langle F(z_{\alpha_1}, \ldots, z_{\alpha_m}) \Big\rangle_g \equiv \int \left[\frac{\prod_{\alpha \in D} dz_\alpha}{\sqrt{|2\pi g|}} \right] \exp\left(-\frac{1}{2} \sum_{\beta_1, \beta_2 \in D} g^{\beta_1 \beta_2} z_{\beta_1} z_{\beta_2} \right) F(z_{\alpha_1}, \ldots, z_{\alpha_m})$$

e denotamos ainda:

$$\sigma_\alpha := \sigma(z_\alpha)$$

Com a expressão do final, reintroduzimos a notação:

$$\left\langle F(z_{\alpha_1}, \ldots, z_{\alpha_m}) \right\rangle_g \equiv \int \left[\frac{\prod_{\alpha \in D} dz_\alpha}{\sqrt{|2\pi g|}} \right] \exp\left(-\frac{1}{2} \sum_{\beta_1, \beta_2 \in D} g^{\beta_1 \beta_2} z_{\beta_1} z_{\beta_2} \right) F(z_{\alpha_1}, \ldots, z_{\alpha_m})$$

e denotamos ainda:

$$\sigma_\alpha := \sigma(z_\alpha)$$

Assim, podemos escrever

$$\mathbb{E}\left[\sigma\left(z^{(1)}_{i_1;\alpha_1} \right) \sigma\left(z^{(1)}_{i_1;\alpha_2} \right) \right] = \langle \sigma_{\alpha_1} \sigma_{\alpha_2} \rangle_{G^{(1)}}$$

É possível ainda generalizar para correlatores de mais pontos. Para todas as saídas do mesmo neurônio i_1:

$$\mathbb{E}\left[\sigma(z^{(1)}_{i_1;\alpha_1})\sigma(z^{(1)}_{i_1;\alpha_2})\sigma(z^{(1)}_{i_1;\alpha_3})\sigma(z^{(1)}_{i_1;\alpha_4})\right] = \langle \sigma_{\alpha_1}\sigma_{\alpha_2}\sigma_{\alpha_3}\sigma_{\alpha_4}\rangle_{G^{(1)}}$$

fazendo as exatas mesmas etapas.

Para $i_1 \neq i_2$

$$\mathbb{E}\left[\sigma\left(z^{(1)}_{i_1;\alpha_1}\right)\sigma\left(z^{(1)}_{i_1;\alpha_2}\right)\sigma\left(z^{(1)}_{i_2;\alpha_3}\right)\sigma\left(z^{(1)}_{i_2;\alpha_4}\right)\right]$$

$$= \left\{\prod_{i\notin\{i_1,i_2\}}\int\left[\prod_{\alpha\in D}\frac{dz_{i;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left(-\frac{1}{2}\sum_{\beta_1,\beta_2\in D}G^{\beta_1\beta_2}_{(1)}z_{i;\beta_1}z_{i;\beta_2}\right)\right\}$$

$$\times\int\left[\frac{\prod_{\alpha\in D}dz_{i_1;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left(-\frac{1}{2}\sum_{\beta_1,\beta_2\in D}G^{\beta_1\beta_2}_{(1)}z_{i_1;\beta_1}z_{i_1;\beta_2}\right)\sigma(z_{i_1;\alpha_1})\sigma(z_{i_1;\alpha_2})$$

$$\times\int\left[\frac{\prod_{\alpha\in D}dz_{i_2;\alpha}}{\sqrt{|2\pi G^{(1)}|}}\right]\exp\left(-\frac{1}{2}\sum_{\beta_1,\beta_2\in D}G^{\beta_1\beta_2}_{(1)}z_{i_2;\beta_1}z_{i_2;\beta_2}\right)\sigma(z_{i_2;\alpha_3})\sigma(z_{i_2;\alpha_4})$$

$$= \left\langle\sigma(z_{\alpha_1})\sigma(z_{\alpha_2})\right\rangle_{G^{(1)}}\left\langle\sigma(z_{\alpha_3})\sigma(z_{\alpha_4})\right\rangle_{G^{(1)}}$$

I.e.:

$$\mathbb{E}\left[\sigma\left(z^{(1)}_{i_1;\alpha_1}\right)\sigma\left(z^{(1)}_{i_1;\alpha_2}\right)\sigma\left(z^{(1)}_{i_2;\alpha_3}\right)\sigma\left(z^{(1)}_{i_2;\alpha_4}\right)\right] = \left\langle\sigma(z_{\alpha_1})\sigma(z_{\alpha_2})\right\rangle_{G^{(1)}}\left\langle\sigma(z_{\alpha_3})\sigma(z_{\alpha_4})\right\rangle_{G^{(1)}}$$

Isso ilustra a independência entre as distribuições de neurônios diferentes.

Obrigado!

Contato: riffel.felipe@grad.ufsc.br

Repositório com os materiais da apresentação:
https://github.com/felipekriffel/Mestrado