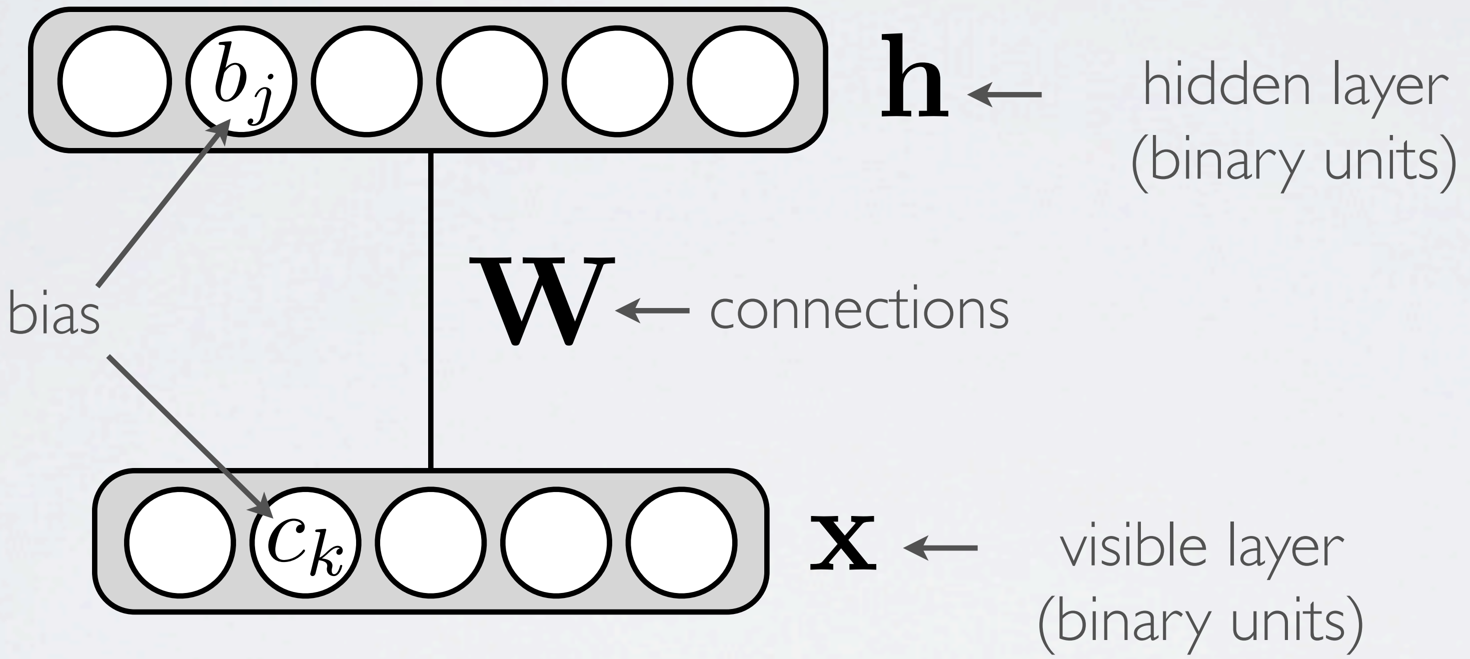
Neural networks

Restricted Boltzmann machine - free energy

RESTRICTED BOLTZMANN MACHINE

Topics: RBM, visible layer, hidden layer, energy function



Energy function:
$$\begin{aligned} E(\mathbf{x},\mathbf{h}) &= -\mathbf{h}^\top \mathbf{W}\mathbf{x} - \mathbf{c}^\top \mathbf{x} - \mathbf{b}^\top \mathbf{h} \\ &= -\sum_j \sum_k W_{j,k} h_j x_k - \sum_k c_k x_k - \sum_j b_j h_j \end{aligned}$$

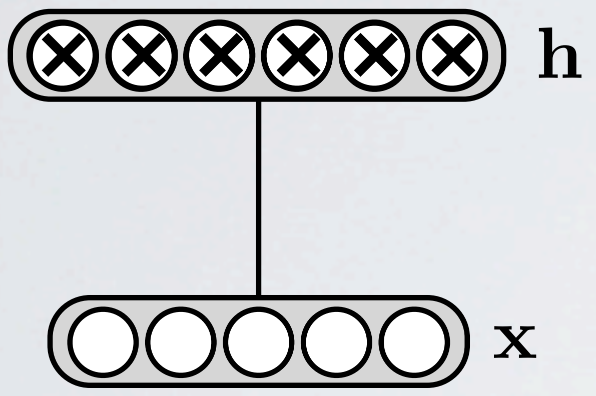
Distribution: $p(\mathbf{x},\mathbf{h}) = \exp(-E(\mathbf{x},\mathbf{h}))/Z$

partition function
(intractable)

FREE ENERGY

Topics: free energy

- What about $p(\mathbf{x})$?



$$
\begin{aligned}
p(\mathbf{x}) &= \sum_{\mathbf{h}\in\{0,1\}^H} p(\mathbf{x},\mathbf{h}) = \sum_{\mathbf{h}\in\{0,1\}^H} \exp(-E(\mathbf{x},\mathbf{h}))/Z \\
&= \exp\left(\mathbf{c}^\top\mathbf{x} + \sum_{j=1}^{H} \log(1+\exp(b_j + \mathbf{W}_{j\cdot}\mathbf{x}))\right)/Z \\
&= \exp(-F(\mathbf{x}))/Z
\end{aligned}
$$

free energy

$p(\mathbf{x})$

$$p(\mathbf{x}) \quad = \quad \sum_{\mathbf{h}\in\{0,1\}^H} \exp(\mathbf{h}^\top \mathbf{W}\mathbf{x} + \mathbf{c}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{h})/Z$$

$$
\begin{aligned}
p(\mathbf{x}) &= \sum_{\mathbf{h}\in\{0,1\}^H} \exp(\mathbf{h}^\top \mathbf{W}\mathbf{x} + \mathbf{c}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{h})/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \sum_{h_1\in\{0,1\}} \cdots \sum_{h_H\in\{0,1\}} \exp\left(\sum_j h_j \mathbf{W}_{j\cdot}\mathbf{x} + b_j h_j\right)/Z
\end{aligned}
$$

$$
\begin{aligned}
p(\mathbf{x}) &= \sum_{\mathbf{h}\in\{0,1\}^H} \exp(\mathbf{h}^\top \mathbf{W}\mathbf{x} + \mathbf{c}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{h})/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \sum_{h_1\in\{0,1\}} \cdots \sum_{h_H\in\{0,1\}} \exp\left(\sum_j h_j \mathbf{W}_{j\cdot}\mathbf{x} + b_j h_j\right) /Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \left(\sum_{h_1\in\{0,1\}} \exp(h_1 \mathbf{W}_{1\cdot}\mathbf{x} + b_1 h_1)\right) \ldots \left(\sum_{h_H\in\{0,1\}} \exp(h_H \mathbf{W}_{H\cdot}\mathbf{x} + b_H h_H)\right) /Z
\end{aligned}
$$

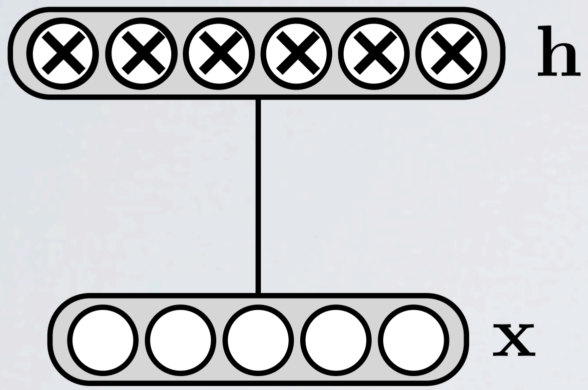
$$
\begin{aligned}
p(\mathbf{x}) &= \sum_{\mathbf{h}\in\{0,1\}^H} \exp(\mathbf{h}^\top \mathbf{W}\mathbf{x} + \mathbf{c}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{h})/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \sum_{h_1\in\{0,1\}} \cdots \sum_{h_H\in\{0,1\}} \exp\left(\sum_j h_j \mathbf{W}_{j\cdot}\mathbf{x} + b_j h_j\right)/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \left(\sum_{h_1\in\{0,1\}} \exp(h_1 \mathbf{W}_{1\cdot}\mathbf{x} + b_1 h_1)\right) \dots \left(\sum_{h_H\in\{0,1\}} \exp(h_H \mathbf{W}_{H\cdot}\mathbf{x} + b_H h_H)\right)/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \left(1 + \exp(b_1 + \mathbf{W}_{1\cdot}\mathbf{x})\right) \dots \left(1 + \exp(b_H + \mathbf{W}_{H\cdot}\mathbf{x})\right)/Z
\end{aligned}
$$

$$
\begin{aligned}
p(\mathbf{x}) &= \sum_{\mathbf{h}\in\{0,1\}^H} \exp(\mathbf{h}^\top \mathbf{W}\mathbf{x} + \mathbf{c}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{h})/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \sum_{h_1\in\{0,1\}} \cdots \sum_{h_H\in\{0,1\}} \exp\left(\sum_j h_j \mathbf{W}_{j\cdot}\mathbf{x} + b_j h_j\right)/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \left(\sum_{h_1\in\{0,1\}} \exp(h_1 \mathbf{W}_{1\cdot}\mathbf{x} + b_1 h_1)\right) \ldots \left(\sum_{h_H\in\{0,1\}} \exp(h_H \mathbf{W}_{H\cdot}\mathbf{x} + b_H h_H)\right)/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \left(1 + \exp(b_1 + \mathbf{W}_{1\cdot}\mathbf{x})\right) \ldots \left(1 + \exp(b_H + \mathbf{W}_{H\cdot}\mathbf{x})\right)/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \exp(\log(1 + \exp(b_1 + \mathbf{W}_{1\cdot}\mathbf{x}))) \ldots \exp(\log(1 + \exp(b_H + \mathbf{W}_{H\cdot}\mathbf{x})))/Z
\end{aligned}
$$

$$
\begin{aligned}
p(\mathbf{x}) &= \sum_{\mathbf{h}\in\{0,1\}^H} \exp(\mathbf{h}^\top \mathbf{W}\mathbf{x} + \mathbf{c}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{h})/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \sum_{h_1\in\{0,1\}} \cdots \sum_{h_H\in\{0,1\}} \exp\left(\sum_j h_j \mathbf{W}_{j\cdot}\mathbf{x} + b_j h_j\right)/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \left(\sum_{h_1\in\{0,1\}} \exp(h_1 \mathbf{W}_{1\cdot}\mathbf{x} + b_1 h_1)\right) \ldots \left(\sum_{h_H\in\{0,1\}} \exp(h_H \mathbf{W}_{H\cdot}\mathbf{x} + b_H h_H)\right)/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \left(1 + \exp(b_1 + \mathbf{W}_{1\cdot}\mathbf{x})\right) \ldots \left(1 + \exp(b_H + \mathbf{W}_{H\cdot}\mathbf{x})\right)/Z \\
&= \exp(\mathbf{c}^\top \mathbf{x}) \exp(\log(1 + \exp(b_1 + \mathbf{W}_{1\cdot}\mathbf{x}))) \ldots \exp(\log(1 + \exp(b_H + \mathbf{W}_{H\cdot}\mathbf{x})))/Z \\
&= \exp\left(\mathbf{c}^\top \mathbf{x} + \sum_{j=1}^{H} \log(1 + \exp(b_j + \mathbf{W}_{j\cdot}\mathbf{x}))\right)/Z
\end{aligned}
$$

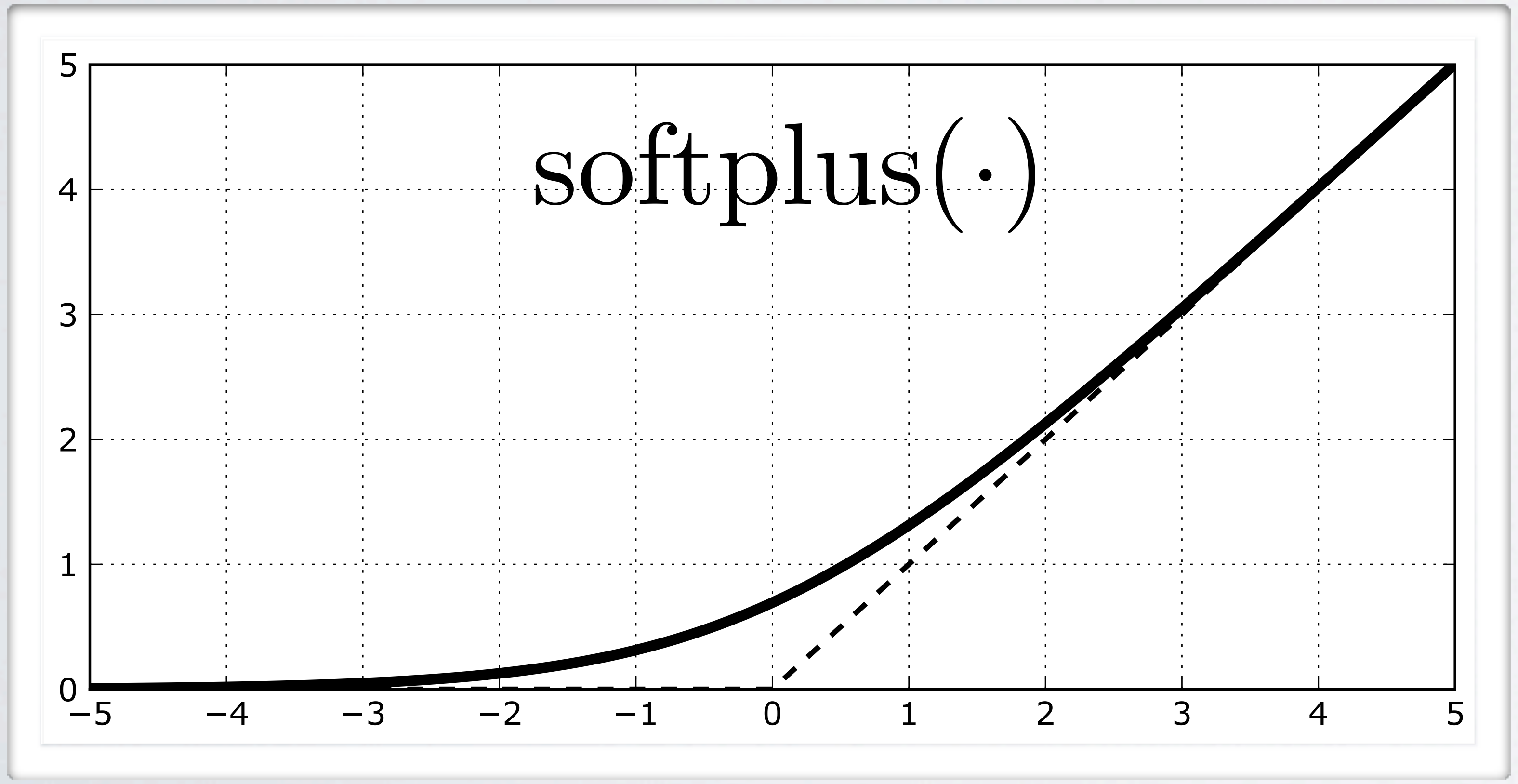
RESTRICTED BOLTZMANN MACHINE

Topics: free energy



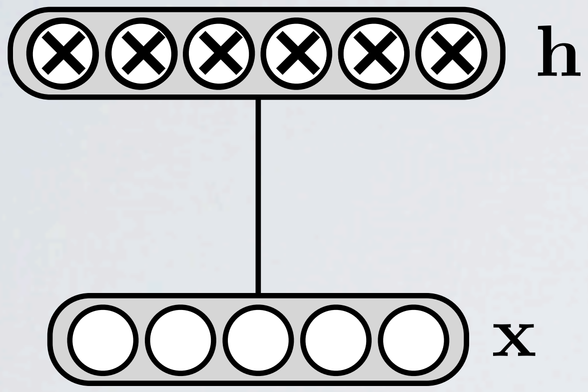
$$p(\mathbf{x}) = \exp\left(\mathbf{c}^\top\mathbf{x} + \sum_{j=1}^{H} \log(1+\exp(b_j + \mathbf{W}_{j\cdot}\mathbf{x}))\right)/Z$$

$$= \exp\left(\mathbf{c}^\top\mathbf{x} + \sum_{j=1}^{H} \mathrm{softplus}(b_j + \mathbf{W}_{j\cdot}\mathbf{x})\right)/Z$$



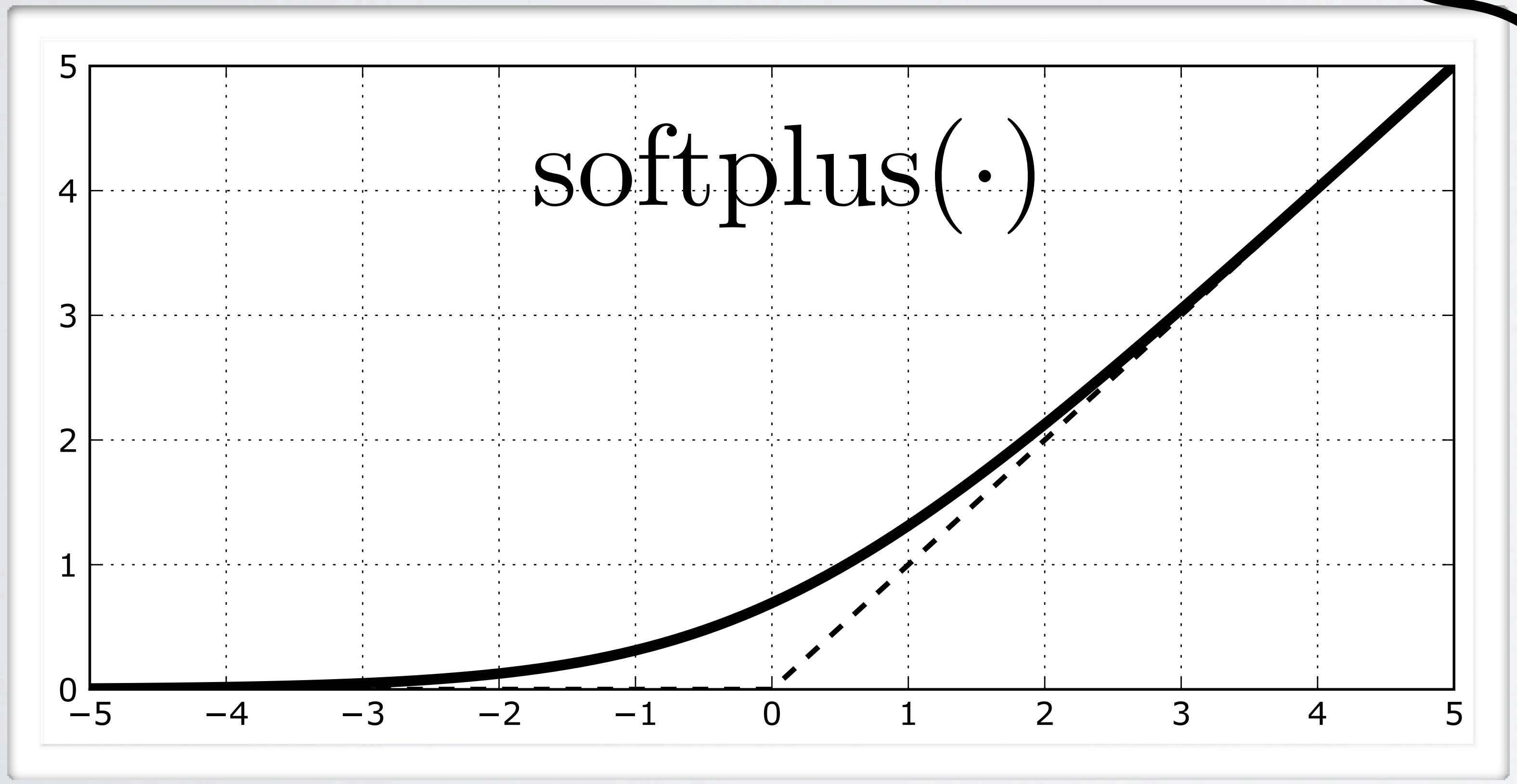
RESTRICTED BOLTZMANN MACHINE

Topics: free energy



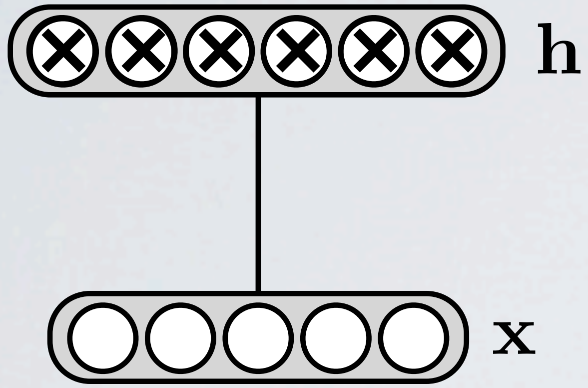
$$
\begin{aligned}
p(\mathbf{x}) &= \exp\left(\mathbf{c}^\top\mathbf{x} + \sum_{j=1}^{H} \log(1+\exp(b_j + \mathbf{W}_{j\cdot}\mathbf{x}))\right) / Z \\
&= \exp\left(\mathbf{c}^\top\mathbf{x} + \sum_{j=1}^{H} \mathrm{softplus}(b_j + \mathbf{W}_{j\cdot}\mathbf{x})\right) / Z
\end{aligned}
$$

bias the prob of each x_i



RESTRICTED BOLTZMANN MACHINE

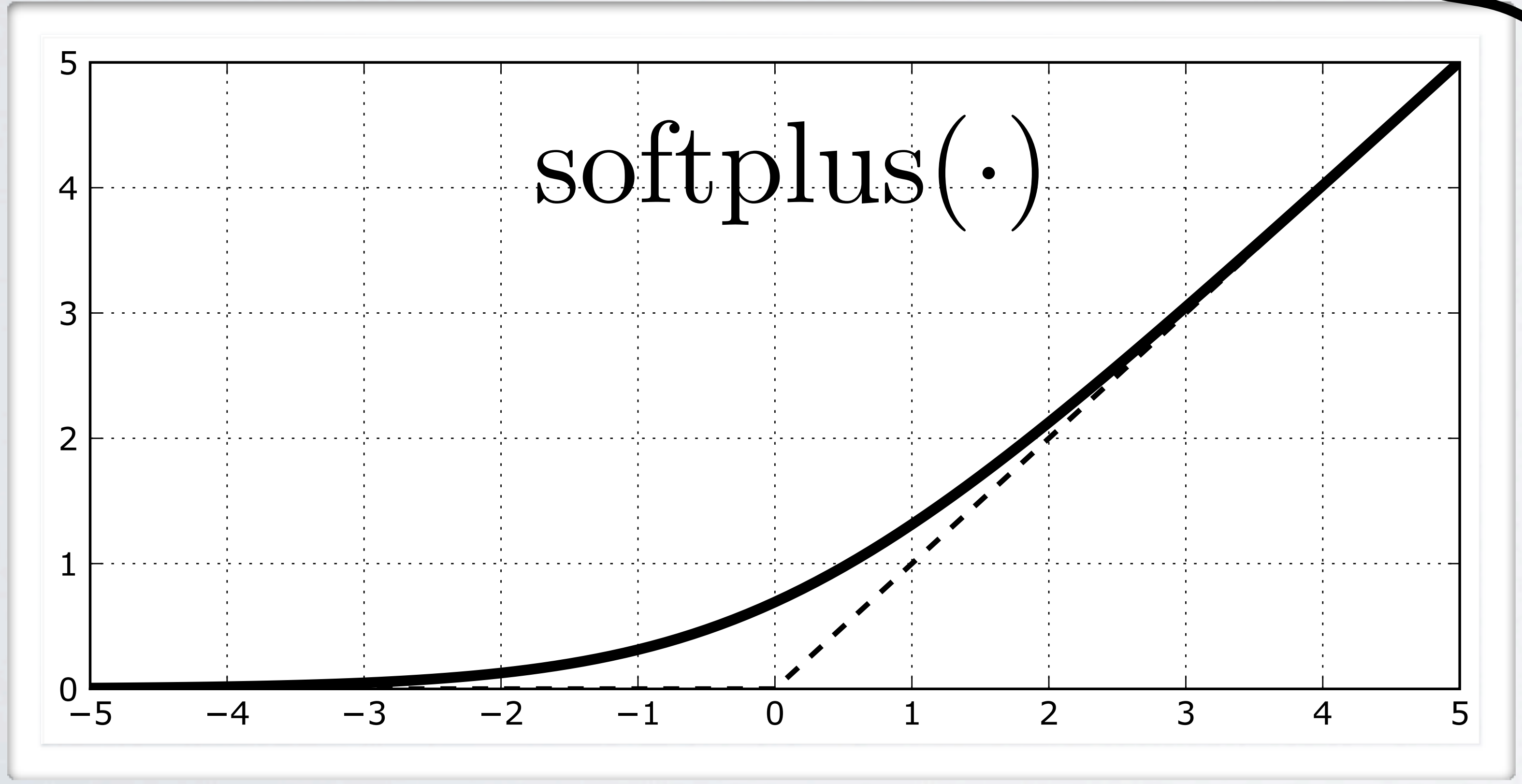
Topics: free energy



$$p(\mathbf{x}) = \exp\left(\mathbf{c}^\top \mathbf{x} + \sum_{j=1}^{H} \log(1 + \exp(b_j + \mathbf{W}_{j\cdot}\mathbf{x}))\right) / Z$$

$$= \exp\left(\mathbf{c}^\top \mathbf{x} + \sum_{j=1}^{H} \mathrm{softplus}(b_j + \mathbf{W}_{j\cdot}\mathbf{x})\right) / Z$$

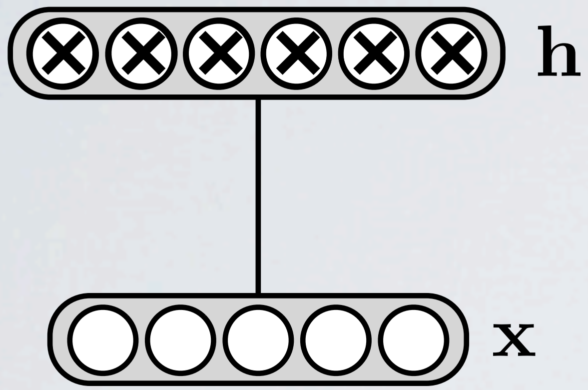
"feature" expected in $\mathbf{x}$



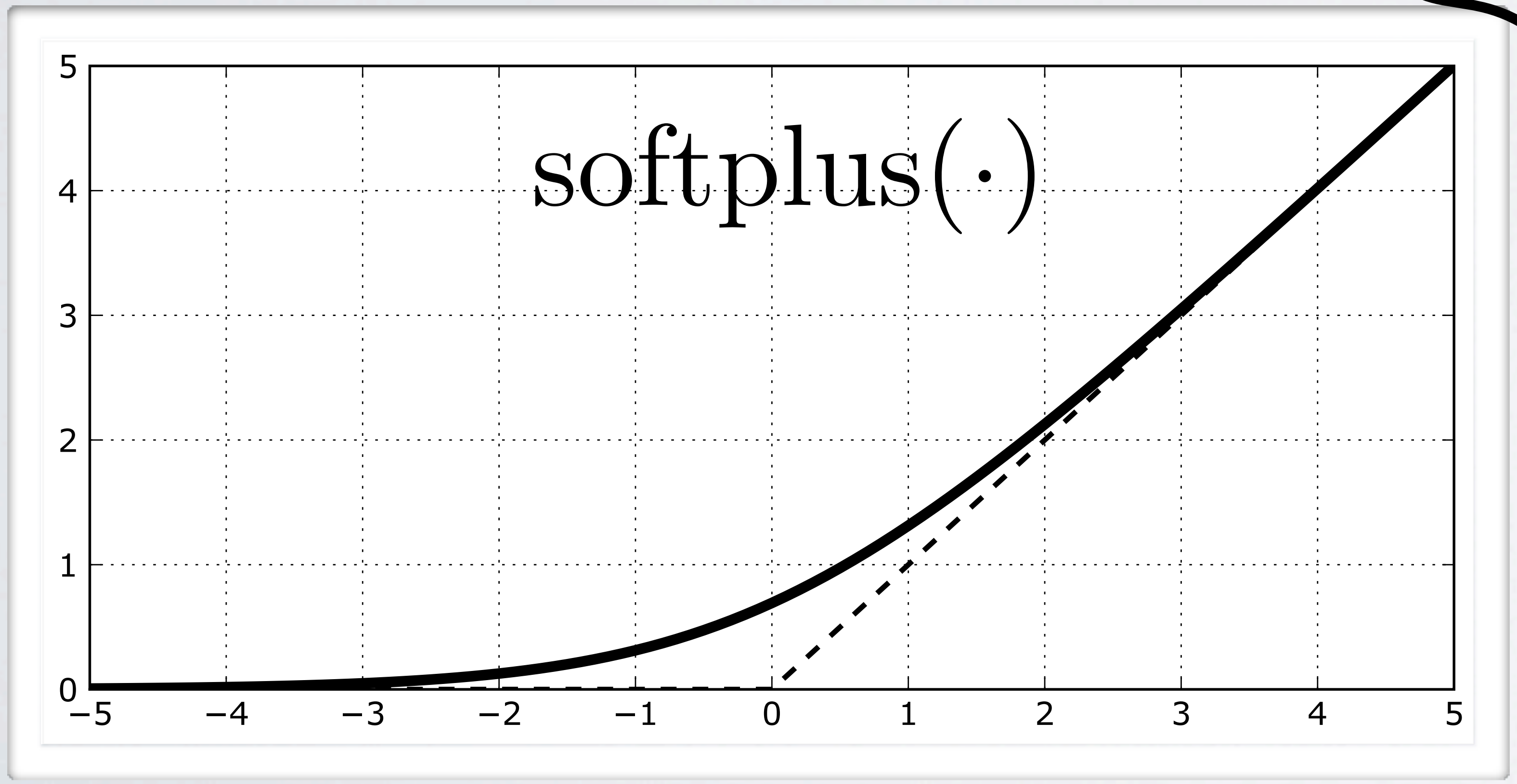
bias the prob of each x_i

RESTRICTED BOLTZMANN MACHINE

Topics: free energy



$$
\begin{aligned}
p(\mathbf{x}) &= \exp\left(\mathbf{c}^\top\mathbf{x} + \sum_{j=1}^{H}\log(1+\exp(b_j+\mathbf{W}_{j\cdot}\mathbf{x}))\right)/Z \\
&= \exp\left(\mathbf{c}^\top\mathbf{x} + \sum_{j=1}^{H}\mathrm{softplus}(b_j+\mathbf{W}_{j\cdot}\mathbf{x})\right)/Z
\end{aligned}
$$



"feature" expected in $\mathbf{x}$

bias of each feature

bias the prob of each x_i