

# Mining and Processing Biomedical Data

Dr. rer. nat. Krisztian Buza

adiunkt naukowy

Faculty of Mathematics, Informatics and Mechanics

University of Warsaw, Poland

[chrisbuza@yahoo.com](mailto:chrisbuza@yahoo.com)

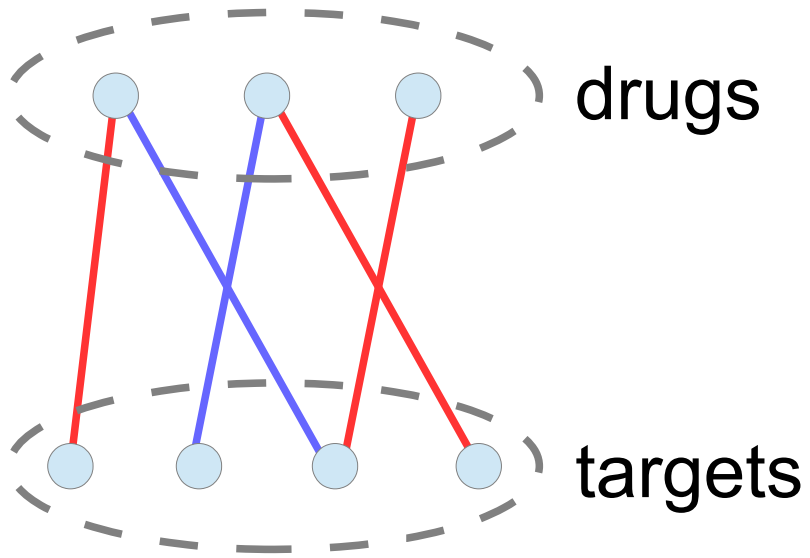
# RBM for drug-target prediction

# Predicting drug-target interactions

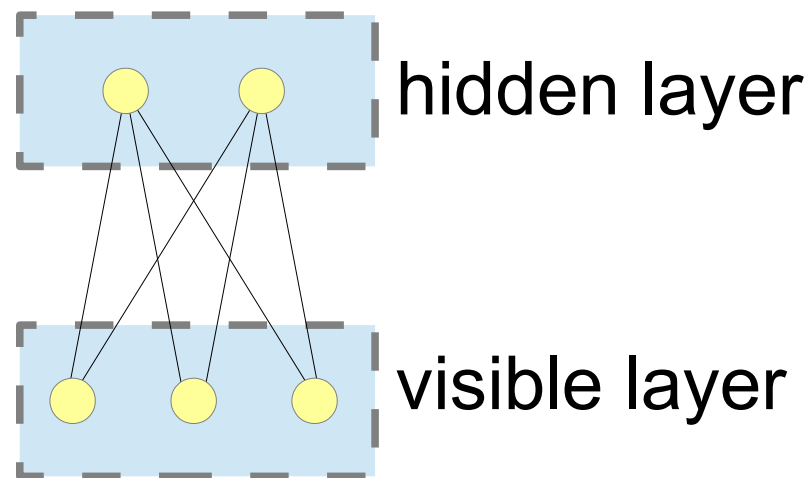
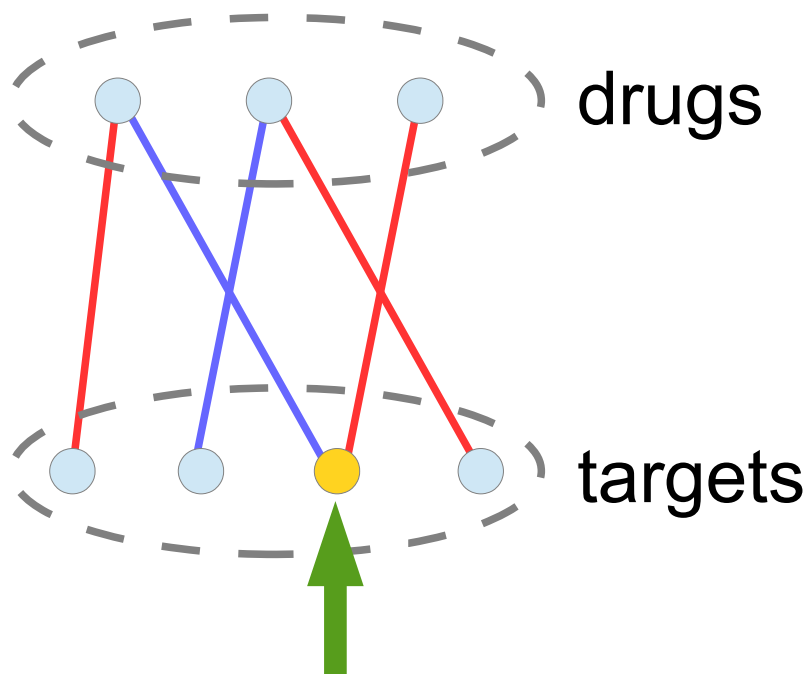
- Understanding how drugs work
- Re-use of existing drugs (for cases in which the particular drug has not been used before)
- “it typically takes 10-15 years and \$800 million – 1 billion to bring a new drug to market”
- “individual drugs can interact with multiple targets rather than a single target (...) serotonin and serotonergic drugs can interact with both 5-HT<sub>1A</sub> G protein-coupled receptors and 5-HT<sub>3A</sub> ion channel proteins, even though these two target proteins are not related in sequence or structure”

Y. Wang & J. Zeng: Predicting drug-target interactions using restricted Boltzmann machines, Bioinformatics, Vol. 29 ISMB/ECCB 2013, pages i126-i134

# Example: RBMs for drug-target prediction

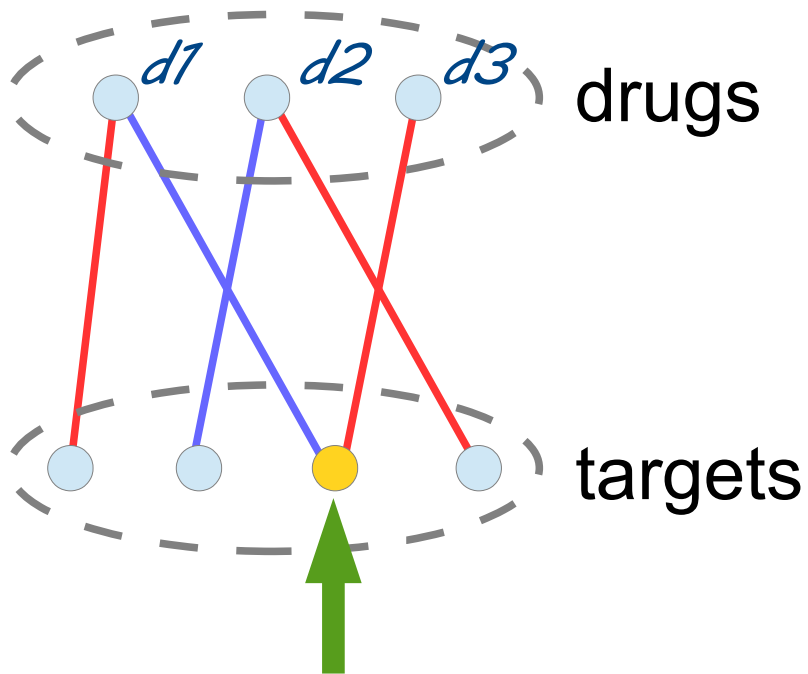


# Example: RBMs for drug-target prediction

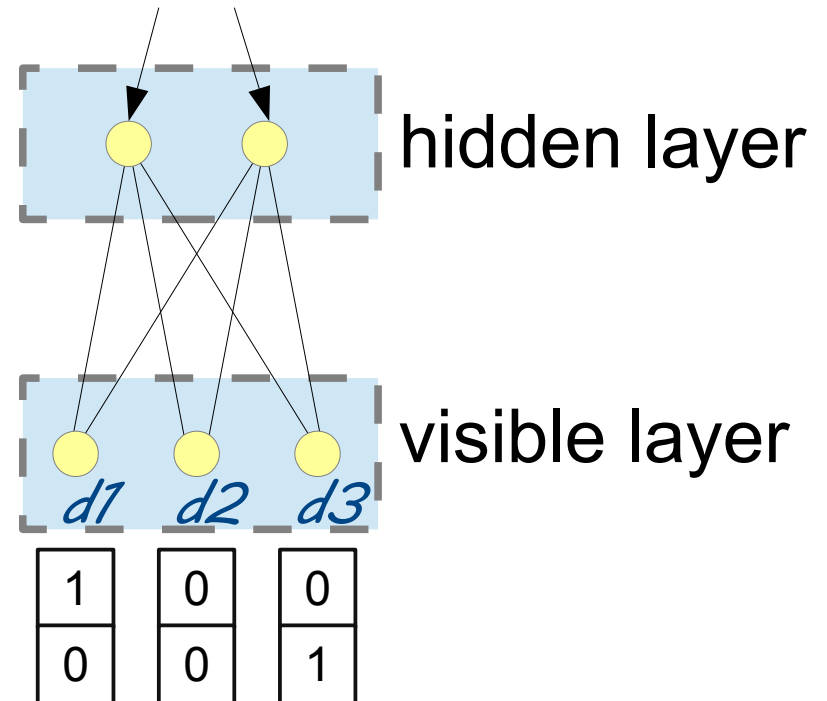


RBM for the highlighted target

# Example: RBMs for drug-target prediction

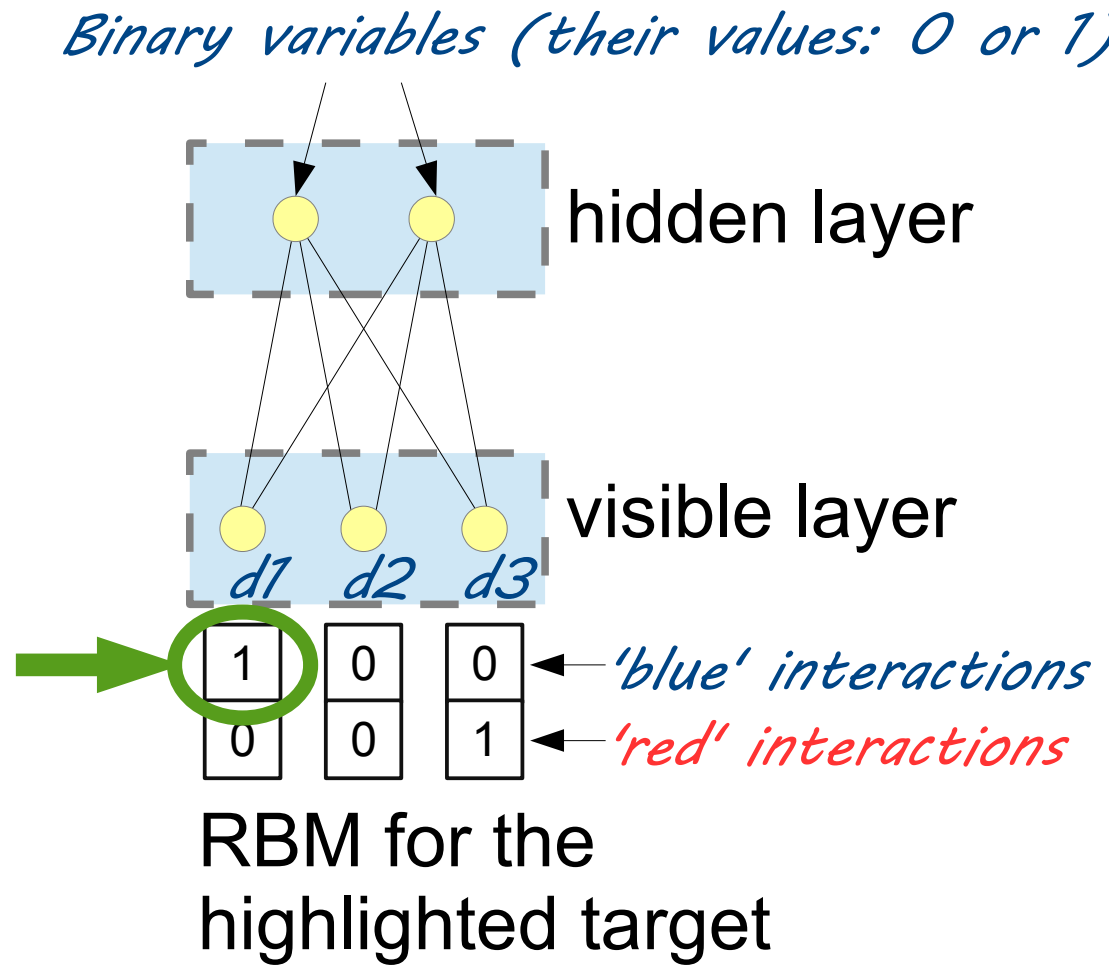
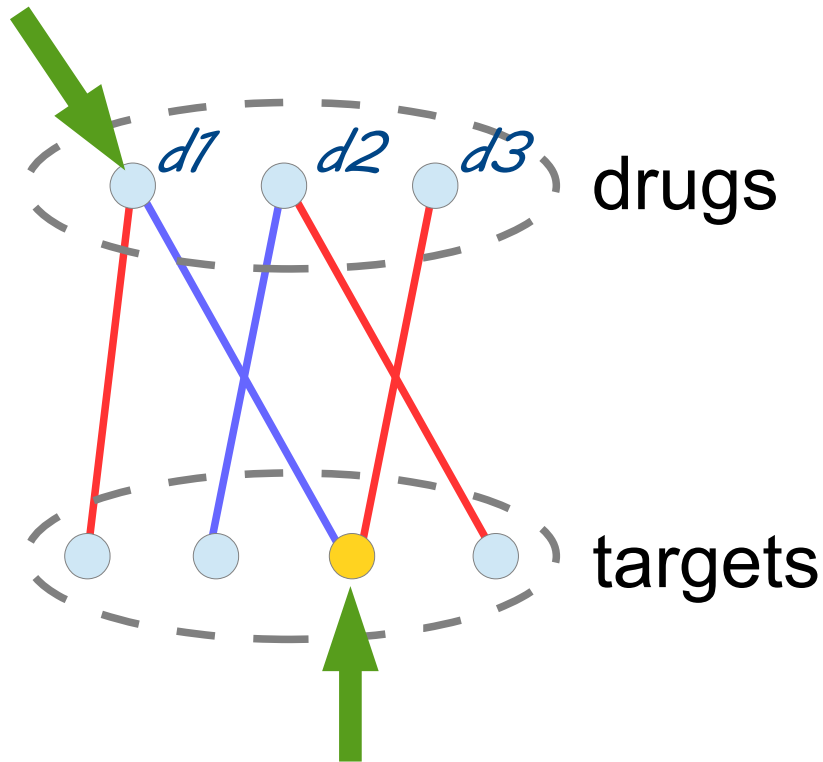


*Binary variables (their values: 0 or 1)*

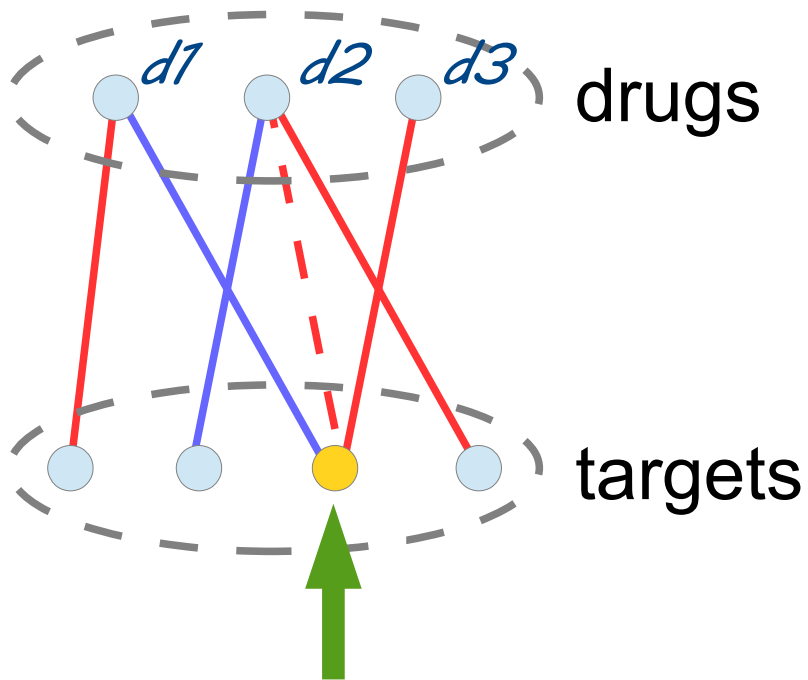


RBM for the highlighted target

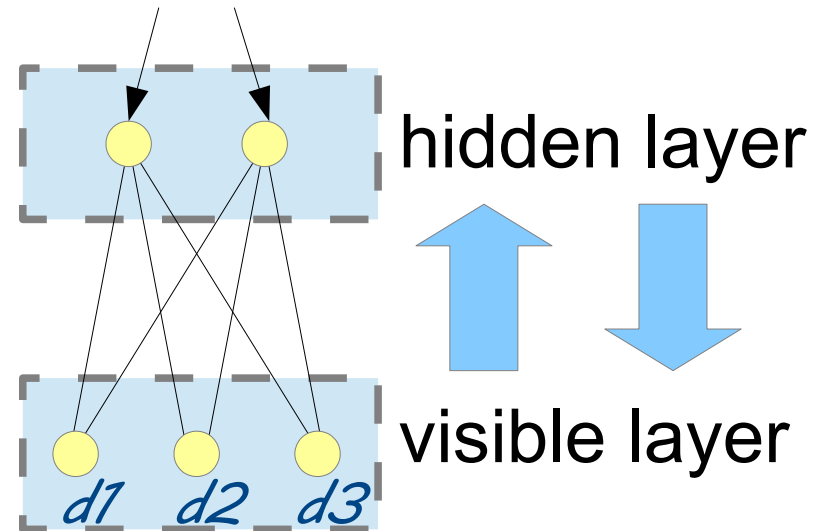
# Example: RBMs for drug-target prediction



# Example: RBMs for drug-target prediction



*Binary variables (their values: 0 or 1)*

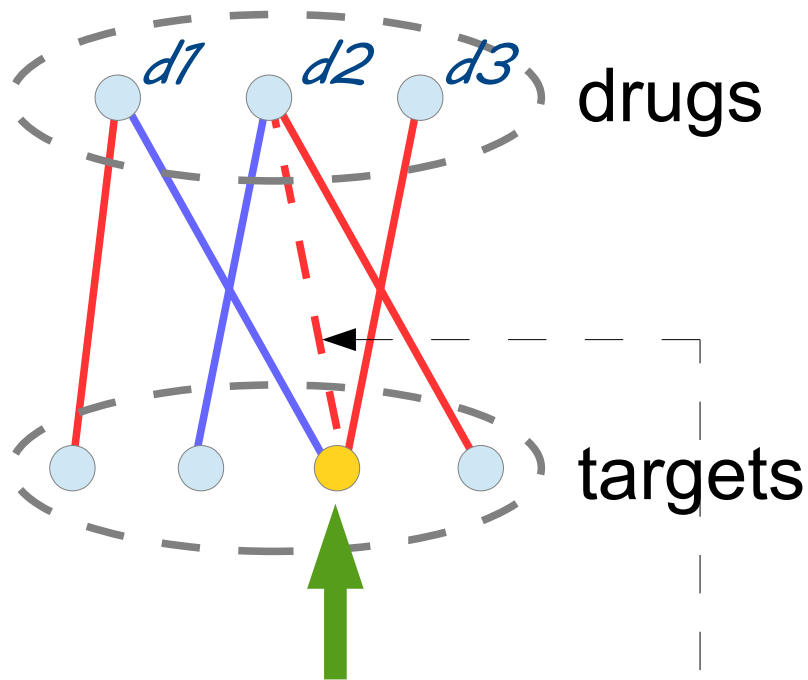


1	<del>1</del>	0	<del>0</del>	0	<del>0</del>	← 'blue' interactions
0	<del>0</del>	1	<del>0</del>	1	<del>1</del>	← 'red' interactions

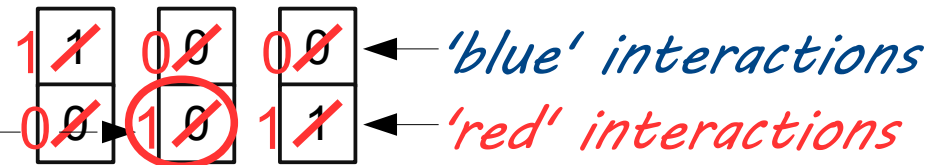
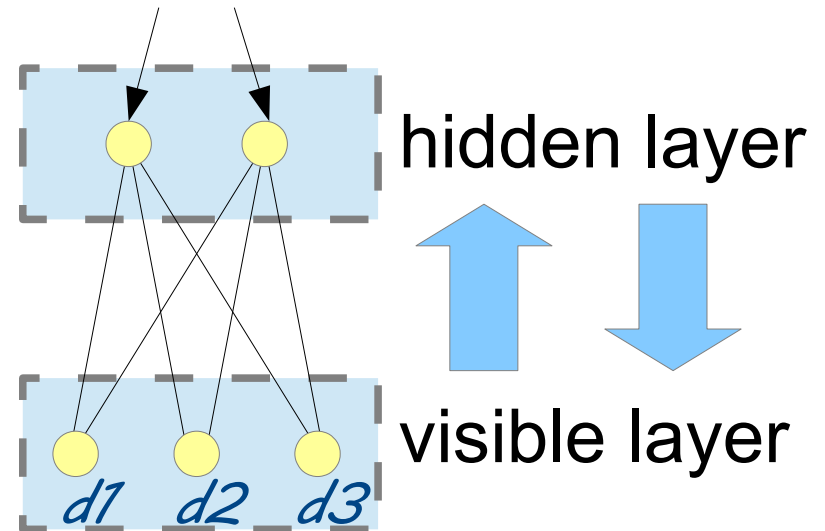
RBM for the highlighted target



# Example: RBMs for drug-target prediction



*Binary variables (their values: 0 or 1)*



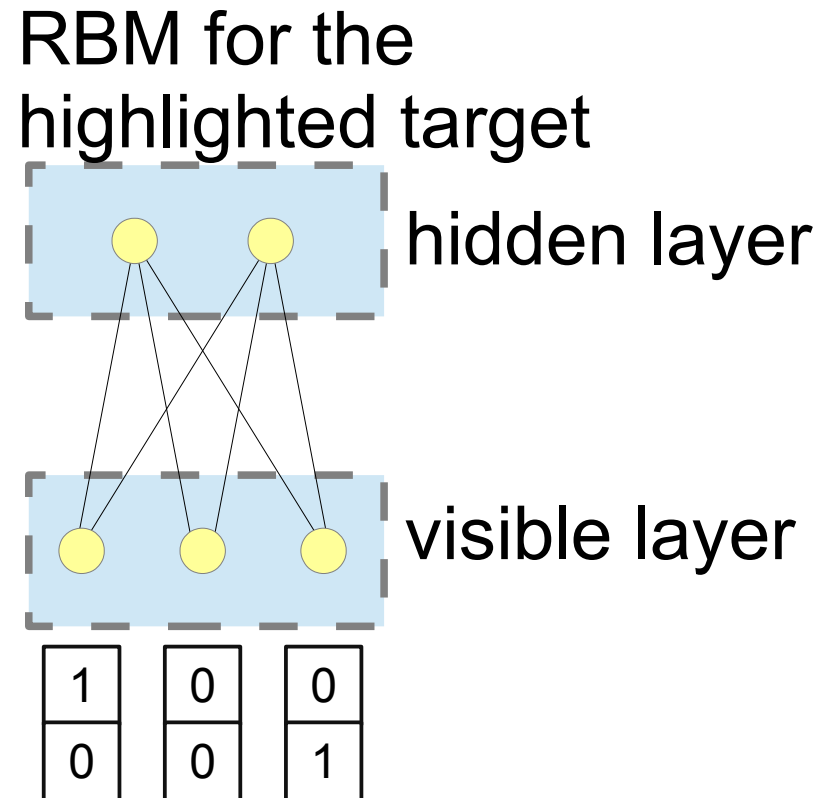
RBM for the highlighted target

*The RBM predicted a new 'red' interaction between d2 and the highlighted target*

# Example: RBMs for drug-target prediction

## Parameters of the RBM:

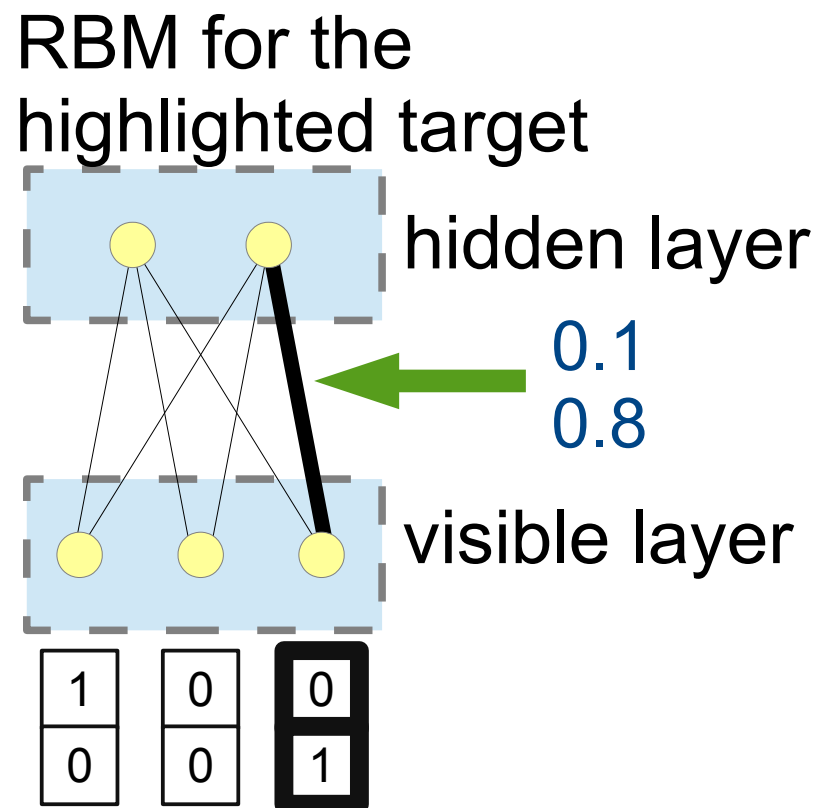
- connection weights  $W^k_{i,j}$
  - bias weights of the visible nodes  $a^k_i$
  - bias weights of the hidden nodes  $b_j$
- (indices:  $k$ -th component of the vector,  $i$ -th visible node,  $j$ -th hidden node)



# Example: RBMs for drug-target prediction

## Parameters of the RBM:

- connection weights  $W^k_{i,j}$
  - bias weights of the visible nodes  $a^k_i$
  - bias weights of the hidden nodes  $b_j$
- (indices:  $k$ -th component of the vector,  $i$ -th visible node,  $j$ -th hidden node)



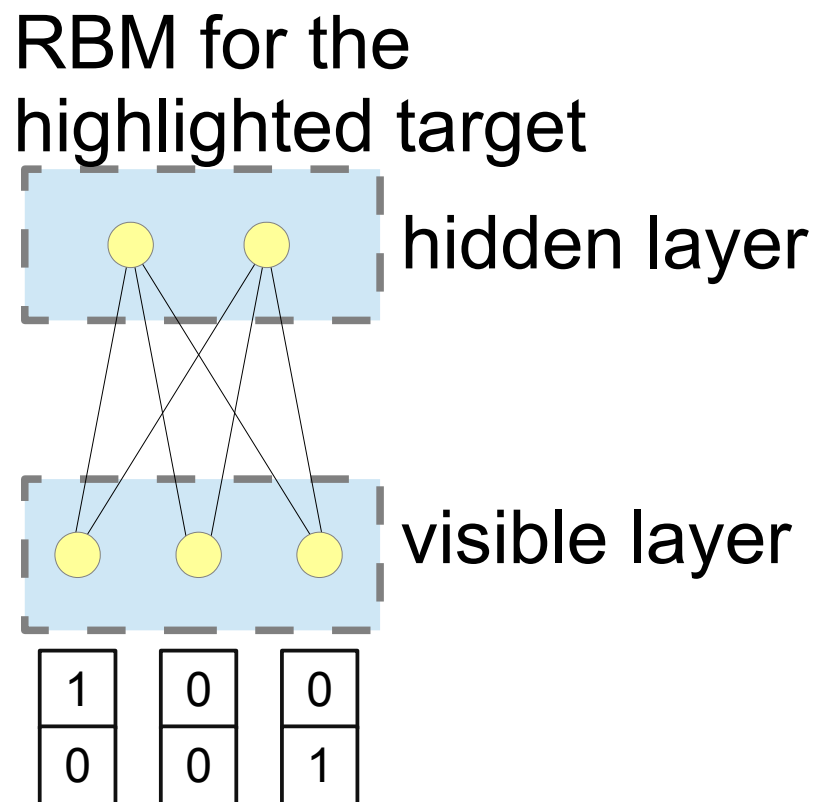
# Example: RBMs for drug-target prediction

## Parameters of the RBM:

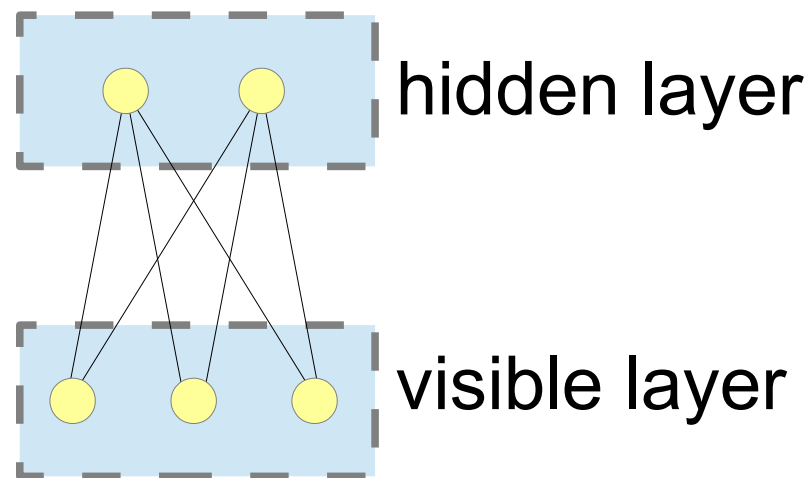
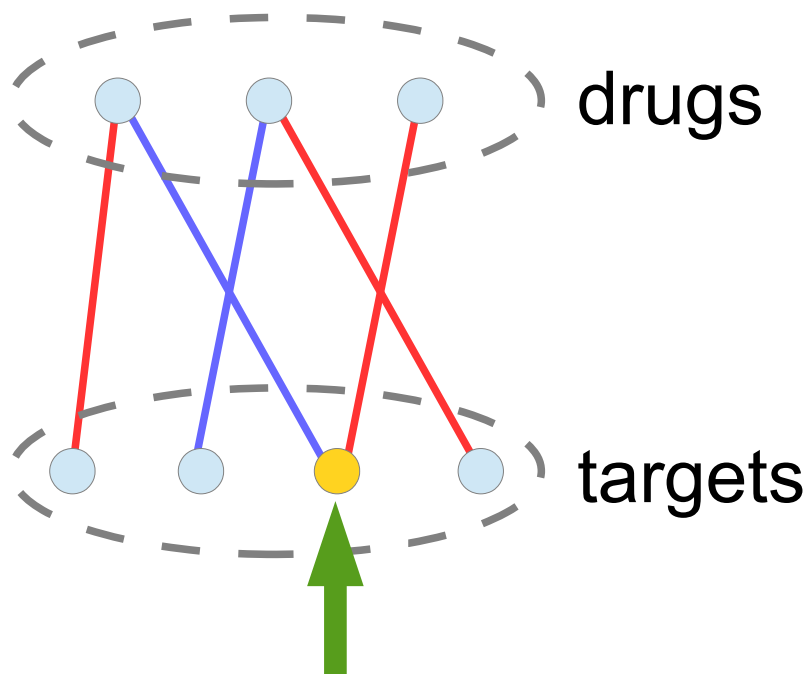
- connection weights  $W^k_{i,j}$
  - bias weights of the visible nodes  $a^k_i$
  - bias weights of the hidden nodes  $b_j$
- (indices:  $k$ -th component of the vector,  $i$ -th visible node,  $j$ -th hidden node)

Number of the parameters:  $n \cdot m \cdot t + n \cdot t + m$

where  $n = \text{\#drugs} = \text{\#visible nodes}$ ,  $m = \text{\#hidden nodes}$ ,  
 $t = \text{\#interaction types}$ , in the example:  $n=3$ ,  $m=2$ ,  $t=2$

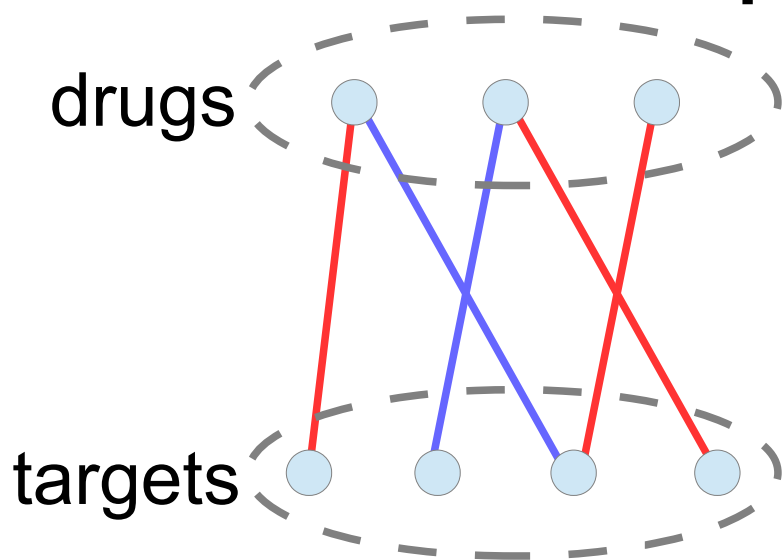


# Example: RBMs for drug-target prediction

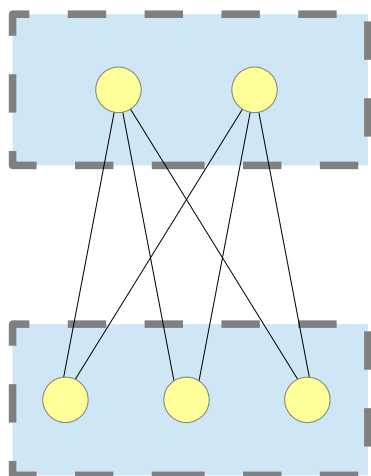


RBM for the highlighted target

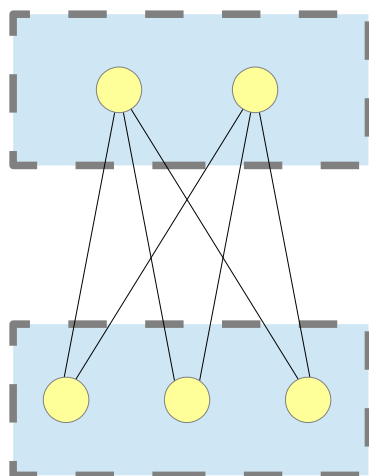
# Example: RBMs for drug-target prediction



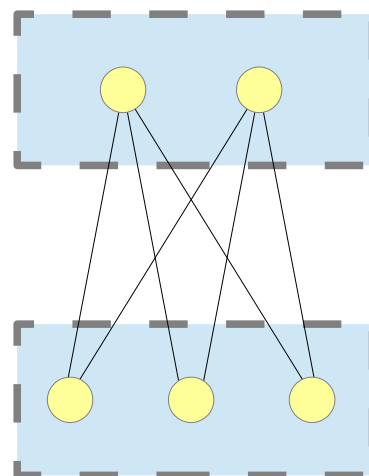
- An RBMs is constructed for each targets
- all of them have the same parameters of the RBM:



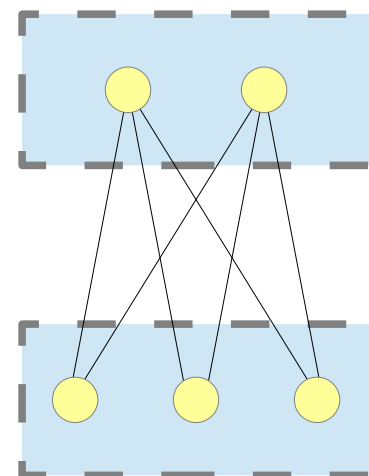
RBM for target 1



RBM for target 2

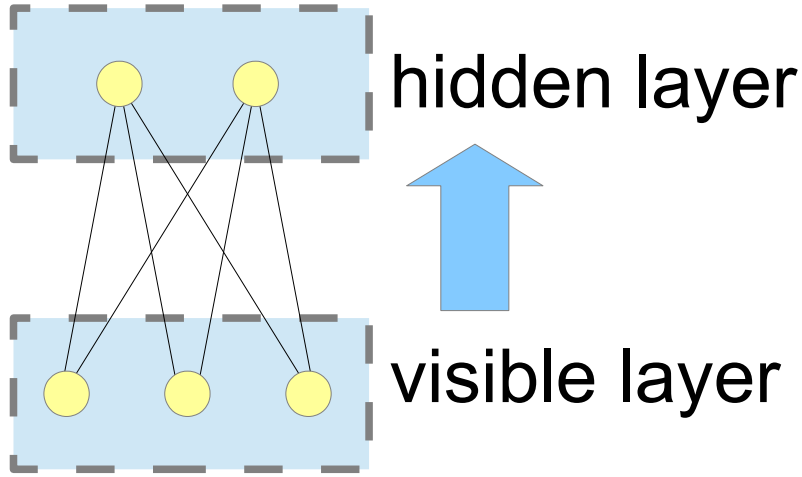


RBM for target 3



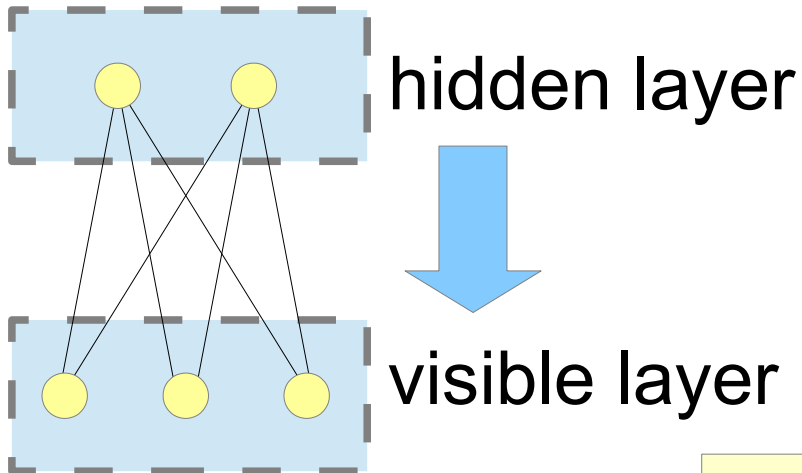
RBM for target 4

# Information flow



$$\Pr(h_j = 1 \mid \mathbf{v}) = \sigma\left(b_j + \sum_{i=1}^n \sum_{k=1}^t W_{ij}^k v_i^k\right)$$

*The value of the  $k$ -th component of the vector associated with the  $i$ -th node of the visible layer*

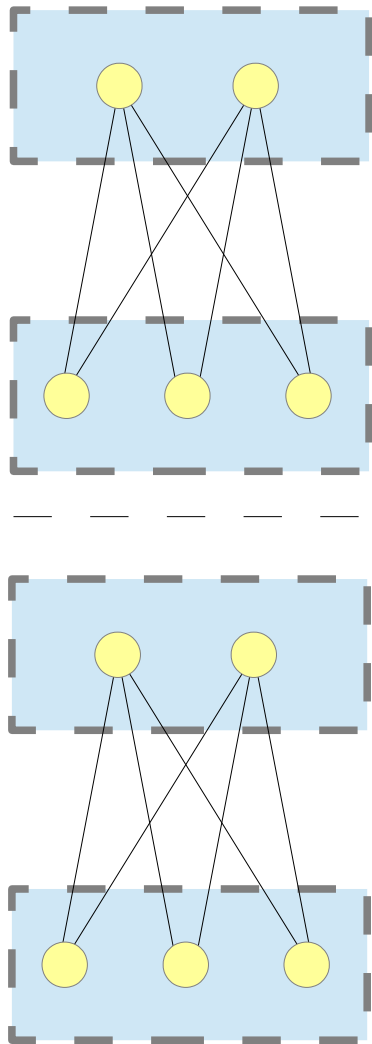


*The value of the  $j$ -th hidden node*

$$\Pr(v_i^k = 1 \mid \mathbf{h}) = \sigma\left(a_i^k + \sum_{j=1}^m W_{ij}^k h_j\right)$$

Y. Wang, J. Zeng: Predicting drug-target interactions using restricted Boltzmann machines, Bioinformatics, Vol. 29 ISMB/ECCB 2013, pages i126-i134

# Information flow



hidden layer

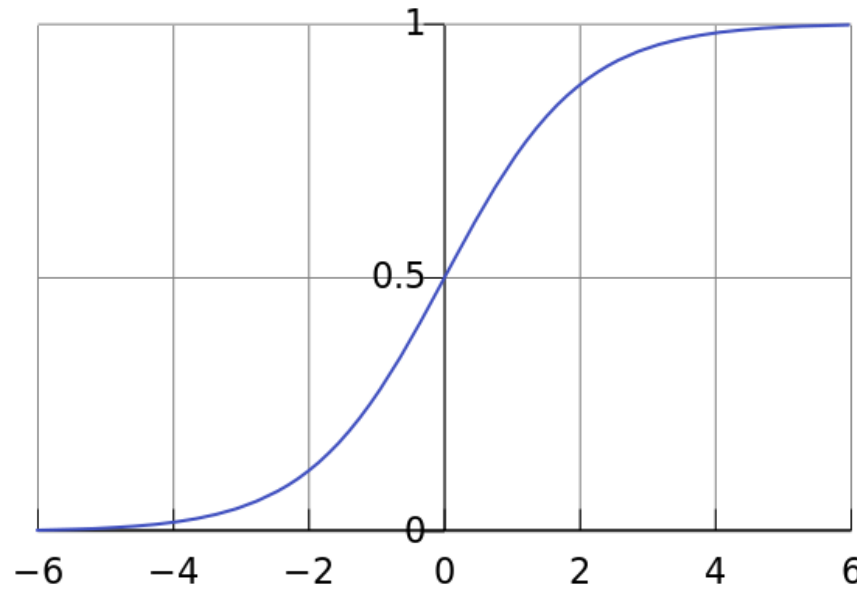
visible

hidden

visible

$$\Pr(h_j = 1 | \mathbf{v}) = \sigma\left(b_j + \sum_{i=1}^n \sum_{k=1}^t W_{ij}^k v_i^k\right)$$

Sigmoid function



[http://en.wikipedia.org/wiki/Sigmoid\\_function](http://en.wikipedia.org/wiki/Sigmoid_function)

each component  
is connected with the  
visible layer

each hidden node

$$+ \sum_{j=1}^m W_{ij}^k h_j$$

Wang, S. Zeng. Predicting drug-target interactions using restricted Boltzmann machines, Bioinformatics, Vol. 29 ISMB/ECCB 2013, pages i126-i134



# How to find the values of the parameters (weights, biases)?

- Contrastive Divergence

1. Initialize the weights and bias weights

2. **Repeat**

3.  $W^{k_{i,j}} \leftarrow W^{k_{i,j}} + \text{eps} * \Delta W^{k_{i,j}}$

4.  $a^{k_i} \leftarrow a^{k_i} + \text{eps} * \Delta a^{k_i}$

5.  $b_j \leftarrow b_j + \text{eps} * \Delta b_j$

6. **until** the weights are “good”

- How to find  $\Delta W^{k_{i,j}}$ ,  $\Delta a^{k_i}$  and  $\Delta b_j$  ?

# Calculation of $\Delta W^k_{i,j}$ , $\Delta a^k_j$ and $\Delta b_j$ by Contrastive Divergence

1. Let  $\Delta W^k_{i,j} = \Delta a^k_j = \Delta b_j = 0$  for all  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$
2. for each target
3.  $\mathbf{v}^{(0)} = \mathbf{v}$  (according to the current target) /\*  $\mathbf{v}=(v_1^1, \dots, v_i^k, \dots, v_n^t)$  \*/
4. for  $s = 0, \dots, r - 1$  **do**
5. for  $j = 1 \dots m$  **do** set  $h_j^{(s)}$  according to  $\Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$
6. for  $i = 1 \dots n$  **do** set  $v_i^{k,(s+1)}$  according to  $\Pr( v_i^k = 1 \mid \mathbf{h}^{(s)} )$
7. for  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$  **do**
8.  $\Delta W^k_{i,j} \leftarrow \Delta W^k_{i,j} + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) \cdot v_i^{k,(0)} - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} ) \cdot v_i^{k,(s)}$
9. for  $i=1..n$ ,  $k=1..t$  **do**
10.  $\Delta a^k_j \leftarrow \Delta a^k_j + v_i^{k,(0)} - v_i^{k,(s)}$
11. for  $j=1..m$  **do**
12.  $\Delta b_j \leftarrow \Delta b_j + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$

See also: Asja Fischer and Christian Igel. Training Restricted Boltzmann Machines: An Introduction, <http://image.diku.dk/igel/paper/TRBMAI.pdf>

# Calculation of $\Delta W^k_{i,j}$ , $\Delta a^k_j$ and $\Delta b_j$ by Contrastive Divergence

1. Let  $\Delta W^k_{i,j} = \Delta a^k_j = \Delta b_j = 0$  for all  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$
2. for each target
3.  $\mathbf{v}^{(0)} = \mathbf{v}$  (according to the current target) /\*  $\mathbf{v}=(v_1^1, \dots, v_i^k, \dots, v_n^t)$  \*/
4. for  $s = 0, \dots, r - 1$  **do**
5. for  $j = 1 \dots m$  **do** set  $h_j^{(s)}$  according to  $\Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$
6. for  $i = 1 \dots n$  **do** set  $v_i^{k,(s+1)}$  according to  $\Pr( v_i^k = 1 \mid \mathbf{h}^{(s)} )$
7. for  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$  **do**
8.  $\Delta W^k_{i,j} \leftarrow \Delta W^k_{i,j} + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) \cdot v_i^{k,(0)} - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} ) \cdot v_i^{k,(s)}$
9. for  $i=1..n$ ,  $k=1..t$  **do**
10.  $\Delta a^k_j \leftarrow \Delta a^k_j + v_i^{k,(0)} - v_i^{k,(s)}$
11. for  $j=1..m$  **do**
12.  $\Delta b_j \leftarrow \Delta b_j + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$

See also: Asja Fischer and Christian Igel. Training Restricted Boltzmann Machines: An Introduction, <http://image.diku.dk/igel/paper/TRBMAI.pdf>

# Calculation of $\Delta W^k_{i,j}$ , $\Delta a^k_j$ and $\Delta b_j$ by Contrastive Divergence

1. Let  $\Delta W^k_{i,j} = \Delta a^k_j = \Delta b_j = 0$  for all  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$
2. for each target
3.  $\mathbf{v}^{(0)} = \mathbf{v}$  (according to the current target) /\*  $\mathbf{v}=(v_1^1, \dots, v_i^k, \dots, v_n^t)$  \*/
4. for  $s = 0, \dots, r - 1$  **do**
5. for  $j = 1 \dots m$  **do** set  $h_j^{(s)}$  according to  $\Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$
6. for  $i = 1 \dots n$  **do** set  $v_i^{k,(s+1)}$  according to  $\Pr( v_i^k = 1 \mid \mathbf{h}^{(s)} )$
7. for  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$  **do**
8.  $\Delta W^k_{i,j} \leftarrow \Delta W^k_{i,j} + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) \cdot v_i^{k,(0)} - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} ) \cdot v_i^{k,(s)}$
9. for  $i=1..n$ ,  $k=1..t$  **do**
10.  $\Delta a^k_j \leftarrow \Delta a^k_j + v_i^{k,(0)} - v_i^{k,(s)}$
11. for  $j=1..m$  **do**
12.  $\Delta b_j \leftarrow \Delta b_j + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$

See also: Asja Fischer and Christian Igel. Training Restricted Boltzmann Machines: An Introduction, <http://image.diku.dk/igel/paper/TRBMAI.pdf>

# Calculation of $\Delta W_{i,j}^k$ , $\Delta a_j^k$ and $\Delta b_j$ by Contrastive Divergence

1. Let  $\Delta W_{i,j}^k = \Delta a_j^k = \Delta b_j = 0$  for all  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$
2. for each target
3.  $\mathbf{v}^{(0)} = \mathbf{v}$  (according to the current target) /\*  $\mathbf{v}=(v_1^1, \dots, v_i^k, \dots, v_n^t)$  \*/
4. for  $s = 0, \dots, r - 1$  **do**
5. for  $j = 1 \dots m$  **do** set  $h_j^{(s)}$  according to  $\Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$
6. for  $i = 1 \dots n$  **do** set  $v_i^{k,(s+1)}$  according to  $\Pr( v_i^k = 1 \mid \mathbf{h}^{(s)} )$
7. for  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$  **do**
8.  $\Delta W_{i,j}^k \leftarrow \Delta W_{i,j}^k + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) \cdot v_i^{k,(0)} - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} ) \cdot v_i^{k,(s)}$
9. for  $i=1..n$ ,  $k=1..t$  **do**
10.  $\Delta a_j^k \leftarrow \Delta a_j^k + v_i^{k,(0)} - v_i^{k,(s)}$
11. for  $j=1..m$  **do**
12.  $\Delta b_j \leftarrow \Delta b_j + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$

See also: Asja Fischer and Christian Igel. Training Restricted Boltzmann Machines: An Introduction, <http://image.diku.dk/igel/paper/TRBMAI.pdf>

# Calculation of $\Delta W^k_{i,j}$ , $\Delta a^k_j$ and $\Delta b_j$ by Contrastive Divergence

1. Let  $\Delta W^k_{i,j} = \Delta a^k_j = \Delta b_j = 0$  for all  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$
2. for each target
3.  $\mathbf{v}^{(0)} = \mathbf{v}$  (according to the current target) /\*  $\mathbf{v}=(v_1^1, \dots, v_i^k, \dots, v_n^t)$  \*/
4. for  $s = 0, \dots, r - 1$  **do**
5. for  $j = 1 \dots m$  **do** set  $h_j^{(s)}$  according to  $\Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$
6. for  $i = 1 \dots n$  **do** set  $v_i^{k,(s+1)}$  according to  $\Pr( v_i^k = 1 \mid \mathbf{h}^{(s)} )$
7. for  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$  **do**
8.  $\Delta W^k_{i,j} \leftarrow \Delta W^k_{i,j} + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) \cdot v_i^{k,(0)} - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} ) \cdot v_i^{k,(s)}$
9. for  $i=1..n$ ,  $k=1..t$  **do**
10.  $\Delta a^k_j \leftarrow \Delta a^k_j + v_i^{k,(0)} - v_i^{k,(s)}$
11. for  $j=1..m$  **do**
12.  $\Delta b_j \leftarrow \Delta b_j + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$

See also: Asja Fischer and Christian Igel. Training Restricted Boltzmann Machines: An Introduction, <http://image.diku.dk/igel/paper/TRBMAI.pdf>

# Calculation of $\Delta W^k_{i,j}$ , $\Delta a^k_j$ and $\Delta b_j$ by Contrastive Divergence

1. Let  $\Delta W^k_{i,j} = \Delta a^k_j = \Delta b_j = 0$  for all  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$
2. for each target
3.  $\mathbf{v}^{(0)} = \mathbf{v}$  (according to the current target)  $/* \mathbf{v}=(v_1^1, \dots, v_i^k, \dots, v_n^t) */$
4. for  $s = 0, \dots, r - 1$  **do**
5. for  $j = 1 \dots m$  **do** set  $h_j^{(s)}$  according to  $\Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$
6. for  $i = 1 \dots n$  **do** set  $v_i^{k,(s+1)}$  according to  $\Pr( v_i^k = 1 \mid \mathbf{h}^{(s)} )$
7. for  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$  **do**
8.  $\Delta W^k_{i,j} \leftarrow \Delta W^k_{i,j} + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) \cdot v_i^{k,(0)} - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} ) \cdot v_i^{k,(s)}$
9. for  $i=1..n$ ,  $k=1..t$  **do**
10.  $\Delta a^k_j \leftarrow \Delta a^k_j + v_i^{k,(0)} - v_i^{k,(s)}$
11. for  $j=1..m$  **do**
12.  $\Delta b_j \leftarrow \Delta b_j + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$

See also: Asja Fischer and Christian Igel. Training Restricted Boltzmann Machines: An Introduction, <http://image.diku.dk/igel/paper/TRBMAI.pdf>

# Calculation of $\Delta W^k_{i,j}$ , $\Delta a^k_j$ and $\Delta b_j$ by Contrastive Divergence

1. Let  $\Delta W^k_{i,j} = \Delta a^k_j = \Delta b_j = 0$  for all  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$
2. for each target
3.  $\mathbf{v}^{(0)} = \mathbf{v}$  (according to the current target) /\*  $\mathbf{v}=(v_1^1, \dots, v_i^k, \dots, v_n^t)$  \*/
4. for  $s = 0, \dots, r - 1$  **do**
5. for  $j = 1 \dots m$  **do** set  $h_j^{(s)}$  according to  $\Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$
6. for  $i = 1 \dots n$  **do** set  $v_i^{k,(s+1)}$  according to  $\Pr( v_i^k = 1 \mid \mathbf{h}^{(s)} )$
7. for  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$  **do**
8.  $\Delta W^k_{i,j} \leftarrow \Delta W^k_{i,j} + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) \cdot v_i^{k,(0)} - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} ) \cdot v_i^{k,(s)}$
9. for  $i=1..n$ ,  $k=1..t$  **do**
10.  $\Delta a^k_j \leftarrow \Delta a^k_j + v_i^{k,(0)} - v_i^{k,(s)}$
11. for  $j=1..m$  **do**
12.  $\Delta b_j \leftarrow \Delta b_j + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$

See also: Asja Fischer and Christian Igel. Training Restricted Boltzmann Machines: An Introduction, <http://image.diku.dk/igel/paper/TRBMAI.pdf>



# Calculation of $\Delta W_{i,j}^k$ , $\Delta a_j^k$ and $\Delta b_j$ by Contrastive Divergence

1. Let  $\Delta W_{i,j}^k = \Delta a_j^k = \Delta b_j = 0$  for all  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$
2. for each target
3.  $\mathbf{v}^{(0)} = \mathbf{v}$  (according to the current target) /\*  $\mathbf{v}=(v_1^1, \dots, v_i^k, \dots, v_n^t)$  \*/
4. for  $s = 0, \dots, r - 1$  **do**
5. for  $j = 1 \dots m$  **do** set  $h_j^{(s)}$  according to  $\Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$
6. for  $i = 1 \dots n$  **do** set  $v_i^{k,(s+1)}$  according to  $\Pr( v_i^k = 1 \mid \mathbf{h}^{(s)} )$
7. for  $i=1..n$ ,  $j=1..m$ ,  $k=1..t$  **do**
8.  $\Delta W_{i,j}^k \leftarrow \Delta W_{i,j}^k + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) \cdot v_i^{k,(0)} - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} ) \cdot v_i^{k,(s)}$
9. for  $i=1..n$ ,  $k=1..t$  **do**
10.  $\Delta a_j^k \leftarrow \Delta a_j^k + v_i^{k,(0)} - v_i^{k,(s)}$
11. for  $j=1..m$  **do**
12.  $\Delta b_j \leftarrow \Delta b_j + \Pr( h_j = 1 \mid \mathbf{v}^{(0)} ) - \Pr( h_j = 1 \mid \mathbf{v}^{(s)} )$

See also: Asja Fischer and Christian Igel. Training Restricted Boltzmann Machines: An Introduction, <http://image.diku.dk/igel/paper/TRBMAI.pdf>

# Predictions

*Information flow: visible  $\rightarrow$  hidden layer:*

$$\hat{p}_j = \Pr(h_j | \mathbf{v}) = \sigma\left(b_j + \sum_{i=1}^n \sum_{k=1}^l W_{ij}^k v_i^k\right)$$

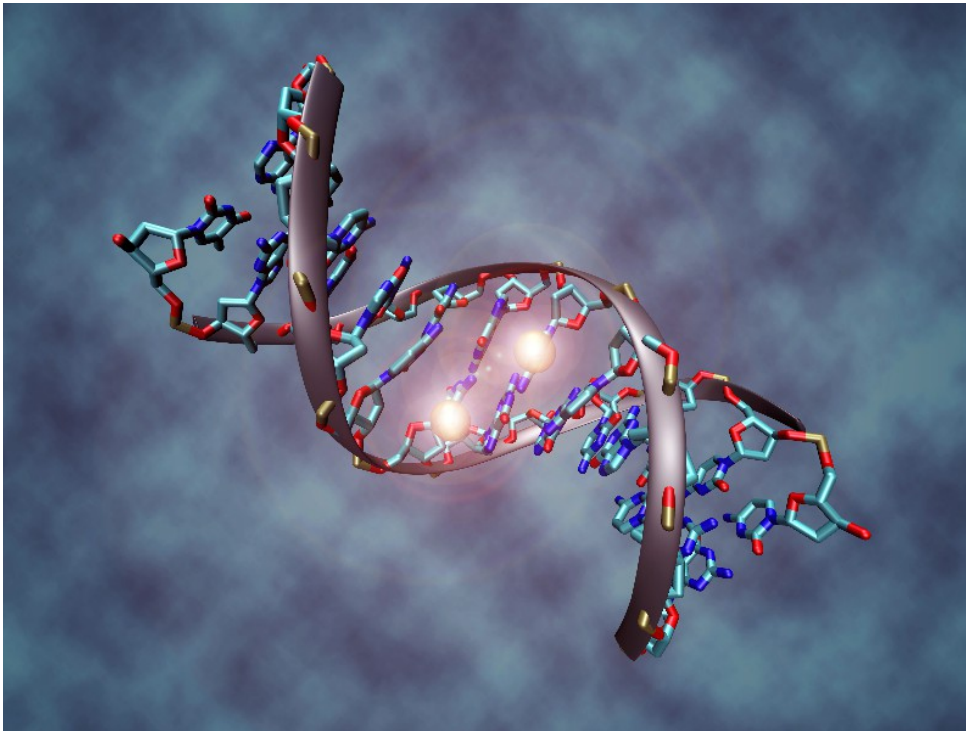
*Information flow: hidden  $\rightarrow$  visible layer:*

$$\Pr(v_i^k = 1 | \hat{p}_1, \dots, \hat{p}_m) = \sigma\left(a_i^k + \sum_{j=1}^m W_{ij}^k \hat{p}_j\right)$$

*“Probability” of an interaction*

# Matrix factorisation for the analysis of DNA methylation

# DNA methylation



“DNA methylation is a biochemical process whereby a methyl group is added to the cytosine or adenine DNA nucleotides. DNA methylation stably alters the expression of genes in cells.”

[http://en.wikipedia.org/wiki/DNA\\_methylation](http://en.wikipedia.org/wiki/DNA_methylation)

Image from Wikipedia. Licence info:  
[http://en.wikipedia.org/wiki/File:DNA\\_methylation.jpg](http://en.wikipedia.org/wiki/File:DNA_methylation.jpg)

# Matrix factorization for the analysis of DNA methylation in cancer

See also *Matrix factorization with binary components*, by M. Slawski et al., NIPS 2013

- A site is either methylated or not (0 or 1)
- Samples: mixture of different types of cells (e.g. healthy cells and cells affected by cancer)

A	B	C
0.21	0.51	0.13
0.17	0.25	0.12
0.34	0.45	0.87
0.51	0.41	0.32
0.23	0.12	0.35
...	...	...

≈

Type1	Type2
0	0
0	1
1	0
1	0
0	1
...	...

	A	B	C
x	0.23	0.7	0.3
	0.77	0.3	0.7

Proportion of cell types in the samples

Measurements: Methylation levels in different samples (columns) and DNA sites (rows)

Methylation in the cell types at different sites