DBM

Лобачева Екатерина

2 мая 2012

Постановка задачи

AB_2X_2

- <u>Вход:</u>

  A, B, X – химические элементы

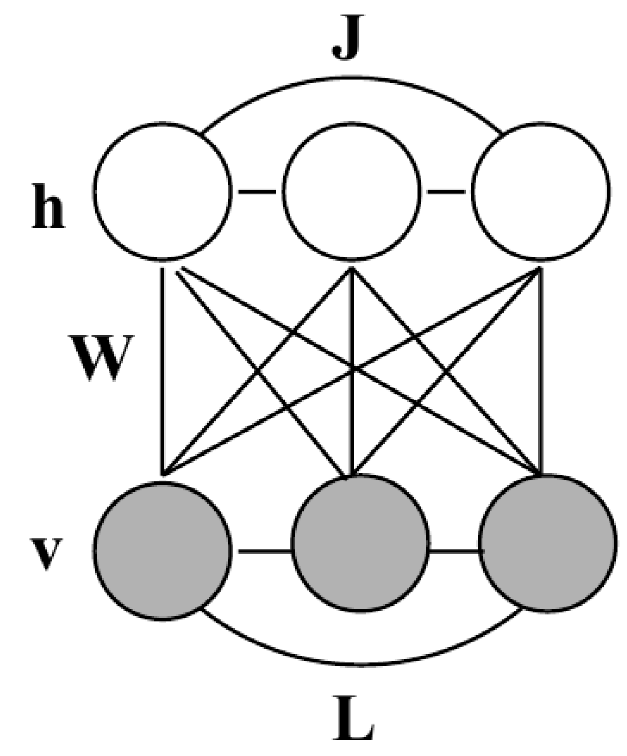
$$F = \left\{ \underbrace{f_1, \ldots, f_{32}}_{A}, \underbrace{f_{33}, \ldots, f_{64}}_{B}, \underbrace{f_{65}, \ldots, f_{96}}_{X} \right\}$$

- <u>Выход:</u> 7 классов
  - 1-5 – определенные структуры
  - 6 – другая структура
  - 7 – не существует

Особенности задачи

- Многоклассовая классификация.
- Плохое признаковое пространство.
- Размеченная выборка ≈ 2000,
  Неразмеченная ≈ 10000.

Модель ВМ

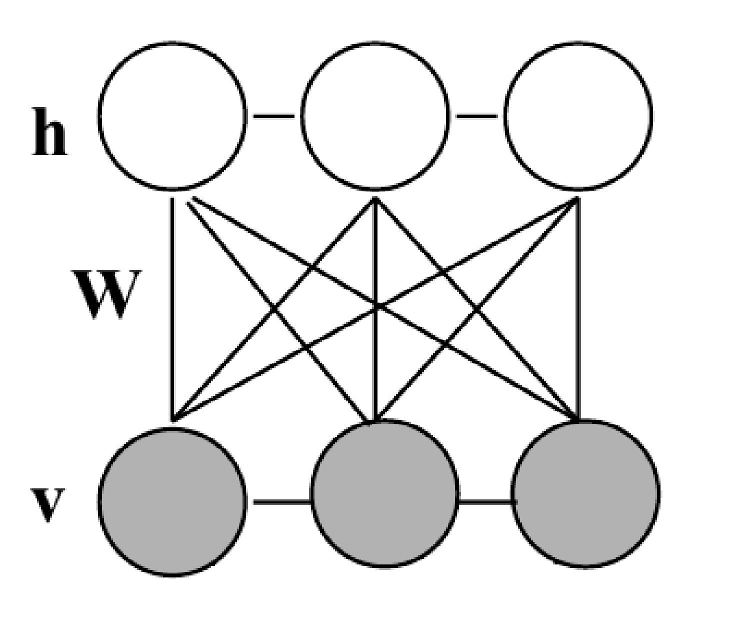


Параметры:

$$\theta = \{W, L, J, b^v, b^h\}$$

$$E(v, h|\theta) = -v^T W h - v^T L v - h^T J h - v^T b^v - h^T b^h$$

Модель RBM



Параметры:

$$\theta = \{W, b^v, b^h\}$$

$$E(v, h|\theta) = -v^T W h - v^T b^v - h^T b^h$$

$$p(v, h|\theta) = \frac{1}{Z(\theta)} \exp\left(-E(v, h|\theta)\right)$$

Вывод в RBM

- Вход:

  Параметры модели – W.

- Выход:

  Выборка из $p(v, h, W)$.

$$RBM \Rightarrow \begin{cases} p(v_i = 1 | h, W) = \dfrac{1}{1 + \exp\left(-\sum_j W_{ij} h_j\right)} \\ p(h_j = 1 | v, W) = \dfrac{1}{1 + \exp\left(-\sum_i W_{ij} v_i\right)} \end{cases}$$

Схема Гиббса

- Вход: W, начальное приближение v^0.
- Выход: Выборка $\{v, h\}$ из $p(v, h|W)$.

1. $h^0 \sim p(h|v^0, W)$;
2. для iter = 1 : #iter
3. $\quad v^{iter} \sim p(v|h^{iter-1}, W)$;
4. $\quad h^{iter} \sim p(h|v^{iter}, W)$;
5. $v = v^{\#iter}$, $h = h^{\#iter}$.

Обучение RBM

- Вспомним:

$$E(v,h|W) = -v^T W h \qquad p(v,h|W) = \frac{1}{Z(W)} \exp\left(-E(v,h|W)\right)$$

$$p(v|W) = \frac{1}{Z(W)} \sum_h \exp\left(-E(v,h|W)\right)$$

- Вход: $\hat{V} = \left\{ \hat{v}^1, \ldots, \hat{v}^N \right\}$

- Задача: $p(\hat{V}|W) \to \max_W$

Экспоненциальное семейство распределений

- Общий вид:

$$p(x|\theta)=\frac{1}{Z(\theta)}f(x)\exp\left(\theta^T u(x)\right)$$

- Фишка:

$$\left(\log Z(\theta)\right)' = Eu(x)$$

$$\Downarrow$$

$$\nabla_\theta \log p(\hat{x}|\theta)=u(\hat{x})-Eu(x)$$

Экспоненциальное семейство распределений

Для выборки $\hat{X} = \left\{\hat{x}^1, \dots, \hat{x}^N\right\}$

$$\frac{1}{N}\nabla_\theta \log p\left(\hat{X}\middle|\theta\right) = E_{data}u(x) - E_{\mathrm{mod}el}u(x),$$

где

$$p_{data}(x) = \frac{1}{N}\sum_{i=1}^{N}\delta\left(x - \hat{x}_i\right)$$

$$p_{\mathrm{mod}el}(x) = p\left(x\middle|\theta\right)$$

Экспоненциальное семейство распределений

Аналогично для наблюдаемых x и ненаблюдаемых t:

$$p(x,t|\theta)=\frac{1}{Z(\theta)}f(x,t)\exp\left(\theta^T u(x,t)\right)$$

$$\frac{1}{N}\nabla_\theta \log p\left(\hat{X}|\theta\right)=E_{data}u(x,t)-E_{\mathrm{model}}u(x,t) \ ,$$

где $p_{data}(x,t)=\frac{1}{N}\sum_{i=1}^{N}p(t|\hat{x}_i,\theta)\delta(x-\hat{x}_i),\ \ p_{\mathrm{model}}(x,t)=p(x,t|\theta)$.

Обучение RBM

$p(v,h|W)$ из экспоненциального семейства

$$\Downarrow$$

$$\frac{1}{N}\nabla_W \log p(\hat{V}|W) = E_{data}(vh^T) - E_{\mathrm{mod}el}(vh^T)$$

причем

$$E_{data}(vh^T) = \sum_{n=1}^{N} \hat{v}^n h^{n^T}$$

$$\{v_{\mathrm{mod}el}, h_{\mathrm{mod}el}\}: \quad E_{\mathrm{mod}el}(vh^T) = \sum_{m=1}^{M} v_{\mathrm{mod}el}^m h_{\mathrm{mod}el}^{m}{}^T$$

Алгоритм обучения RBM

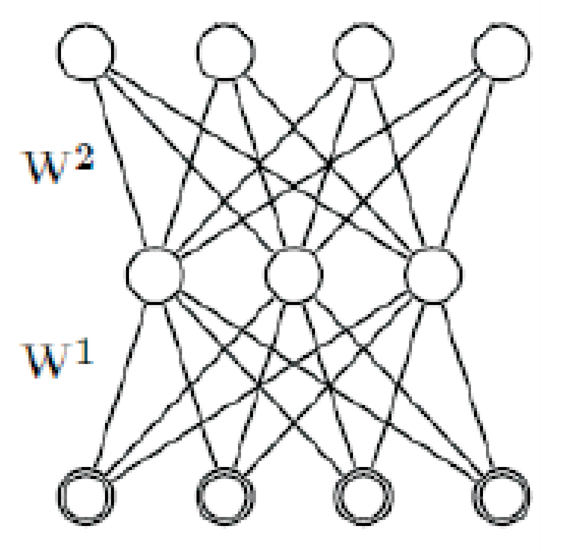
- Вход: $\hat{V} = \{\hat{v}^1, \ldots, \hat{v}^N\}$
- Выход: W

1. Инициализация W_0;
2. для iter = 1 : #iter
3. для n = 1 : N
4. $h^n \sim p\left(h \middle| \hat{v}^n, W_{iter-1}\right);$
5. $v^n_{\text{model}} \sim p\left(v \middle| h^n, W_{iter-1}\right),\quad h^n_{\text{model}} \sim p\left(h \middle| v^n_{\text{model}}, W_{iter-1}\right);$
6. $E_{data}\left(vh^T\right) = \sum_{n=1}^{N} \hat{v}^n h^{nT};$
7. $E_{\text{model}}\left(vh^T\right) = \sum_{n=1}^{N} v^n_{\text{model}} h^{n\,T}_{\text{model}};$
8. $W_{iter} = W_{iter-1} + \alpha_{iter}\left(E_{data}\left(vh^T\right) - E_{\text{model}}\left(vh^T\right)\right).$

Замечания: RBM

- Увеличение числа шагов по схеме Гиббса при генерации $\{v_{\mathrm{mod}el}, h_{\mathrm{mod}el}\}$ улучшает качество, но увеличивает время. Поэтому лучше увеличивать их число по мере приближения к оптимуму.
- Для оценки статистик лучше брать вероятности $h^n, v_{\mathrm{mod}el}, h_{\mathrm{mod}el}$.
- Чувствителен к подбору параметров обучения $(W_0, \alpha_{iter}, \ldots)$.

Модель DBM



Параметры:

$$\theta = \{W^1, W^2\}$$

$$E(v, h^1, h^2 | \theta) = -v^T W^1 h^1 - {h^1}^T W^2 h^2$$

$$p(v, h^1, h^2 | \theta) = \frac{1}{Z(\theta)} \exp\left(-E(v, h^1, h^2 | \theta)\right)$$

Обучение DBM

- Задача: $p(\hat{V}|\theta) \to \max_{\theta}$
- Аналогично:

$p(v, h^1, h^2 | \theta)$ из экспоненциального семейства

$$\Downarrow$$

$$\frac{1}{N} \nabla_{\theta} \log p(\hat{V}|\theta) = E_{data} u(v, h^1, h^2) - E_{\mathrm{model}} u(v, h^1, h^2)$$

$$u(v, h^1, h^2) = \begin{pmatrix} v h^{1^T} & h^1 h^{2^T} \end{pmatrix}$$

$$p_{data}(v, h^1, h^2) = \frac{1}{N} \sum_{i=1}^{N} p(h^1, h^2 | \hat{v}_i, \theta) \delta(v - \hat{v}_i),$$

$$p_{\mathrm{model}}(v, h^1, h^2) = p(v, h^1, h^2 | \theta).$$

Оценка E_{data}

$p(h^1, h^2 | \hat{v}, \theta)$: Схема Гиббса $\longrightarrow$ Вариационный подход

$$\log p(\hat{v}|\theta) = \underbrace{\int \log \frac{p(\hat{v}, h^1, h^2 | \theta)}{q(h^1, h^2 | \mu^1, \mu^2)} q(h^1, h^2 | \mu^1, \mu^2) dh^1 dh^2}_{L(q)} + KL(q \| p(h^1, h^2 | \hat{v}, \theta))$$

Факторизованное семейство:

$$q(h^1, h^2 | \mu^1, \mu^2) = q(h^1 | \mu^1) q(h^2 | \mu^2) = \prod_j q(h_j^1 | \mu_j^1) \prod_k q(h_k^2 | \mu_k^2)$$

Оценка E$_{data}$

$$KL\left(q \,\|\, p\left(h^1, h^2 \mid \hat{v}, \theta\right)\right) > 0 \quad \Longrightarrow \quad L(q) \to \max_{\mu^1, \mu^2}$$

$$\mu_j^{1^n} = \frac{1}{1 + \exp\left(-\sum_i W_{ij}^1 v_j^n - \sum_k W_{jk}^2 \mu_k^{2^n}\right)}$$

$$\mu_k^{2^n} = \frac{1}{1 + \exp\left(-\sum_j W_{jk}^2 \mu_j^{1^n}\right)}$$

$$E_{data}\left(v h^{1^T}\right) = \sum_{n=1}^{N} \hat{v}^n \mu^{1^n T} \qquad E_{data}\left(h^1 h^{2^T}\right) = \sum_{n=1}^{N} \mu^{1^n} \mu^{2^n T}$$

Оценка E_{model}

$p\left(v, h^1, h^2 \middle| \theta\right)$: Схема Гиббса $\longrightarrow$ Продолжающиеся цепи Гиббса

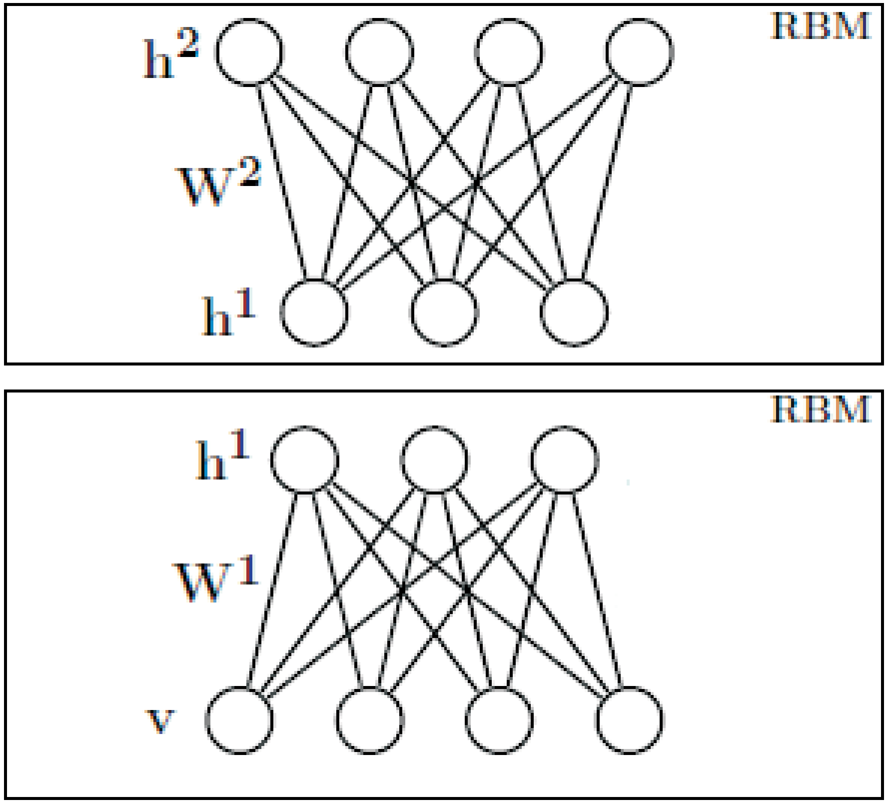
Инициализируем M цепей: $\left(v_0{}^m, h_0^{1\,m}, h_0^{2\,m}\right)$

На каждой итерации один шаг по схеме Гиббса:

$$\left(v_{iter}{}^m, h_{iter}^{1}{}^m, h_{iter}^{2}{}^m\right)$$

$$E_{\mathrm{mod}el}\left(vh^{1^T}\right) = \sum_{m=1}^{M} v_{iter}^m h_{iter}^{1}{}^{mT} \qquad E_{\mathrm{mod}el}\left(h^1 h^{2^T}\right) = \sum_{m=1}^{M} h_{iter}^{1}{}^m h_{iter}^{2}{}^{mT}$$

Предобучение весов



Алгоритм обучения DBM

- Вход: $\hat{V} = \{\hat{v}^1, \dots, \hat{v}^N\}$
- Выход: W^1, W^2

1. Предобучение W^1, W^2
2. Инициализация M цепей Гиббса
3. для iter = 1 : #iter
4. Оценить E_{data} с помощью вариационного подхода
5. Оценить E_{model} с помощью цепей Гиббса
6. $W^1_{iter} = W^1_{iter-1} + \alpha_{iter}\left(E_{data}\left(vh^{1^T}\right) - E_{model}\left(vh^{1^T}\right)\right)$
7. $W^2_{iter} = W^2_{iter-1} + \alpha_{iter}\left(E_{data}\left(h^1 h^{2^T}\right) - E_{model}\left(h^1 h^{2^T}\right)\right)$

Оценка качества обучения

- Задача: $p\left(\hat{V}\middle|\theta\right) \to \max\limits_{W}$

- Проверяем: $p\left(\hat{V}\middle|\theta\right), \quad p\left(V_{generate}\middle|\theta\right)$

- Вариационная оценка:

$$\log p\left(v\middle|\theta\right) \geq L(q) = \underbrace{-\sum\nolimits_h q\left(h\middle|\mu\right) E\left(v, h\middle|\theta\right) + H(q)}_{F} - \log Z(q)$$

$$L(q) \to \max_{\mu} \quad \Longrightarrow \quad F$$

Схема Гиббса для оценки Z

- Даны: $p_A(t) = \dfrac{\tilde{p}_A(t)}{Z_A}$ $\qquad p_B(t) = \dfrac{\tilde{p}_B(t)}{Z_B}$
- Если они близки:

$$\frac{Z_B}{Z_A} = \frac{\int \tilde{p}_B(t)dt}{Z_A} = \int \frac{\tilde{p}_B(t)}{Z_A}dt = \int \frac{\tilde{p}_B(t)}{\tilde{p}_A(t)} p_A(t)dt \approx \frac{1}{M}\sum_{i=1}^{M} \frac{\tilde{p}_B(t^m)}{\tilde{p}_A(t^m)}$$

- Иначе:

$$0 = \beta_1 < \beta_2 < \ldots < \beta_l = 1 \quad \Longrightarrow \quad \tilde{p}_l(t) = [\tilde{p}_A(t)]^{1-\beta_l}[\tilde{p}_B(t)]^{\beta_l}$$

$$\frac{Z_B}{Z_A} = \frac{Z_l}{Z_1} = \frac{Z_l}{Z_{l-1}} * \ldots * \frac{Z_2}{Z_1}$$

Оценка Z: RBM

$$p_B(v,h|W)=\frac{\exp(-E(v,h|W))}{Z_B(W)} \qquad \tilde{p}_B(v,h|W)=\exp(-E(v,h|W))$$

$p_A(v,h|W)$ - равномерное

Выборка из $\tilde{p}_l(v,h|W)=\tilde{p}_B{}^{\beta_l}(v,h|W)$ по схеме Гиббса:

$$p(h_j=1|v,W)=\frac{1}{1+\exp\left(-\sum_i W_{ij}v_i\right)}$$

$$p(v_i=1|h,W)=\frac{1}{1+\exp\left(-\sum_j W_{ij}h_j\right)}$$

Оценка Z: DBM

$$p_B\left(v,h^1,h^2|\theta\right)=\frac{\exp\left(-E\left(v,h^1,h^2|\theta\right)\right)}{Z_B(\theta)} \qquad \tilde{p}_B\left(v,h^1,h^2|\theta\right)=\exp\left(-E\left(v,h^1,h^2|\theta\right)\right)$$

$p_A\left(v,h^1,h^2|\theta\right)$ - равномерное

$$\tilde{p}_B\left(v,h^2|\theta\right)=\sum_{h^1}\exp\left(-E\left(v,h^1,h^2|\theta\right)\right)=\prod_j\left(1+\exp\left(\sum_i W_{ij}^1 v_i+\sum_k W_{jk}^2 h_k^2\right)\right)$$

$$\tilde{p}_l\left(v_i=1|\theta,h^2,v_{\setminus i}\right)$$

$$\tilde{p}_l\left(h_k^2=1|\theta,h_{\setminus k}^2,v\right)$$

Эксперименты: RBM и DBM

Порождающая модель:

Обучение

$\Downarrow$

Генерация по схеме Гиббса

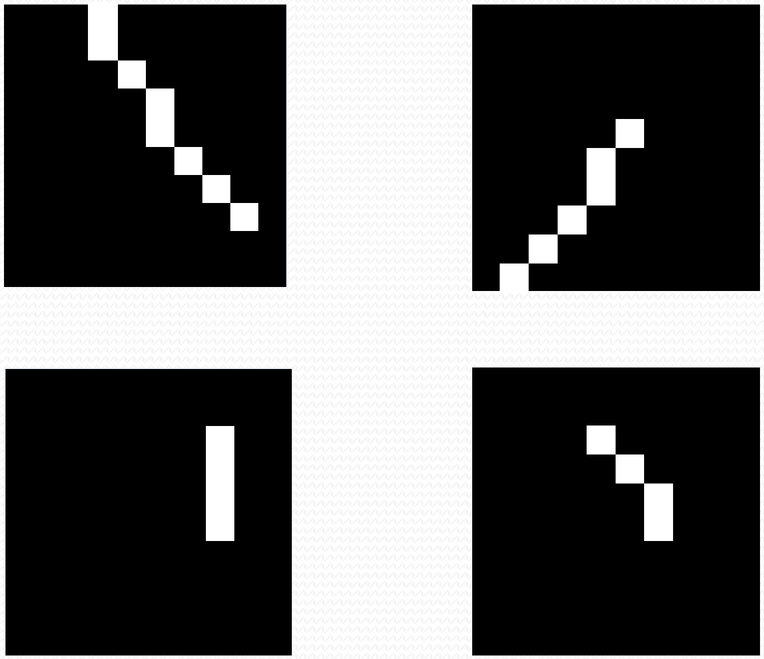
$\Downarrow$

Сравнение обучающих и полученных объектов

Задача

«Линии»

10000 картинок 10*10



RBM

- Веса лучше всего инициализировать случайными значениями из $N\left(x \middle| 0, \sigma^2\right)$, где $\sigma^2 \leq 0.1$.

- Примерное время работы при dim(v)=100 , N=10000, без деления выборки.

dim(h)	1 шаг	ceil(iter/10) шагов
50	160	1809
100	236	2890
200	496	4497

RBM и «линии»

- dim(h)=80, без деления выборки.

Вид	t	p_learn	p_generate
1 шаг	302	6.5173e-009	2.5784e-006
ceil(iter/10) шагов	2167	2.3352e-007	1.7733e-007

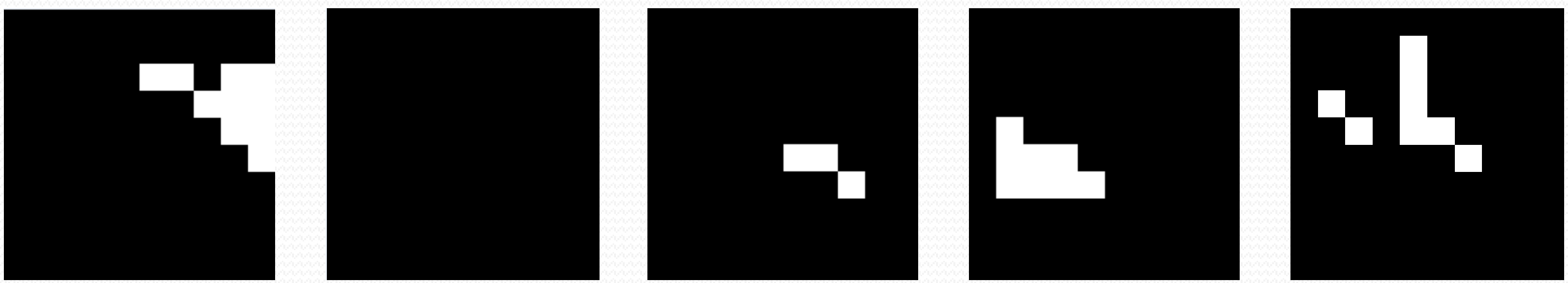
- dim(h)=80, с делением выборки.

Вид	t	p_learn	p_generate
1 шаг	236	4.4137e-006	2.4552e-005
ceil(iter/10) шагов	1356	9.5253e-008	6.4714e-005

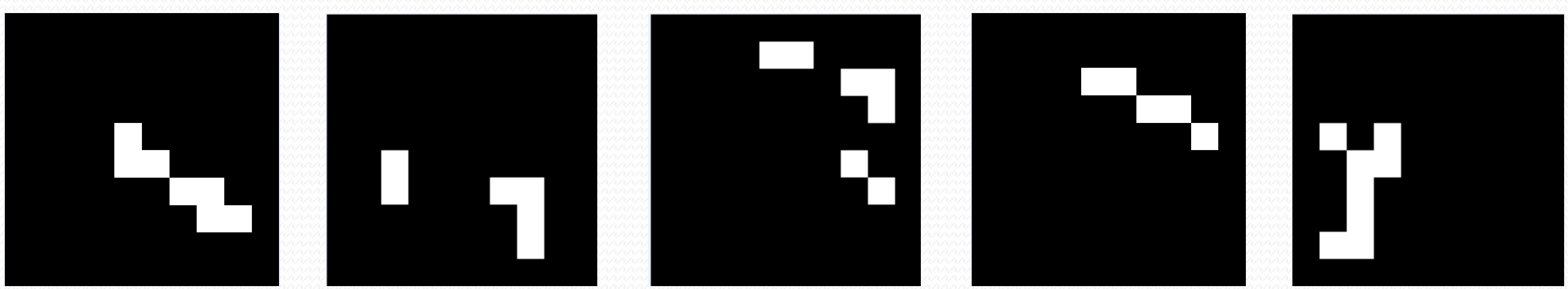
RBM и «линии»

Без деления выборки:

- 1 шаг:



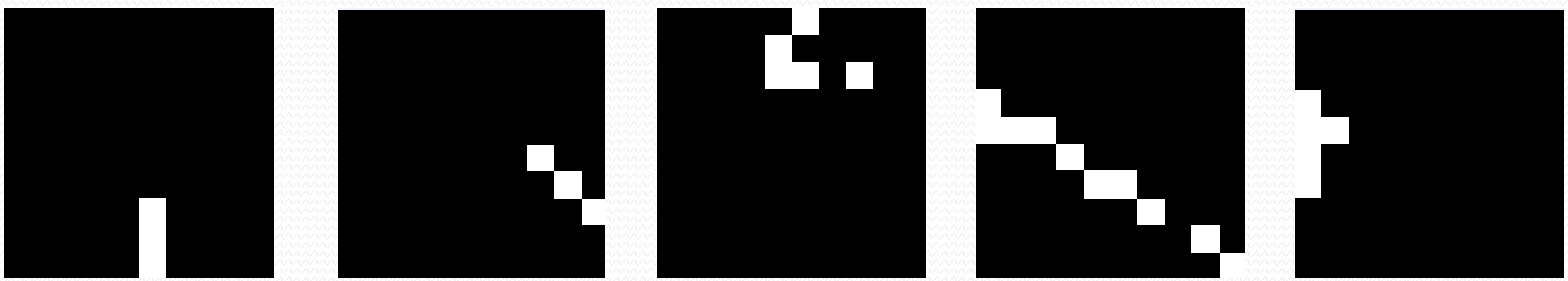
- ceil(iter/10) шагов:



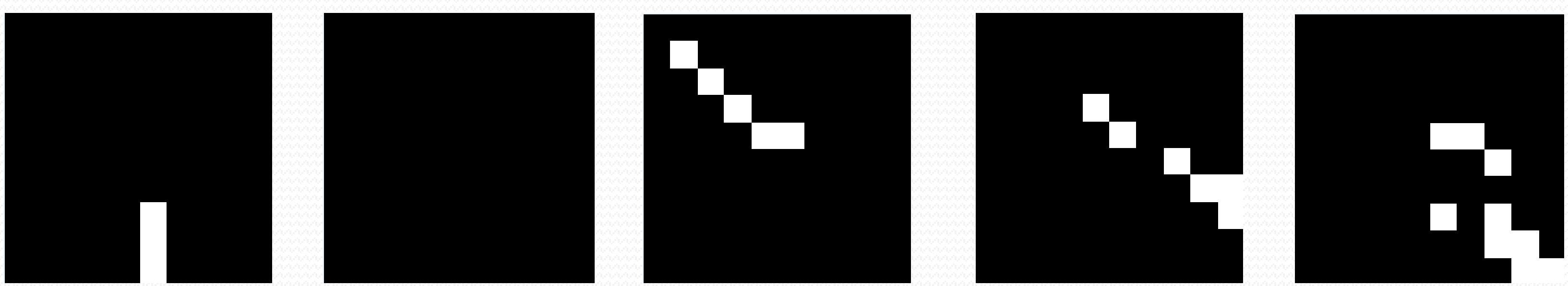
RBM и «линии»

С делением выборки:

- 1 шаг:



- ceil(iter/10) шагов:



Химическая задача

- Обучение DBM с непрерывными переменными;
- Оценка значений переменных последнего скрытого слоя;
- Использование переменных последнего скрытого слоя как новых признаков;