

Max-Margin Boltzmann Machines for object segmentation

Игорь Гитман

ВМК МГУ



Jimei Yang, Simon Safar, Ming-Hsuan Yang
University of California, Merced

План доклада

Введение

- RBM
- DBM
- ShapeBM

Max-Margin Boltzmann Machines

- Общая структура
- ICM алгоритм
- CCCP алгоритм
- Обучение MMBM
- Эксперименты

Выводы

Введение

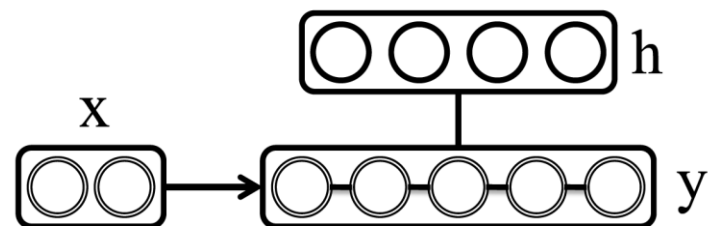
Задача: сегментация изображения

Классические методы



CRF (2011)

Superspixel-based
Conditional Random Fields



CHOPPs (2013)

Compositional High Order
Pattern Potentials

Введение



CRF



MMBM

Введение



CHOPPs

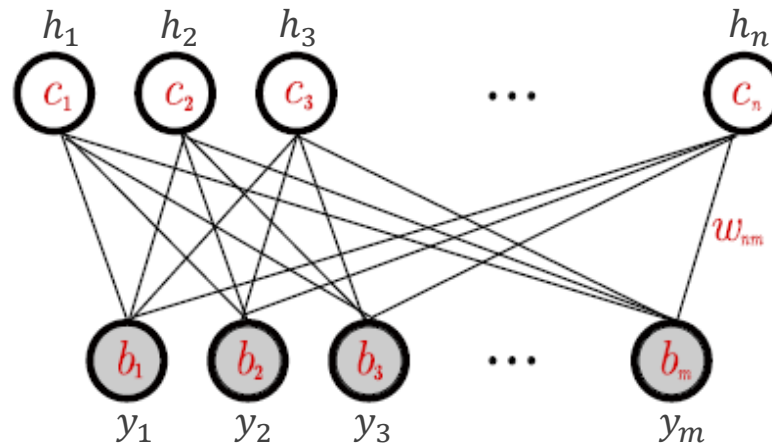


MMBM

Restricted Boltzmann Machines

Задача: генерация изображений

$$p(y, h) = \frac{e^{-E(y, h)}}{\sum_{y, h} e^{-E(y, h)}} = \frac{1}{Z} e^{-E(y, h)} = \frac{1}{Z} e^{y^T W h + b^T y + c^T h}$$



Restricted Boltzmann Machines

Задача: генерация изображений

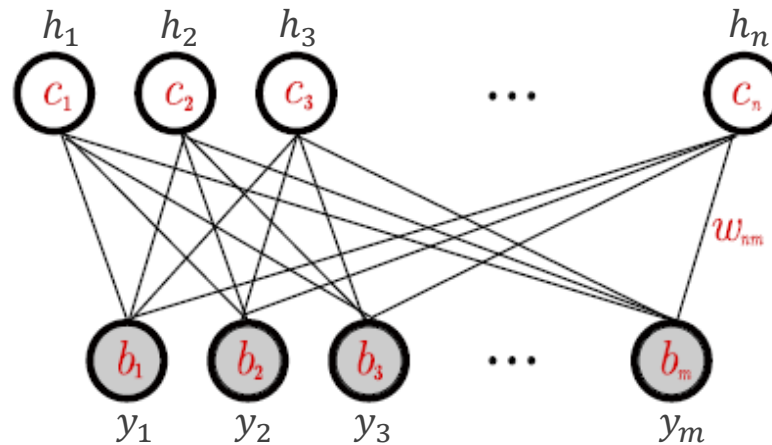
$$p(y|h) = \prod_{i=1}^m p(y_i|h)$$

и

$$p(h|y) = \prod_{i=1}^n p(h_i|y)$$

$$p(y_i = 1|h) = \frac{1}{1 + e^{-\sum_{j=1}^m w_{ij}h_j - b_i}}$$

$$p(h_j = 1|y) = \frac{1}{1 + e^{-\sum_{i=1}^n w_{ij}y_i - c_j}}$$



Restricted Boltzmann Machines

Задача: $\theta^* = \operatorname{argmax}_{\theta} [\log(p(y; \theta))]$

$$\frac{\partial \log p(y; \theta)}{\partial \theta} = ?$$

Restricted Boltzmann Machines

Задача: $\theta^* = \operatorname{argmax}_{\theta} [\log(p(y; \theta))]$

\Downarrow^{θ}

$$\frac{\partial \log p(y; \theta)}{\partial w_{ij}} = - \sum_h p(h|y) \frac{\partial E(y, h)}{\partial w_{ij}} + \sum_{y, h} p(h, y) \frac{\partial E(y, h)}{\partial w_{ij}}$$

Restricted Boltzmann Machines

Задача: $\theta^* = \operatorname{argmax}_{\theta} [\log(p(y; \theta))]$

\Downarrow^{θ}

$$\begin{aligned} \frac{\partial \log p(y; \theta)}{\partial w_{ij}} &= - \sum_h p(h|y) \frac{\partial E(y, h)}{\partial w_{ij}} + \sum_{y, h} p(h, y) \frac{\partial E(y, h)}{\partial w_{ij}} = \\ &= -\mathbb{E}_{p(h|y)} y_i h_j + \mathbb{E}_{p(y, h)} y_i h_j \end{aligned}$$

Restricted Boltzmann Machines

Задача: $\theta^* = \operatorname{argmax}_{\theta} [\log(p(y; \theta))]$

\Downarrow^{θ}

$$\begin{aligned} \frac{\partial \log p(y; \theta)}{\partial w_{ij}} &= - \sum_h p(h|y) \frac{\partial E(y, h)}{\partial w_{ij}} + \sum_{y, h} p(h, y) \frac{\partial E(y, h)}{\partial w_{ij}} = \\ &= -\mathbb{E}_{p(h|y)} y_i h_j + \mathbb{E}_{p(y, h)} y_i h_j = \\ &= p(h_j = 1|y) y_i + \sum_y p(y) \sum_h p(h|y) y_i h_j \end{aligned}$$

Restricted Boltzmann Machines

Задача: $\theta^* = \operatorname{argmax} [\log(p(y; \theta))]$

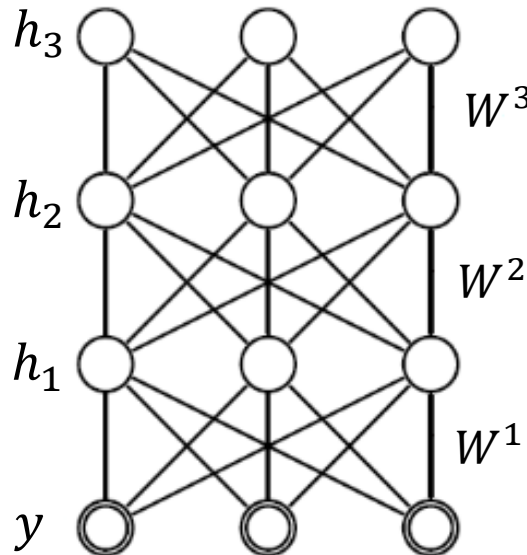
\Downarrow^{θ}

$$\begin{aligned} \frac{\partial \log p(y; \theta)}{\partial w_{ij}} &= - \sum_h p(h|y) \frac{\partial E(y, h)}{\partial w_{ij}} + \sum_{y, h} p(h, y) \frac{\partial E(y, h)}{\partial w_{ij}} = \\ &= -\mathbb{E}_{p(h|y)} y_i h_j + \mathbb{E}_{p(y, h)} y_i h_j = \\ &= p(h_j = 1|y) y_i + \sum_y p(y) \sum_h p(h|y) y_i h_j \approx \\ &\approx p(h_j = 1|y) y_i + \sum_h p(h|y^{(k)}) y_i^{(k)} h_j = CD_k(w_{ij}, y) \end{aligned}$$

Deep Boltzmann Machines

Задача: генерация изображений

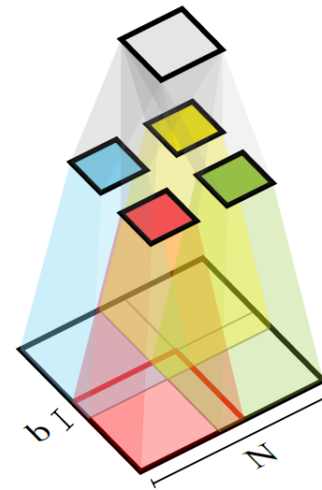
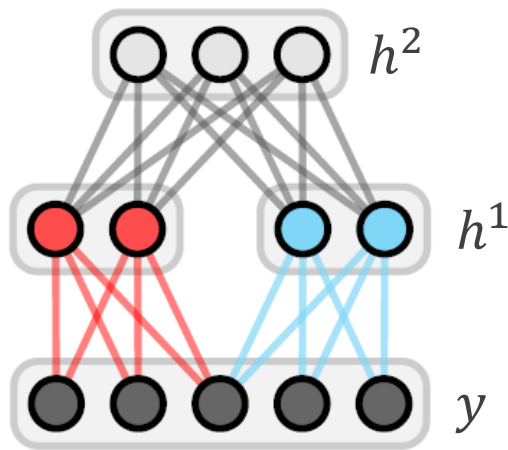
$$p(y, h) = \frac{e^{-E(y, h)}}{\sum_{y, h} e^{-E(y, h)}} = \frac{1}{Z} e^{-E(y, h)} = \frac{1}{Z} e^{y^T W^1 h^1 + h^1 T W^2 h^2 + h^2 T W^3 h^3 + a^T y + b^T h^1 + c^T h^2 + d^T h^3}$$



Shape Boltzmann Machines

$$E(y, h^1, h^2) = -y^T W^1 h^1 - b^T y - c^{1T} h^1 - h^{1T} W^2 h^2 - c^{2T} h^2$$

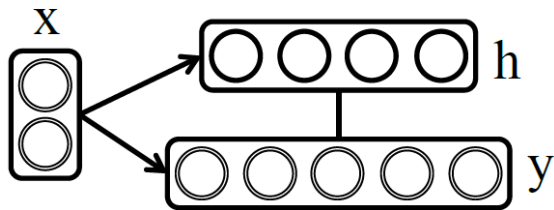
$p(y|h^1)$, $p(h^1|y, h^2)$, $p(h^2|h^1)$ — условно независимы



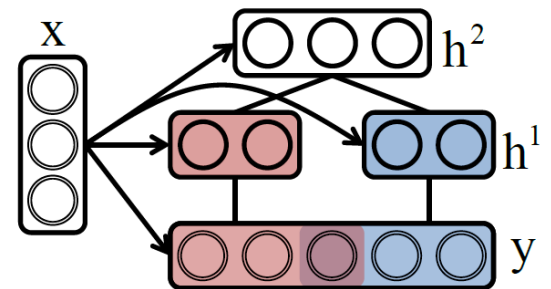
Max-Margin Boltzmann Machines

$$E(y, h, x) = -y^T W h - h^T (V^1 x^1 + c) - y^T (V^0 x^0 + b)$$

$$E(y, h^1, h^2, x) = -y^T W^1 h^1 - h^1{}^T W^2 h^2 - h^1{}^T (V^1 x^1 + c^1) - \\ - h^2{}^T (V^2 x^2 + c^2) - y^T (V^0 x^0 + b)$$



MMBM1: $p(y, h|x)$



MMBM2: $p(y, h^1, h^2|x)$

Max-Margin Boltzmann Machines

$$E(y, h, x) = -y^T W h - h^T (V^1 x^1 + c) - y^T (V^0 x^0 + b)$$

$$E(y, h^1, h^2, x) = -y^T W^1 h^1 - h^{1T} W^2 h^2 - h^{1T} (V^1 x^1 + c^1) - \\ - h^{2T} (V^2 x^2 + c^2) - y^T (V^0 x^0 + b)$$

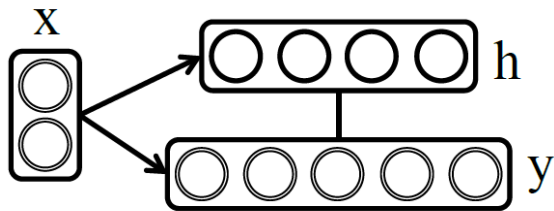
x^0 — SIFT, гистограммы цвета и границ по суперпикселям

x^1 — HOG дескрипторы всей картинки или 4 частей

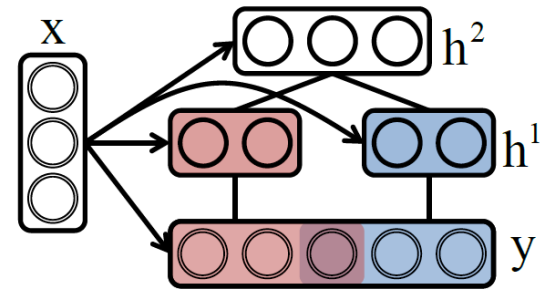
x^2 — HOG дескрипторы всей картинки

Max-Margin Boltzmann Machines

Задача вывода: $y^* = \operatorname{argmax}_y p(y|x) = \operatorname{argmax}_y \sum_H p(y, H|x)$



MMBM1: $p(y, h|x)$

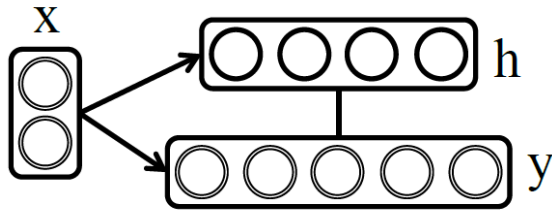


MMBM2: $p(y, h^1, h^2|x)$

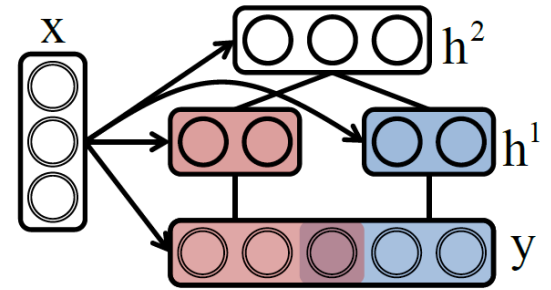
Max-Margin Boltzmann Machines

Задача вывода: $y^* = \operatorname{argmax}_y p(y|x) = \operatorname{argmax}_y \sum_H p(y, H|x)$

$$\{y^*, H^*\} = \operatorname{argmax}_{y, H} p(y, H|x)$$



MMBM1: $p(y, h|x)$



MMBM2: $p(y, h^1, h^2|x)$

Iterated Conditional Modes

Задача: $\{y^*, H^*\} = \operatorname{argmax}_{y, H} p(y, H|x)$

MMBM1

1. Инициализируем h
 2. До сходимости:
 3. $y \leftarrow \max p(y|h, x)$
 4. $h \leftarrow \max p(h|y, x)$
 5. Завершаемся
-

MMBM2

1. Инициализируем h^1
 2. До сходимости:
 3. $h^2 \leftarrow \max p(h^2|h^1, x)$
 4. $y \leftarrow \max p(y|h^1, x)$
 5. $h^1 \leftarrow \max p(h^1|y, h^2, x)$
 6. Завершаемся
-

$$p(h_{i+1}, y_i) = p(h_{i+1}|y_i)p(y_i) \geq p(h_i|y_i)p(y_i) = p(h_i, y_i)$$

$$p(y_{i+1}, h_i) = p(y_{i+1}|h_i)p(h_i) \geq p(y_i|h_i)p(h_i) = p(y_i, h_i)$$

Обучение ММВМ

1. Предобучение: минимизация contrastive divergence для RBM и для каждой пары слоев ShapeVM.
2. Max-Margin оптимизационная задача.

Обозначения

$\{x_i, y_i\}_{i=1}^t$ — обучающая выборка изображение/сегментация

Для ММВМ1: $H = h$ и $w = \{W, V^{0,1}, c, b\}$

Для ММВМ2: $H = \{h^1, h^2\}$ и $w = \{W^{1,2}, V^{0,1,2}, c^{1,2}, b\}$

Обучение ММВМ

$$\text{Найти } w : p(y_i, H|x_i; w) \geq p(y, H|x_i; w) \forall H, \forall y, \forall i \quad \Leftrightarrow$$

$$-E(y_i, H, x_i; w) \geq -E(y, H, x_i; w) \forall H, \forall y, \forall i \quad \Leftrightarrow$$

$$\left| H_i^* = \operatorname{argmax}_H -E(y_i, H|x_i; w) \right| \quad \Leftrightarrow$$

$$\max_H \left[-E(y_i, H|x_i; w) \right] \geq \max_{y, H} \left[-E(y, H, x_i; w) + \Delta(y, y_i, H, H_i^*) \right] \forall i$$

Обучение ММВМ

Найти w : $\min_{\xi, w} \frac{\gamma}{2} \|w\|^2 + \sum_i \xi_i$, при

$$\max_H [-E(y_i, H | x_i; w)] \geq \max_{y, H} [-E(y, H, x_i; w) + \Delta(y, y_i, H, H_i^*)] - \xi_i \quad \forall i$$
$$\xi_i \geq 0 \quad \forall i$$

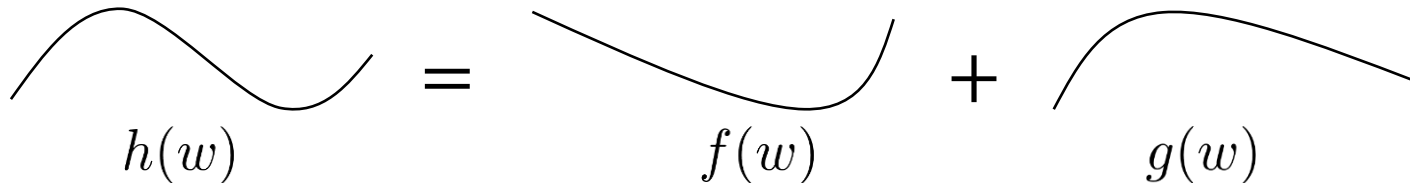


$$\min_w \underbrace{\frac{\gamma}{2} \|w\|^2 + \sum_i \left(\max_{y, H} [-E(y, H, x_i; w) + \Delta(y, y_i, H, H_i^*)] \right)}_{\text{Выпуклая часть}} - \underbrace{\sum_i \left(\max_H [-E(y_i, H, x_i, w)] \right)}_{\text{Вогнутая часть}}$$

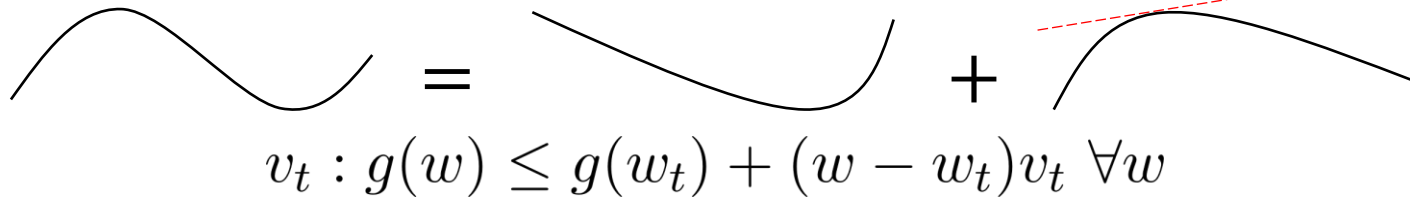
Concave-Convex Procedure

Задача: $\min_w h(w)$

1. Раскладываем функцию на выпуклую и вогнутую части:


$$h(w) = f(w) + g(w)$$

2. Ограничиваем вогнутую часть сверху гиперплоскостью:


$$v_t : g(w) \leq g(w_t) + (w - w_t)v_t \quad \forall w$$

3. Минимизируем получившуюся выпуклую функцию:


$$w_{t+1} = \operatorname{argmin}_w [f(w) + wv_t]$$

4. Повторяем до сходимости: $[f(w_{t+1}) + g(w_{t+1})] - [f(w_t) + g(w_t)] < \epsilon$

Concave-Convex Procedure

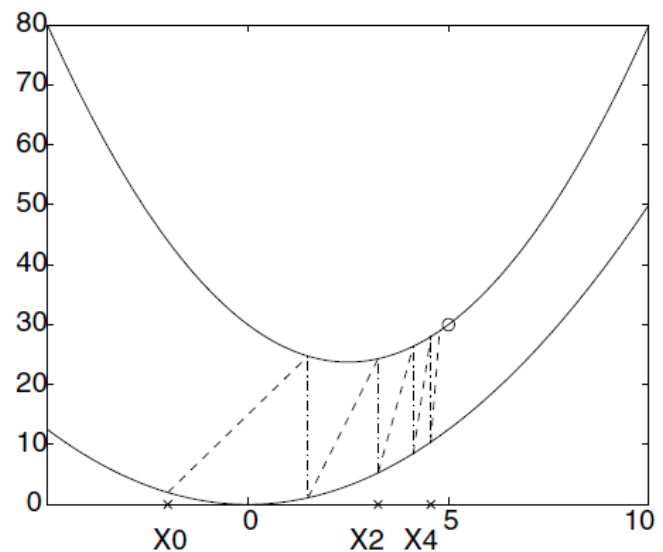
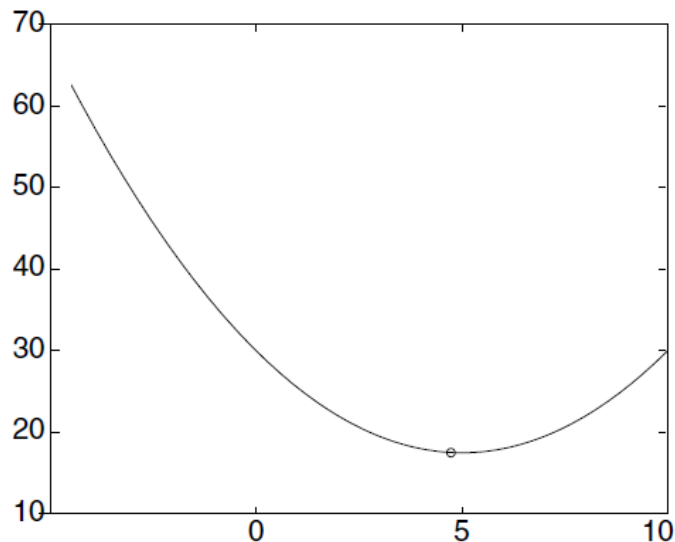
$$\nabla f(w_{t+1}) = -v_t = -\nabla g(w_t)$$

$$f(w_t) \geq f(w_{t+1}) + (w_t - w_{t+1})\nabla f(w_{t+1})$$

$$g(w_{t+1}) \leq g(w_t) + (w_{t+1} - w_t)\nabla g(w_t)$$

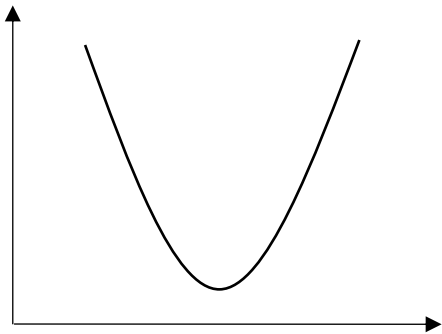


$$f(w_{t+1}) + g(w_{t+1}) \leq f(w_t) + g(w_t)$$

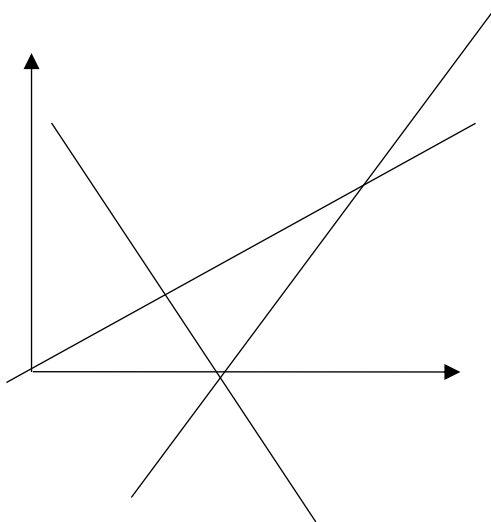


Обучение ММВМ

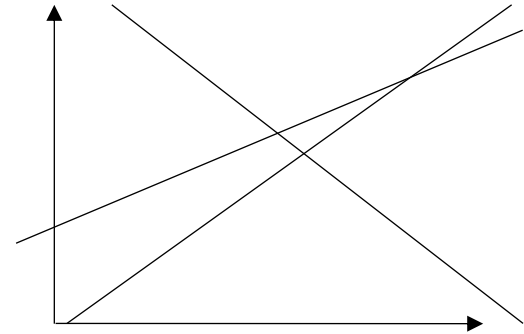
$$\min_w \underbrace{\frac{\gamma}{2} \|w\|^2 + \max_{y, H} [-E(y, H, x_i; w) + \Delta(y, y_i, H, H_i^*)]}_{\text{Выпуклая часть}} - \underbrace{\max_H [-E(y_i, H, x_i, w)]}_{\text{Вогнутая часть}}$$



+

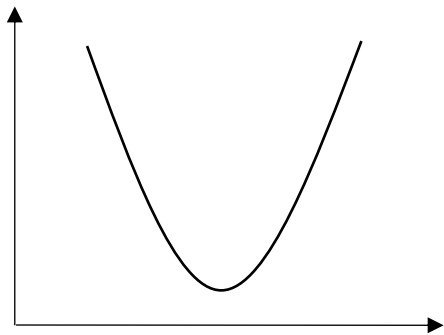


+

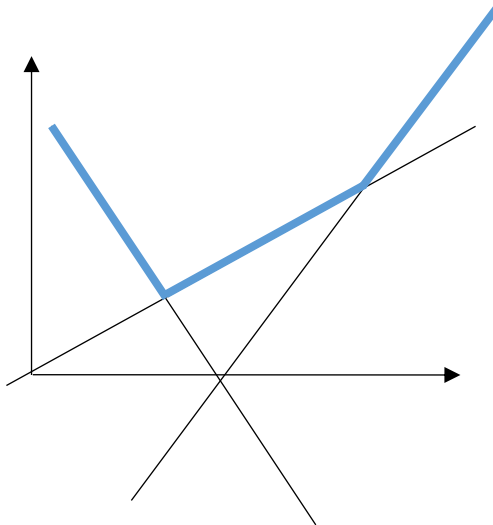


Обучение ММВМ

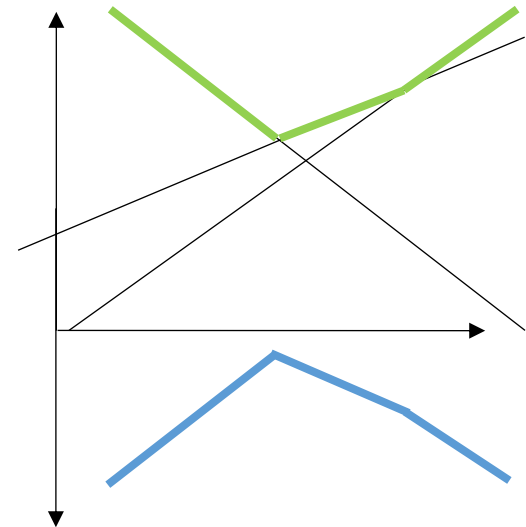
$$\min_w \underbrace{\frac{\gamma}{2} \|w\|^2 + \max_{y, H} [-E(y, H, x_i; w) + \Delta(y, y_i, H, H_i^*)]}_{\text{Выпуклая часть}} - \underbrace{\max_H [-E(y_i, H, x_i, w)]}_{\text{Вогнутая часть}}$$



+

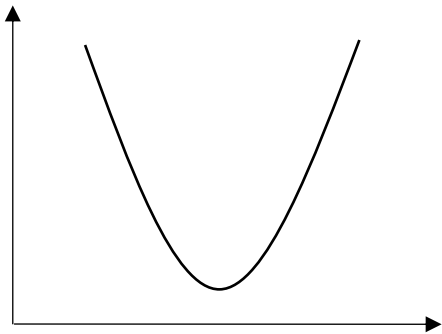


+

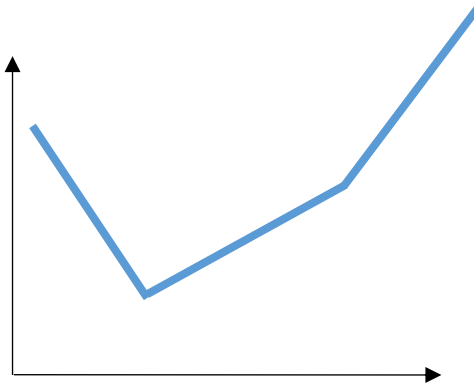


Обучение ММВМ

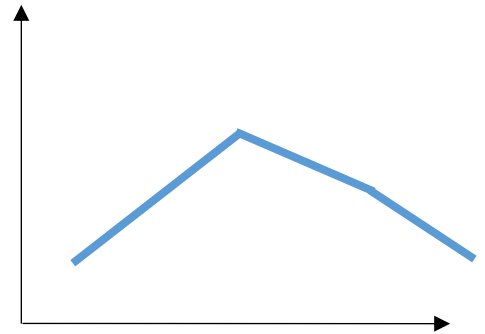
$$\min_w \underbrace{\frac{\gamma}{2} \|w\|^2 + \max_{y, H} [-E(y, H, x_i; w) + \Delta(y, y_i, H, H_i^*)]}_{\text{Выпуклая часть}} - \underbrace{\max_H [-E(y_i, H, x_i, w)]}_{\text{Вогнутая часть}}$$



+

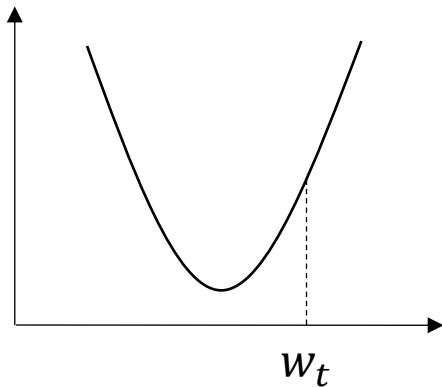


+

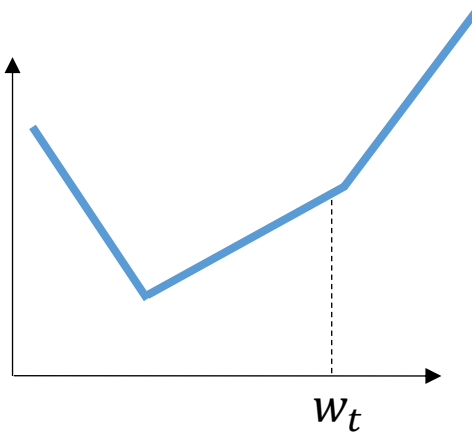


Обучение ММВМ

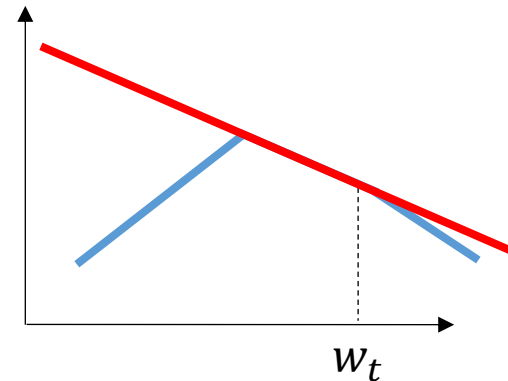
$$\min_w \underbrace{\frac{\gamma}{2} \|w\|^2 + \max_{y, H} [-E(y, H, x_i; w) + \Delta(y, y_i, H, H_i^*)]}_{\text{Выпуклая часть}} - \underbrace{\max_H [-E(y_i, H, x_i, w)]}_{\text{Вогнутая часть}}$$



+

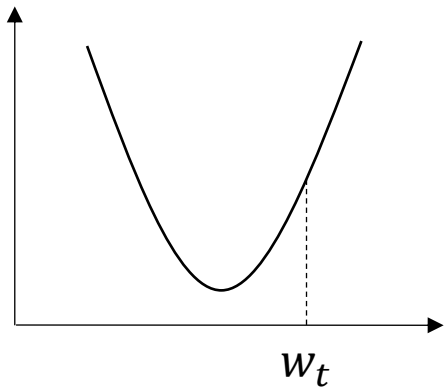


+

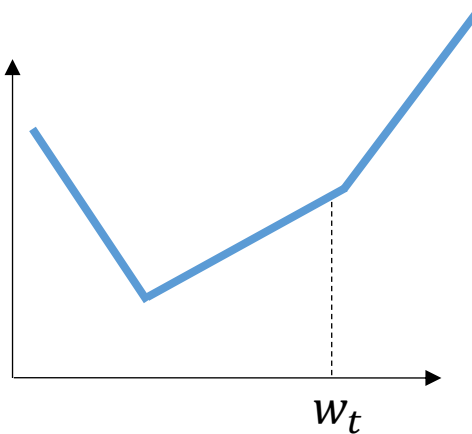


Обучение ММВМ

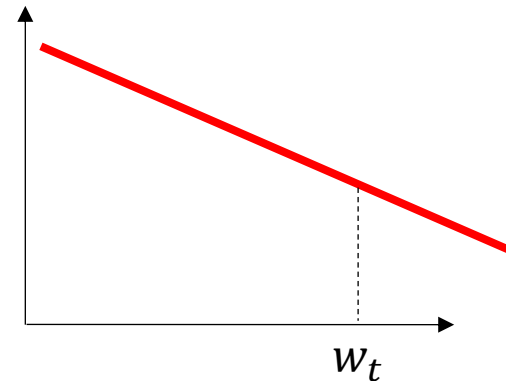
$$\min_w \underbrace{\frac{\gamma}{2} \|w\|^2 + \max_{y, H} [-E(y, H, x_i; w) + \Delta(y, y_i, H, H_i^*)]}_{\text{Выпуклая часть}} - \underbrace{[-E(y_i, H_i^*, x_i, w)]}_{\text{Линейная часть}}$$



+

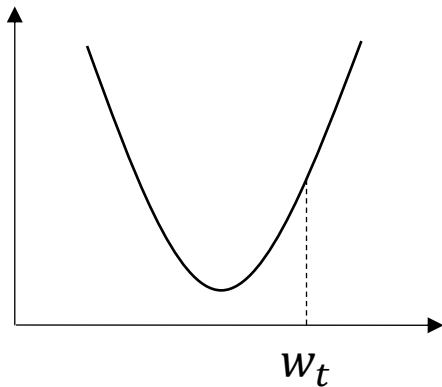


+

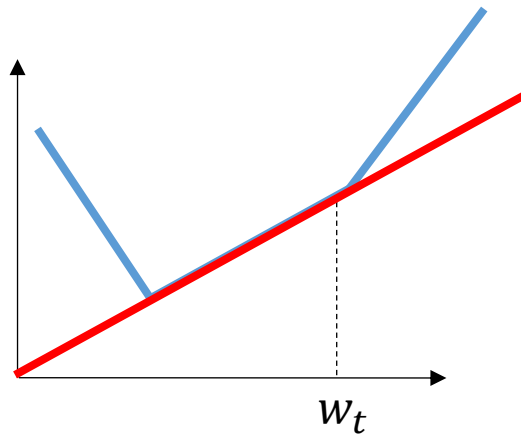


Обучение ММВМ

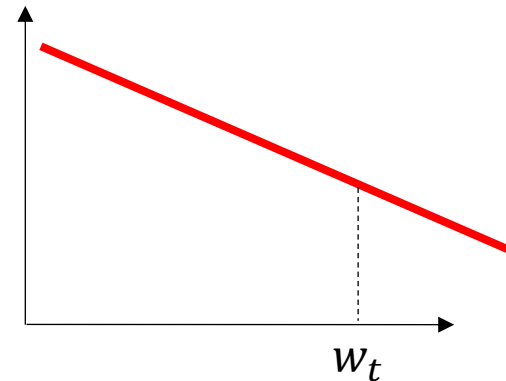
$$\min_w \underbrace{\frac{\gamma}{2} \|w\|^2 + \max_{y, H} [-E(y, H, x_i; w) + \Delta(y, y_i, H, H_i^*)]}_{\text{Выпуклая часть}} - \underbrace{[-E(y_i, H_i^*, x_i, w)]}_{\text{Линейная часть}}$$



+

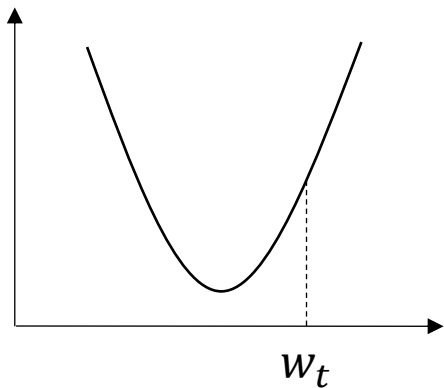


+

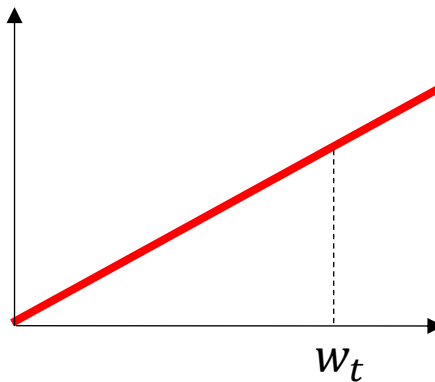


Обучение ММВМ

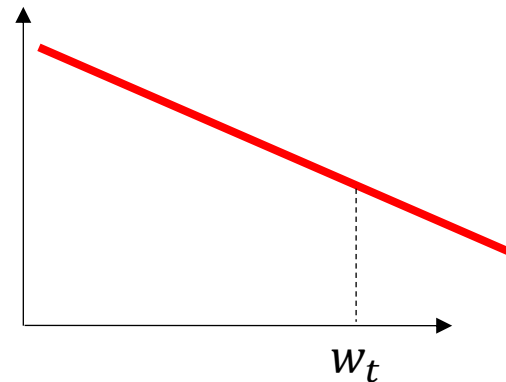
$$\min_w \underbrace{\frac{\gamma}{2} \|w\|^2 + [-E(\tilde{y}_i, \tilde{H}_i, x_i; w) + \Delta(\tilde{y}_i, y_i, \tilde{H}_i, H_i^*)]}_{\text{Выпуклая часть}} - \underbrace{[-E(y_i, H_i^*, x_i, w)]}_{\text{Линейная часть}}$$



+



+



Общий алгоритм обучения ММВМ

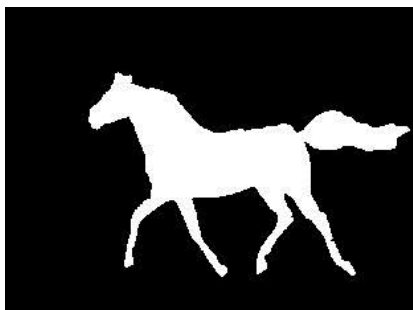
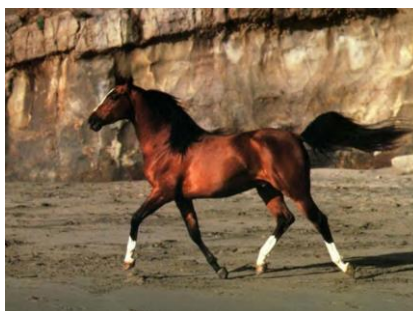
1. Выбираем $w_0, \alpha, \gamma; t = 0$
2. Пока $t < T$:
 3. Случайно выбираем (x_i, y_i)
 4. Решаем: $H_i^* = \operatorname{argmax}_H [-E(y_i, H, x_i, w_t)]$
 5. Решаем: $\tilde{y}_i, \tilde{H}_i = \operatorname{argmax}_{y, H} [-E(y, H, x_i, w_t) + \Delta(y, y_i, H, H_i^*)]$
 6. $w_{t+1} = w_t - \alpha \left[\gamma w_t + \frac{\partial E(\tilde{y}_i, \tilde{H}_i, x_i, w)}{\partial w} - \frac{\partial E(y_i, H_i^*, x_i, w)}{\partial w} \right]$
 7. Уменьшаем α
8. Завершаемся, или переходим на шаг 2

Эксперименты

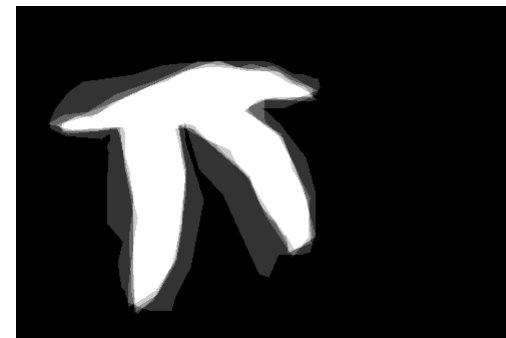
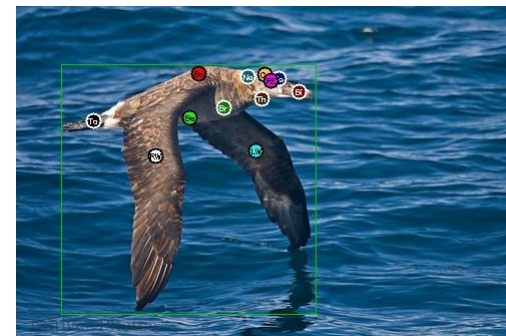
Penn-Fudan Pedestrians
423 картинки



Weizmann Horses
328 картинок

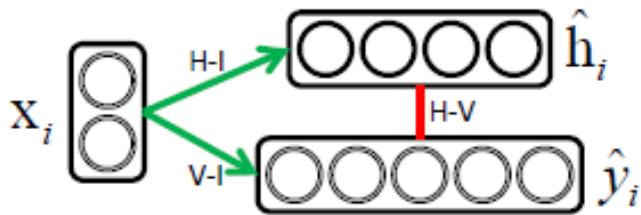


Caltech-UCSD Birds 200
6033 картинки

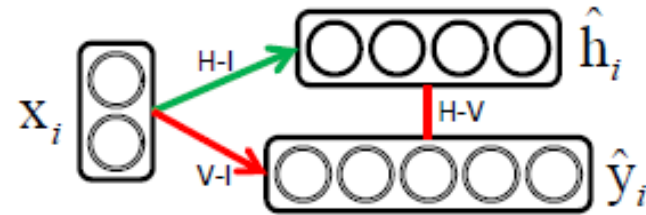


Эксперименты

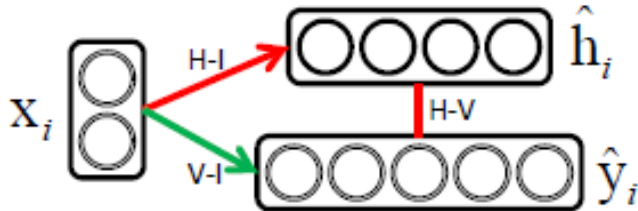
Case 1: $\Delta(\cdot) = 0$



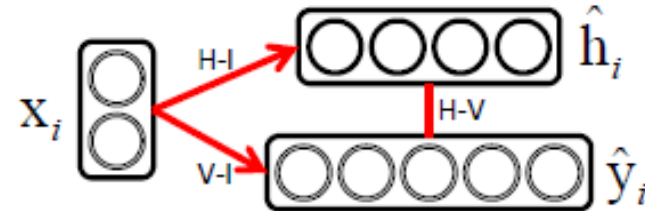
Case 2: $\Delta(\cdot) = \Delta(\mathbf{y}, \mathbf{y}_i)$



Case 3: $\Delta(\cdot) = \Delta(\mathbf{H}, \mathbf{H}_i^*)$



Case 4: $\Delta(\cdot) = \Delta(\mathbf{y}, \mathbf{y}_i) + \Delta(\mathbf{H}, \mathbf{H}_i^*)$



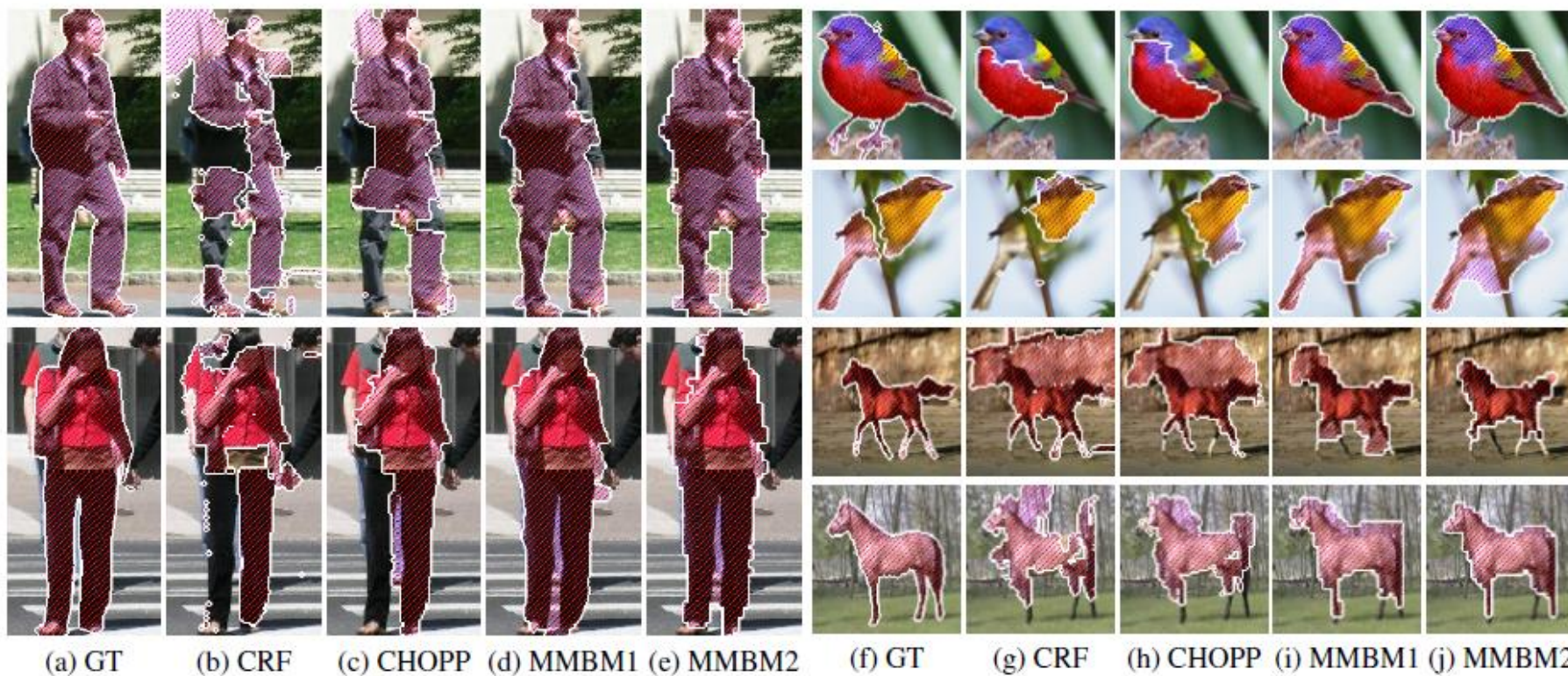
Average Pixel Accuracy (AP): количество «верных» пикселей ко всем;

Intersection-over-Union (IoU): $\frac{|Y \cap Y^*|}{|Y \cup Y^*|}$

Эксперименты

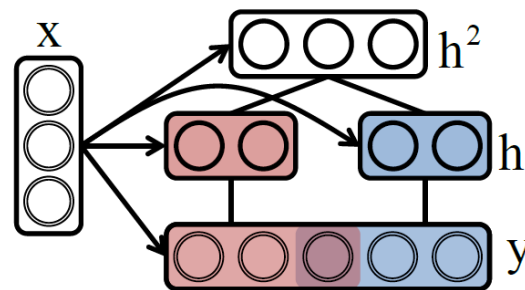
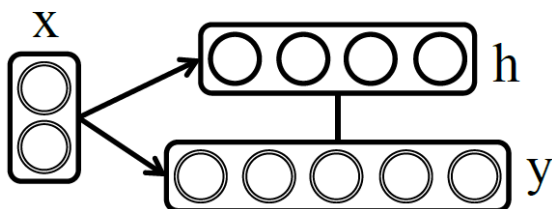
		Pedestrains		Horses		Birds	
		AP	IoU	AP	IoU	AP	IoU
CRF		84.87	68.35	87.46	67.44	83.5	38.45
CHOPPs		86.55	71.33	88.67	71.6	74.52	48.84
MMBM1	Case 1	82.66	64.8	70.59	38.01	80.96	60.37
	Case 2	85.27	69.2	85.87	62.97	87.73	72.45
	Case 3	83.35	65.78	85.37	59.35	75.73	63.22
	Case 4	89.91	76.92	89.43	69.59	88.07	72.96
MMBM2	Case 4	89.74	77.3	89.8	72.09	86.38	69.87
MMBM1 w/ GC	Case 4	90.42	77.97	90.62	74.12	90.42	75.92
MMBM2 w/ GC	Case 4	90.77	79.42	90.71	75.78	90.77	72.4

Эксперименты



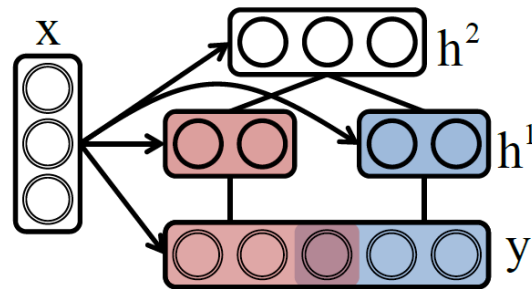
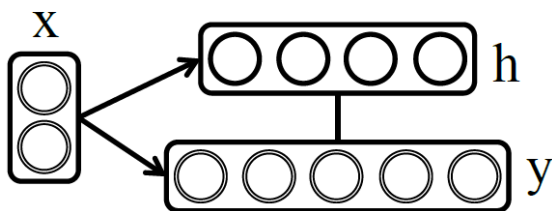
Выводы

1. Max-Margin Boltzmann Machines показывают лучшие результаты для решения задач сегментации.
2. Есть возможности развития:
 1. Добавить парные потенциалы на y -слой.
 2. Вместо «ручной работы» с признаками поставить сверточную сеть.

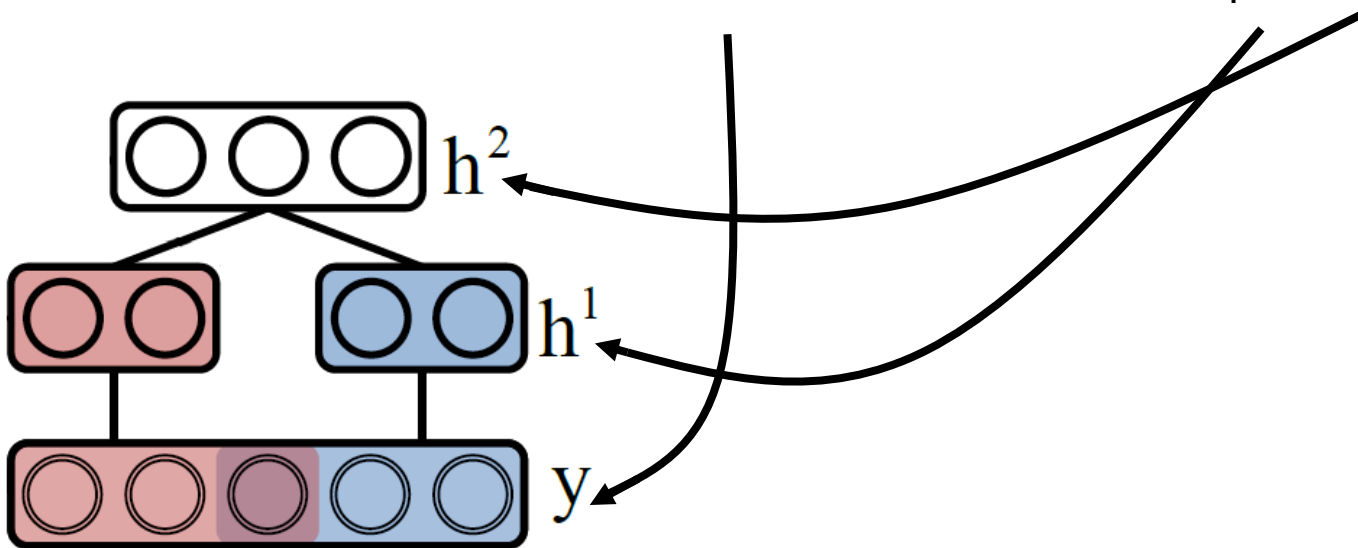
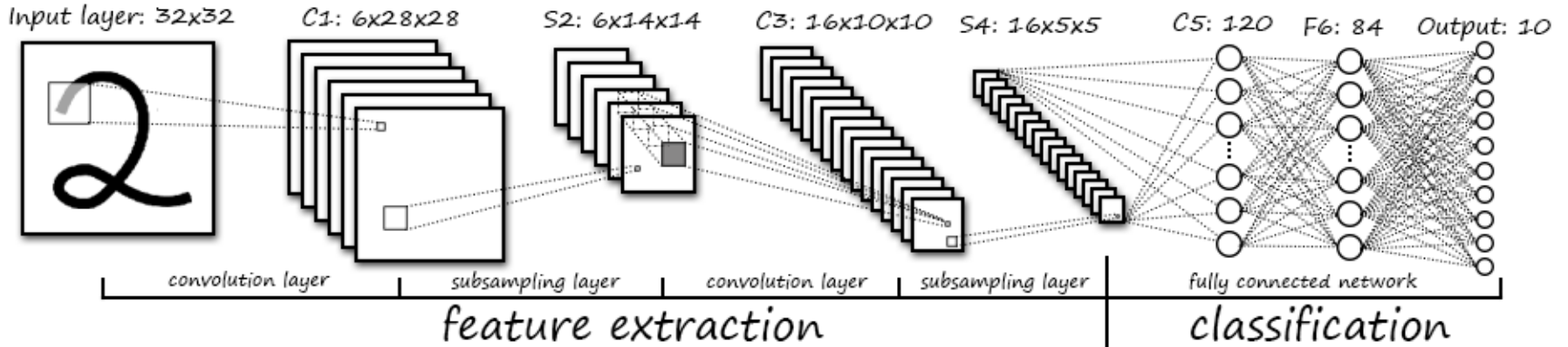


Выводы

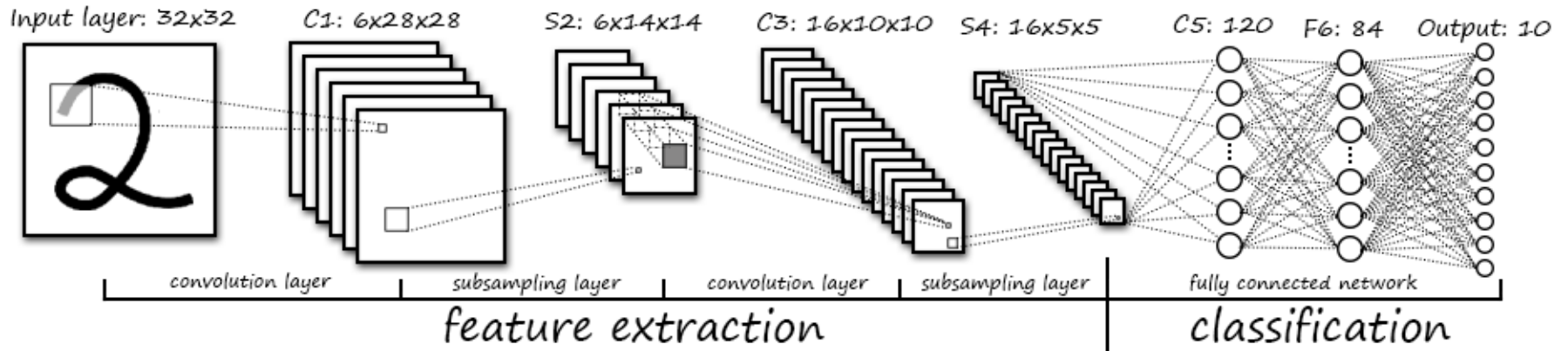
1. Max-Margin Boltzmann Machines показывают лучшие результаты для решения задач сегментации.
2. Есть возможности развития:
 1. Добавить парные потенциалы на y -слой.
 2. **Вместо «ручной работы» с признаками поставить сверточную сеть.**



Выводы



Выводы



Обычный алгоритм + backpropagation:

$$w_{t+1} = w_t - \alpha \left[\gamma w_t + \frac{\partial E(\tilde{y}_i, \tilde{H}_i, x_i, w)}{\partial w} - \frac{\partial E(y_i, H_i^*, x_i, w)}{\partial w} \right]$$