Структурное обучение с использованием логистической регресии

Justin Domke

2014

Table of Contents

Структурное обучение

$$y^* = \underset{y}{\operatorname{argmax}} \, F(x, y)$$

$$x \in \mathbb{R}^M$$

$$y \in \{1, \ldots, L\}^M$$

x y

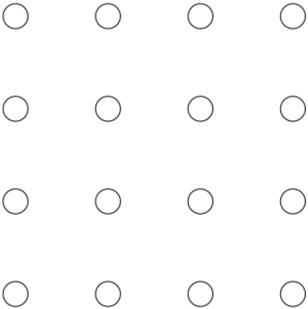
Очень часто F представляется в линейной форме:

$$F(x, y) = \sum_{\alpha} w^T \Phi_\alpha(x, y_\alpha)$$

Стурктурное обучение

Очень часто F представляется в линейной форме:

$$F(x, y) = \sum_{\alpha} w^T \Phi_\alpha(x, y_\alpha)$$

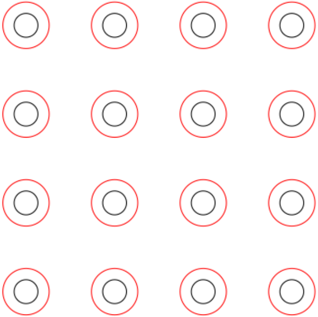


$$\{\alpha\} = \{ \qquad \}$$

Стурктурное обучение

Очень часто F представляется в линейной форме:

$$F(x, y) = \sum_{\alpha} w^T \Phi_\alpha(x, y_\alpha)$$

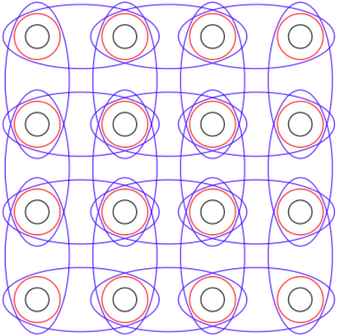


$$\{\alpha\} = \{\{1\}, \ldots, \{M\}, \qquad \}$$

Стурктурное обучение

Очень часто F представляется в линейной форме:

$$F(x,y)=\sum_{\alpha} w^T \Phi_\alpha(x,y_\alpha)$$



$$\{\alpha\}=\{\{1\},\ldots,\{M\},\{1,2\},\ldots,\{M-1,M\}\}$$

Структурное обучение

Дана выборка $(x^1, y^1), \ldots, (x^N, y^N)$

$$y^k \approx argmax_y F(x^k, y)$$

В случае линейной формы

$$F(x, y) = w^T \Phi_\alpha(x, y_\alpha)$$

проблема обучения сводится к нахождению $w \in \mathbb{R}^n$.
В данной работе возьмем

$$F(x, y) = \sum_\alpha f_\alpha(x, y_\alpha)$$

и проблема обучения сводится к нахождению $f_\alpha \in \mathbf{F}_\alpha$.

Table of Contents

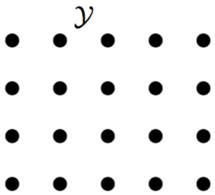
Функция потерь

Для выборки $\{(x^k, y^k)\}$ хотим минимизировать $R(F) = \sum_k (l(x^k, y^k, F))$.

Функция потерь

Для выборки $\{(x^k, y^k)\}$ хотим минимизировать $R(F) = \sum_k (l(x^k, y^k, F))$.
Классическая функция потерь:

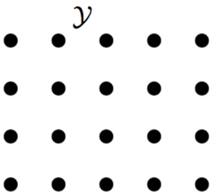
$$l(x^k, y^k, F) = -F(x^k, y^k) + max_{y \in \mathcal{Y}} F(x^k, y)$$



Функция потерь

Для выборки $\{(x^k, y^k)\}$ хотим минимизировать
$R(F) = \sum_k (l(x^k, y^k, F))$.
Классическая функция потерь:

$$l(x^k, y^k, F) = -F(x^k, y^k) + max_{y \in \mathcal{Y}} F(x^k, y) + \triangle(y^k, y)$$

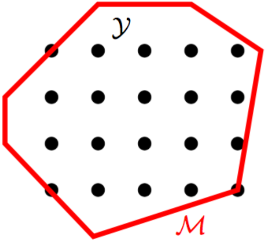


Функция потерь

Для выборки $\{(x^k, y^k)\}$ хотим минимизировать $R(F) = \sum_k (l(x^k, y^k, F))$.

Классическая функция потерь:

$$l(x^k, y^k, F) = -F(x^k, y^k) + max_{y \in \mathcal{Y}} F(x^k, y) + \triangle(y^k, y)$$



LP-релаксация:

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} F(x^k, \mu) + \triangle(y^k, \mu)$$

Релаксация функции потерь

$$l(x^k, y^k, F) = -F(x^k, y^k) + max_{y \in \mathcal{Y}} F(x^k, y) + \triangle(y^k, y)$$

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} F(x^k, \mu) + \triangle(y^k, \mu)$$

Релаксация функции потерь

$$l(x^k, y^k, F) = -F(x^k, y^k) + max_{y \in \mathcal{Y}} F(x^k, y) + \triangle(y^k, y)$$

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} F(x^k, \mu) + \triangle(y^k, \mu)$$

- $\mu = \{\mu_\alpha(y_\alpha)\}$

Релаксация функции потерь

$$l(x^k, y^k, F) = -F(x^k, y^k) + max_{y \in \mathcal{Y}} F(x^k, y) + \triangle(y^k, y)$$

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} F(x^k, \mu) + \triangle(y^k, \mu)$$

- $\mu = \{\mu_\alpha(y_\alpha)\}$
- $\mathcal{M} = \{\mu | \mu_\alpha(y_\alpha) \in [0, 1], \sum_{y_\alpha} \mu_\alpha(y_\alpha) = 1, \mu_{\alpha\beta}(y_\beta) = \mu_\beta(y_\beta)\}$ $\quad \mu_{\alpha\beta}(y_\beta) = \sum_{y_{\alpha \setminus \beta}} \mu_\alpha(y_\alpha)$

Релаксация функции потерь

$$l(x^k, y^k, F) = -F(x^k, y^k) + max_{y \in \mathcal{Y}} F(x^k, y) + \triangle(y^k, y)$$

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} F(x^k, \mu) + \triangle(y^k, \mu)$$

- $\mu = \{\mu_\alpha(y_\alpha)\}$
- $\mathcal{M} = \{\mu | \mu_\alpha(y_\alpha) \in [0,1], \sum_{y_\alpha} \mu_\alpha(y_\alpha) = 1, \mu_{\alpha\beta}(y_\beta) = \mu_\beta(y_\beta)\}$ $\quad \mu_{\alpha\beta}(y_\beta) = \sum_{y_{\alpha \setminus \beta}} \mu_\alpha(y_\alpha)$
- $F(x^k, \mu) = \sum_\alpha \sum_{y_\alpha} f(x^k, y_\alpha)\mu_\alpha,$
  $\triangle(y^k, \mu) = \sum_\alpha \sum_{y_\alpha} \triangle_\alpha(y^k_\alpha, y_\alpha)\mu(y_\alpha)$

Изменение нотации, добавление энтропии

Функцию потерь:

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} F(x^k, \mu) + \triangle(y^k, \mu)$$

Изменение нотации, добавление энтропии

Функцию потерь:

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} F(x^k, \mu) + \triangle(y^k, \mu)$$

Можно переписать:

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} \theta_F^k \cdot \mu$$

$$\theta_F^k = f_\alpha(x^k, y_\alpha) + \triangle_\alpha(y_\alpha^k, y_\alpha)$$

Изменение нотации, добавление энтропии

Функцию потерь:

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} F(x^k, \mu) + \triangle(y^k, \mu)$$

Можно переписать:

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} \theta_F^k \cdot \mu$$

$$\theta_F^k = f_\alpha(x^k, y_\alpha) + \triangle_\alpha(y_\alpha^k, y_\alpha)$$

Добавим энтропию для сглаживания функционала

$$l_1(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}}(\theta_F^k \cdot \mu + \epsilon \sum_\alpha H(\mu_\alpha))$$

Изменение нотации, добавление энтропии

Функцию потерь:

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} F(x^k, \mu) + \triangle(y^k, \mu)$$

Можно переписать:

$$l_0(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} \theta_F^k \cdot \mu$$

$$\theta_F^k = f_\alpha(x^k, y_\alpha) + \triangle_\alpha(y_\alpha^k, y_\alpha)$$

Добавим энтропию для сглаживания функционала

$$l_1(x^k, y^k, F) = -F(x^k, y^k) + max_{\mu \in \mathcal{M}} (\theta_F^k \cdot \mu + \epsilon \sum_\alpha H(\mu_\alpha))$$

Оценка для $l_1(x^k, y^k, F), H_{max} = \sum_\alpha \log(|y_\alpha|)$:

$$l_0(x^k, y^k, F) \leqslant l_1(x^k, y^k, F) \leqslant l_0(x^k, y^k, F) + \epsilon H_{max}$$

Table of Contents

Задача минимизации

$$\min_F R(F) = \min_F \sum_k \left[-F(x^k, y^k) + A(\theta_F^k) \right]$$

$$A(\theta_F^k) = max_{\mu \in \mathcal{M}} (\theta_F^k \cdot \mu + \epsilon \sum_\alpha H(\mu_\alpha))$$

Задача минимизации

$$\min_F R(F) = \min_F \sum_k \left[-F(x^k, y^k) + A(\theta_F^k) \right]$$

$$A(\theta_F^k) = max_{\mu \in \mathcal{M}} (\theta_F^k \cdot \mu + \epsilon \sum_\alpha H(\mu_\alpha))$$

Седловая точка!

Задача минимизации

$$\min_F R(F) = \min_F \sum_k \left[-F(x^k, y^k) + A(\theta_F^k) \right]$$

$$A(\theta_F^k) = max_{\mu \in \mathcal{M}} (\theta_F^k \cdot \mu + \epsilon \sum_\alpha H(\mu_\alpha))$$

Седловая точка! Но что, если мы сможем получить $A(\theta) = min_\lambda (A(\lambda, \theta))$.
Получаем задачу:

$$\min_F \min_{\{\lambda^k\}} \sum_k \left[-F(x^k, y^k) + A(\lambda^k, \theta_F^k) \right]$$

Двойственная задача

$$A(\theta_F^k) = max_{\mu \in \mathcal{M}}(\theta_F^k \cdot \mu + \epsilon \sum_{\alpha} H(\mu_\alpha))$$

$$A(\theta) = min_{\lambda}(A(\lambda, \theta))$$

$$A(\lambda, \theta) = \max_{\mu \in \mathcal{N}} \theta_F^k \cdot \mu + \epsilon \sum_{\alpha} H(\mu_\alpha) + \\ + \sum_{\alpha} \sum_{\beta \subset \alpha} \sum_{y_\beta} \lambda_\alpha(y_\beta)(\mu_{\alpha\beta}(y_\beta) - \mu_\beta(y_\beta))$$

$$\mathcal{N} = \{\mu | \sum_{y_\alpha} \mu_\alpha(y_\alpha) = 1, \mu_\alpha(y_\alpha) \geqslant 0\}$$

двойственная задача

$$A(\lambda, \theta) = \max_{\mu \in \mathcal{N}} \theta_F^k \cdot \mu + \epsilon \sum_{\alpha} H(\mu_\alpha) +$$
$$+ \sum_{\alpha} \sum_{\beta \subset \alpha} \sum_{y_\beta} \lambda_\alpha(y_\beta)(\mu_{\alpha\beta}(y_\beta) - \mu_\beta(y_\beta))$$

Для фиксированных λ μ, которых достигается макимум:

$$\mu_\alpha(y_\alpha) = \frac{1}{Z_\alpha} exp \left(\frac{1}{\epsilon} \left(\theta(y_\alpha) + \sum_{\beta \subset \alpha} \lambda_\alpha(y_\beta) - \sum_{\gamma \supset \alpha} \lambda_\gamma(y_\alpha) \right) \right)$$

Блочно-координатный спуск для $\min_\lambda A(\lambda, \theta)$:

$$\lambda'_\alpha(y_\nu) = \lambda_\alpha(y_\nu) + \frac{\epsilon}{1 + N_\nu} (\log \mu_\nu(y_n u) + \sum_{\alpha' \supset \nu} \log \mu_{\alpha'}(y_\nu)) -$$
$$- \epsilon \log \mu_\alpha(y_\nu)$$

Table of Contents

Логистическая регрессия

Дана выборка $\{(x^k, y^k)\}$

$$\max_W \sum_k \left[(Wx^k)_{y^k} - log \sum_y exp(Wx^k)_y \right]$$

$$x \in \mathbb{R}^M$$

$$y \in \{1, \ldots, L\}$$

Логистическая регрессия

Дана выборка $\{(x^k, y^k)\}$

$$\max_W \sum_k \left[(Wx^k)_{y^k} - log \sum_y exp(Wx^k)_y \right]$$

$$x \in \mathbb{R}^M$$

$$y \in \{1, \dots, L\}$$

$$\max_{f \in \mathcal{F}} \sum_k \left[(f(x^k, y^k) - log \sum_y exp(f(x^k, y)) \right]$$

Логистическая регрессия

Дана выборка $\{(x^k, y^k)\}$

$$\max_W \sum_k \left[(Wx^k)_{y^k} - log \sum_y exp(Wx^k)_y \right]$$

$$x \in \mathbb{R}^M$$

$$y \in \{1, \ldots, L\}$$

$$\max_{f \in \mathcal{F}} \sum_k \left[(f(x^k, y^k) - log \sum_y exp(f(x^k, y)) \right]$$

$$\max_{f \in \mathcal{F}} \sum_k \left[(f(x^k, y^k) + b^k(y^k) - log \sum_y exp(f(x^k, y) + b^k(y)) \right]$$

Сведение к логистической регрессии

$$f_\alpha^* = \arg\min_{f_\alpha \in \mathcal{F}_\alpha} \sum_k \left[-F(x^k, y^k) + A(\lambda^k, \theta_F^k) \right], \quad F(x,y) = \sum_\alpha f_\alpha(x, y_\alpha).$$

Эквивалентно

$$f_\alpha^* = \epsilon \arg\max_{f_\alpha \in \mathcal{F}_\alpha} \sum_k \Bigg[\left(f_\alpha(x^k, y_\alpha^k) + b_\alpha^k(y_\alpha^k) \right) - \log \sum_{y_\alpha} \exp\left(f_\alpha(x^k, y_\alpha) + b_\alpha^k(y_\alpha) \right) \Bigg]$$

$$b_\alpha^k(y_\alpha) = \frac{1}{\epsilon} \left(\Delta(y_\alpha^k, y_\alpha) + \sum_{\beta \subset \alpha} \lambda_\alpha(y_\beta) - \sum_{\gamma \supset \alpha} \lambda_\gamma(y_\alpha) \right)$$

Сведение к логистической регрессии

Theorem

$$\max_{x:1^T x=1, x\geqslant 0} \theta \cdot x - p \sum_i x_i \log x_i = p \log \sum_i exp\left(\frac{\theta_i}{p}\right)$$

$$A(\lambda^k, \theta_F^k) = \max_{\mu \in \mathcal{N}} \sum_\alpha \sum_{y_\alpha} \left(f_\alpha(x^k, y_\alpha) + \Delta_\alpha(y_\alpha^k, y_\alpha)\right) \mu(y_\alpha) + \epsilon \sum_\alpha H(\mu_\alpha)$$
$$+ \sum_\alpha \sum_{\beta \subset \alpha} \sum_{x_\beta} \lambda_\alpha^k(x_\beta) \left(\mu_{\alpha\beta}(y_\beta) - \mu_\beta(y_\beta)\right).$$

$$A(\lambda^k, \theta_F^k) = \sum_\alpha \max_{\mu_\alpha \in \mathcal{N}_\alpha} \left(\sum_{y_\alpha} \left(f_\alpha(x, y_\alpha) + \epsilon b_\alpha(y_\alpha)\right) \mu_\alpha(y_\alpha) + \epsilon H(\mu_\alpha)\right),$$

$$A(\lambda^k, \theta_F^k) = \sum_\alpha \epsilon \log \sum_{y_\alpha} \exp\left(\frac{1}{\epsilon} f_\alpha(x, y_\alpha) + b_\alpha(y_\alpha)\right).$$

Алгоритм

- Задаем фиксируем b_α^k

$$b_\alpha^k(y_\alpha) \leftarrow \frac{1}{\epsilon}\left(\Delta(y_\alpha^k, y_\alpha) + \sum_{\beta\subset\alpha} \lambda_\alpha^k(y_\beta) - \sum_{\gamma\supset\alpha} \lambda_\gamma^k(y_\alpha)\right).$$

- Находим F_α при фиксированных λ^k

$$f_\alpha \leftarrow \arg\max_{f_\alpha\in\mathcal{F}_\alpha} \sum_{k=1}^{K}\left[\left(f_\alpha(x^k, y_\alpha^k) + b_\alpha^k(y_\alpha^k)\right) - \log\sum_{y_\alpha}\exp\left(f_\alpha(x^k, y_\alpha) + b_\alpha^k(y_\alpha)\right)\right]$$

- Меняем параметры

$$\theta^k(y_\alpha) \leftarrow \epsilon f_\alpha(x^k, y_\alpha) + \Delta(y_\alpha^k, y_\alpha).$$

- Для фиксированных f_α ищем λ^k

$$\lambda'_\alpha(y_\nu) = \lambda_\alpha(y_\nu) + \frac{\epsilon}{1+N_\nu}(\log\mu_\nu(y_n u) + \sum_{\alpha'\supset\nu}\log\mu_{\alpha'}(y_\nu)) - \\ -\epsilon\log\mu_\alpha(y_\nu)$$

Table of Contents

Эксперименты

Five sets of functions $\mathcal{F}_\alpha$:

- Zero
- Constant (ignores x)
- Linear
- Single Hidden-Layer MLP
- Boosted Decision Trees

Эксперименты

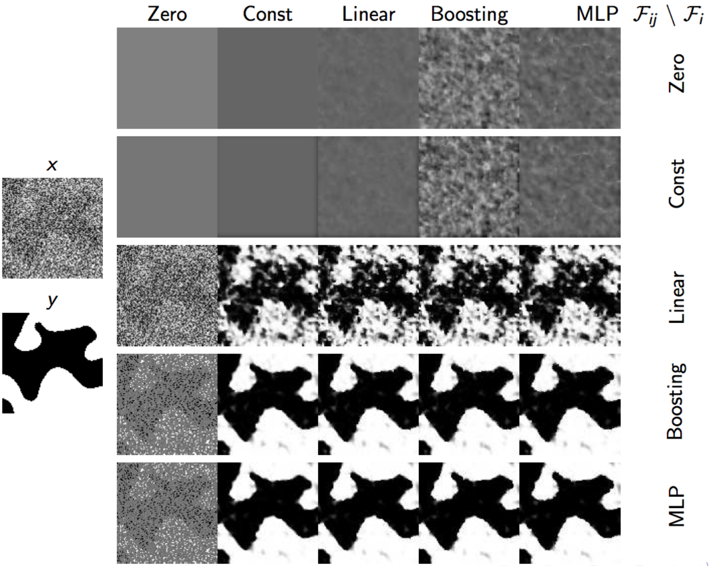
$f_{\{i\}}$ - univariate features ϕ_i
$f_{\{i,j\}}$ - pairwise features ϕ_{ij}

x y

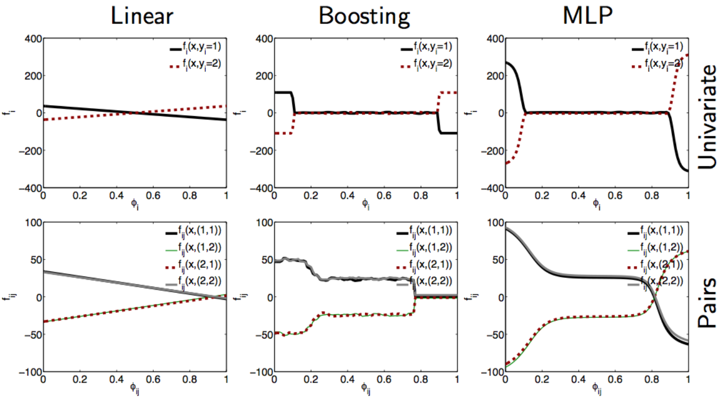
y - convolve gaussian noise with a gaussian and threshold.

- For $\{\alpha\} = \{i\}$, if $y_i = 0$, $\phi_i \in [0, .9]$, else $\phi_i \in [.1, 1]$.
- For $\{\alpha\} = \{i, j\}$, if $y_i = y_j$, $\phi_{ij} \in [0, .8]$, else $\phi_{ij} \in [.2, 1]$

Эксперименты



Эксперименты



Эксперименты

Denoising - Train

$\mathcal{F}_i \setminus \mathcal{F}_{ij}$	Zero	Const.	Linear	Boost.	MLP
Zero	.490	.490	.490	.465	.490
Const.	.490	.490	.490	.465	.490
Linear	.443	.077	.059	.056	.033
Boost.	.429	.032	.014	.012	.008
MLP	.435	.031	.014	.011	.008

Denoising - Test

$\mathcal{F}_i \setminus \mathcal{F}_{ij}$	Zero	Const.	Linear	Boost.	MLP
Zero	.502	.502	.502	.502	.502
Const.	.502	.502	.502	.504	.502
Linear	.444	.077	.059	.057	.034
Boost.	.445	.033	.015	.013	.008
MLP	.445	.032	.015	.013	.008

Эксперименты

$f_{\{i\}}$ - univariate features ϕ_i
$f_{\{i,j\}}$ - pairwise features ϕ_{ij}

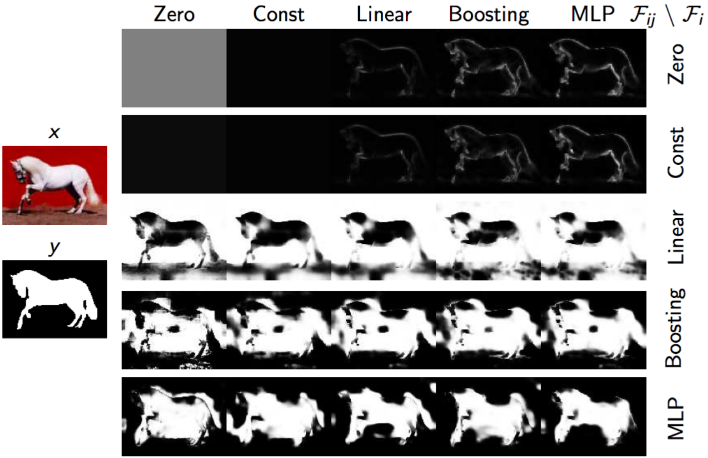
x

y

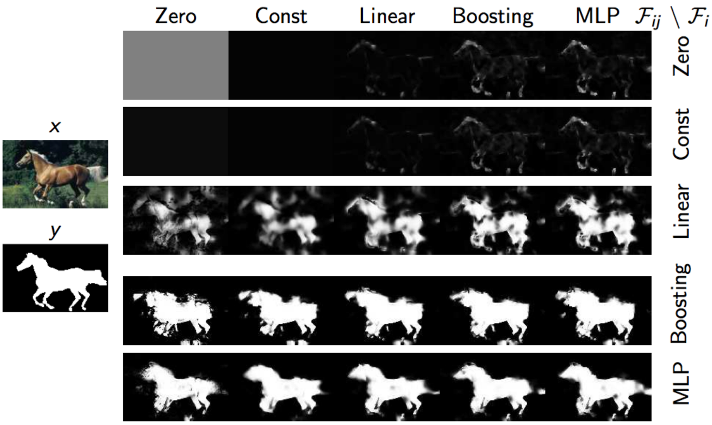
y - horse or not?

- For $\{\alpha\} = \{i\}$, (1) RBG values (2), vertical and horizontal position (3) histogram of gradients.
- For $\{\alpha\} = \{i, j\}$, (1) l_2 distance of RGB for pixels i and j (2) Sobel edge filter

Эксперименты



Эксперименты



Эксперименты

Horses - Train

$\mathcal{F}_i \setminus \mathcal{F}_{ij}$	Zero	Const.	Linear	Boost.	MLP
Zero	.211	.211	.212	.211	.210
Const.	.211	.211	.212	.211	.210
Linear	.141	.139	.126	.111	.113
Boost.	.087	.079	.074	.069	.068
MLP	.054	.051	.046	.043	.041

Horses - Test

$\mathcal{F}_i \setminus \mathcal{F}_{ij}$	Zero	Const.	Linear	Boost.	MLP
Zero	.246	.246	.247	.246	.245
Const.	.246	.246	.247	.246	.245
Linear	.185	.185	.168	.158	.156
Boost.	.115	.107	.100	.095	.094
MLP	.096	.094	.087	.085	.081