Ozon Masters ML1
Теор. задание 2.

Арешин Станислав Олегович.

стр 1

Задание 1

а) $f(x) = \log(1 + \|Ax\|^2)$

$$df(x) = \frac{d(1+\|Ax\|^2)}{1+\|Ax\|^2} = \frac{d(\langle Ax, Ax\rangle)}{1+\|Ax\|^2}$$

Найдём дифференциал $d(\langle Ax, Ax\rangle)$:

$$d(\langle Ax, Ax\rangle) = \langle dAx, Ax\rangle + \langle Ax, dAx\rangle =$$

$$= 2\langle Ax, dAx\rangle = 2\langle A^T A x, dx\rangle$$

Таким образом, $df(x) = \dfrac{2\langle A^T A x, dx\rangle}{1+\|Ax\|^2} =$

$$= \left\langle \frac{2A^T A x}{1+\|Ax\|^2}, dx \right\rangle$$

$$\Rightarrow \nabla f(x) = \frac{2A^T A x}{1+\|Ax\|^2}$$

в) $f(x) = \dfrac{-1}{1+x^T A x} = -1\,(1+x^T A x)^{-1}$

$$df(x) = d(1+x^T A x)\cdot(1+x^T A x)^{-2} = \frac{d(\langle x, Ax\rangle)}{(1+x^T A x)^2} =$$

$$= \frac{d(\langle Ax, x\rangle)}{(1+x^T A x)^2}$$

По св-ву $d(\langle Ax, x\rangle) = \langle (A+A^T)x, dx\rangle$

Таким образом, $df(x) = \dfrac{\langle (A+A^T)x, dx\rangle}{(1+x^T A x)^2} =$

$$= \left\langle \frac{(A+A^T)x}{(1+x^T A x)^2}, dx \right\rangle$$

$$\Rightarrow \nabla f(x) = \frac{(A+A^T)x}{(1+x^T A x)^2}$$

см. след лист №2.

№1 (продолжение)

c) $f(x) = x^T A x^2$, где $x^2 = (x_1^2 \dots x_n^2)^T$

$f(x) = x^T A x^2 = \langle x, A x^2 \rangle$

$df(x) = d(\langle x, A x^2 \rangle) = \langle dx, A x^2 \rangle + $
$+ \langle x, dA x^2 \rangle = \langle A x^2, dx \rangle + \langle A^T x, dx^2 \rangle.$

Найдём дифференциал dx^2:

Пусть $y(x) = x^2$, $x = (x_1 \dots x_n)^T$

$$\frac{\partial y}{\partial x} = \begin{pmatrix} \frac{x_1^2}{\partial x_1} & \frac{x_1^2}{\partial x_2} & \dots & \frac{x_1^2}{\partial x_n} \\ \frac{x_2^2}{\partial x_1} & \frac{x_2^2}{\partial x_2} & \dots & \frac{x_2^2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_n^2}{\partial x_1} & \frac{x_n^2}{\partial x_2} & \dots & \frac{x_n^2}{\partial x_n} \end{pmatrix} =$$

$$= \begin{pmatrix} 2x_1 & 0 & \dots & 0 \\ 0 & 2x_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2x_n \end{pmatrix} = 2\,\mathrm{diag}(x)$$

Таким образом, $df(x) = \langle A x^2, dx \rangle +$
$+ \langle A^T x, 2\,\mathrm{diag}(x)\, dx \rangle = \langle A x^2, dx \rangle + \langle 2\,\mathrm{diag}(x) A^T x, dx \rangle$
$+ \langle 2\,\mathrm{diag}(x) A^T x, dx \rangle = \langle A x^2 + 2\,\mathrm{diag}(x) A^T x, dx \rangle$

$\Rightarrow \nabla f(x) = A x^2 + 2\,\mathrm{diag}(x) A^T x$

Ozon Masters ML1 Аришин Станислав Олегович

Теор. задание 2.

Задание №2

$$f(x) = \sin(x)\cos(2x) \quad x \in \left[-\frac{3\pi}{2} - a;\ \frac{3\pi}{2} + a\right]$$

$$a = \arccos\sqrt{2/3}$$

1) Найдём экстремум функции $f(x)$

$$f'(x) = \cos x \cos 2x - 2\sin x \sin 2x = 0$$

$$\left[\begin{array}{l} \cos 2x = 2\cos^2 x - 1 = 1 - 2\sin^2 x \\ \sin 2x = 2\sin x \cos x \end{array}\right]$$

$$\cos x\,(1 - \sin^2 x) - 4\sin^2 x \cos x = 0$$

$$\cos x - 6\sin^2 x \cos x = 0$$

$$\cos x\,(1 - 6\sin^2 x) = 0$$

$\cos x = 0$ или $1 - 6\sin^2 x = 0$

$x_1 = \pm\frac{\pi}{2} + 2\pi n$, $n \in \mathbb{Z}$

$6\sin^2 x = 1$

$\sin^2 x = 1/6$

$\sin x = \pm\sqrt{1/6}$

$x_2 = \pm\arcsin(1/\sqrt{6}) + \pi n$, $n \in \mathbb{Z}$

2) Отберём нужные экстремумы по ООФ:

1. $x = \frac{\pi}{2} + 2\pi n$:

$n = 0$: $x_1 = \frac{\pi}{2}$; $f(x_1) = -1$

$n = 1$: за пределами ООФ

$n = -1$: $x_2 = -\frac{3\pi}{2}$; $f(x_2) = -1$

$n = -2$: за пределами ООФ

2. $x = -\frac{\pi}{2} + 2\pi n$.

$n = 0$: $x_3 = -\frac{\pi}{2}$; $f(x_3) = 1$

$n = 1$: $x_4 = \frac{3\pi}{2}$; $f(x_3) = 1$

$n = 2$: за пределами ООФ

$n = -1$: за пределами ООФ

3. $x = \arcsin(1/\sqrt{6}) + \pi n$

$n = 0$: $x_5 = \arcsin(1/\sqrt{6})$; $f(x_5) = \frac{\sqrt{2/3}}{3}$

$n = 1$: $x_6 = \arcsin(1/\sqrt{6}) + \pi$; $f(x_6) = -\frac{\sqrt{2/3}}{3}$

$n = 2$: за пределами ООФ

$n = -1$: $x_7 = \arcsin(1/\sqrt{6}) - \pi$; $f(x_7) = -\frac{\sqrt{2/3}}{3}$

см. след лист №2.

4. $x = -\arcsin(1/\sqrt{6}) + \pi n$

$n=0$: $x_8 = -\arcsin(1/\sqrt{6})$; $f(x_8) = -\frac{\sqrt{2/3}}{3}$

$n=1$: $x_9 = -\arcsin(1/\sqrt{6}) + \pi$; $f(x_9) = \frac{\sqrt{2/3}}{3}$

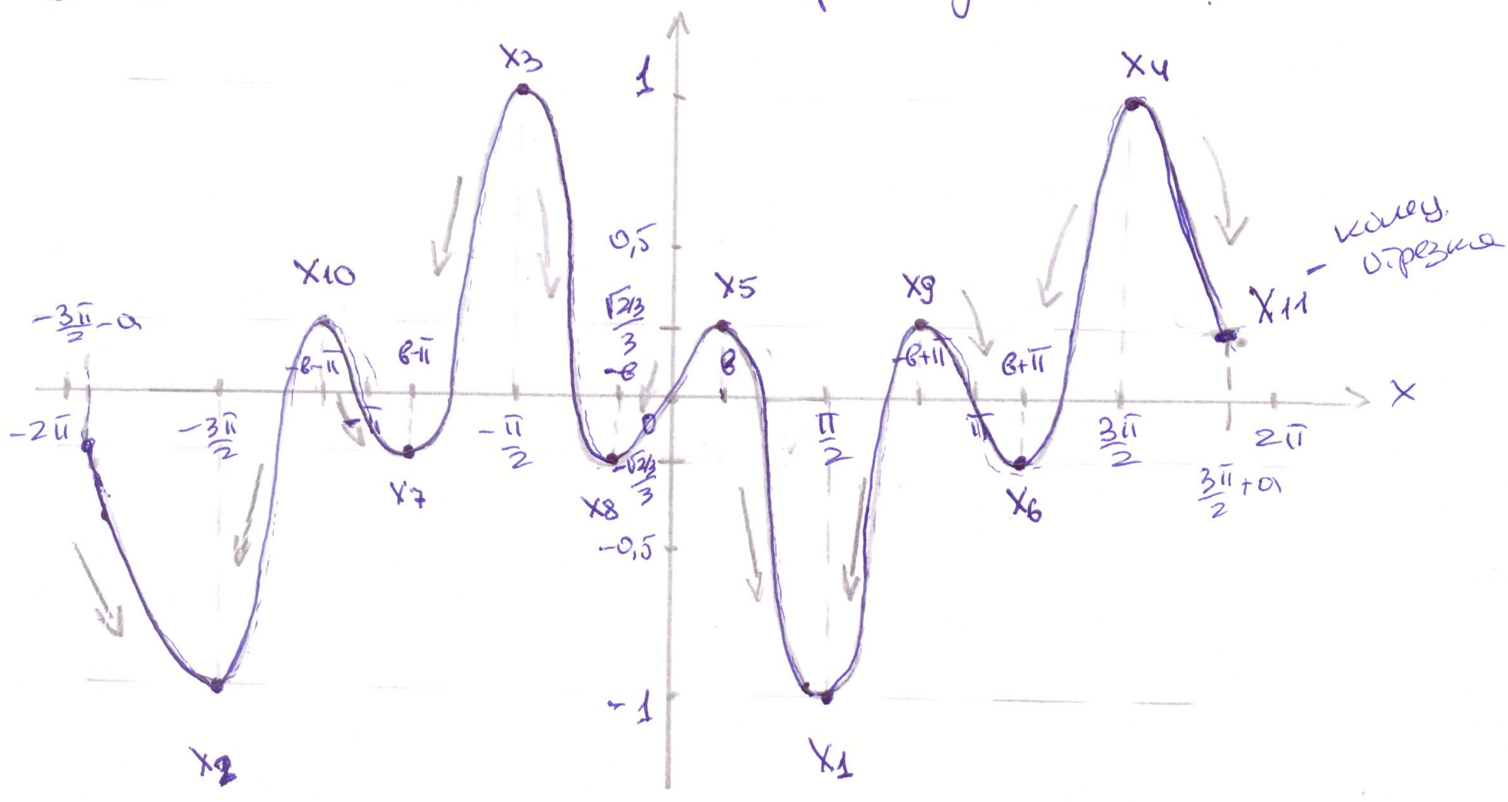
$n=2$: за пределами ООФ

$n=-1$: $x_{10} = -\arcsin(1/\sqrt{6}) - \pi$, $f(x_{10}) = \frac{\sqrt{2/3}}{3}$

$n=-2$: за пределами ООФ.

Таким образом, отобрано 10 экстремумов
Точки глобального минимума x_1, x_2.

3) Построим график функции с отмеченными экстремумами:



$b = \arcsin(1/\sqrt{6})$

правый край: , $x = \frac{3\pi}{2} + a$; $f(x) = \frac{\sqrt{2/3}}{3}$

левый край: , $x = -\frac{3\pi}{2} - a$; $f(x) = -\frac{\sqrt{2/3}}{3}$

см. следующий №3

4) Так как шаг градиентного спуска мал, алгоритм будет останавливаться в разных минимумах в зависимости от точки старта. Точка старта выбирается равномерно: $x_{start} \sim R\left(-\frac{3\pi}{2}-a;\ \frac{3\pi}{2}+a\right)$

Например, чтобы попасть в точку x_2, надо начать спуск в любой точке из интервала $\left(-\frac{3\pi}{2}-a;\ -b-\pi\right)$ и аналогично для других минимумов (см. рисунок.)

Вычислим эти вероятности с помощью функции распред. равномерного распределения

$$F(x) = \frac{x - x_{left}}{x_{right} - x_{left}};\quad x_{left} = -\frac{3\pi}{2}-a;\quad x_{right} = \frac{3\pi}{2}+a.$$

для x_2: $P_2 = P\left(-\frac{3\pi}{2}-a \le x \le -b-\pi\right) =$

$= F\left(-\arcsin(1/\sqrt{6}) - \pi\right) - F\left(-\frac{3\pi}{2} - \arccos(\sqrt{2/3})\right) \approx$

$\approx 0,166 - 0 = 0,166$

для x_7: $P_7 = P\left(-b-\pi \le x \le -\frac{\pi}{2}\right) = F\left(-\frac{\pi}{2}\right) - F\left(-\arcsin(1/\sqrt{6}) - \pi\right) \approx$

$\approx 0,353 - 0,166 = 0,187$

для x_8: $P_8 = P\left(-\frac{\pi}{2} \le x \le b\right) = F\left(\arcsin(1/\sqrt{6})\right) - F\left(-\frac{\pi}{2}\right) \approx$

$\approx 0,539 - 0,353 = 0,186$

для x_1: $P_1 = P\left(b \le x < -b+\pi\right) = F\left(-\arcsin(1/\sqrt{6}) + \pi\right) - F\left(\arcsin(1/\sqrt{6})\right)$

$\approx 0,755 - 0,539 = 0,216$

см. след. лист №4

№2 (продолжение) $\boxed{\text{стр 4}}$

для x_6: $p_6 = P\left(-в+\pi \le x < \frac{3\pi}{2}\right) = F\left(\frac{3\pi}{2}\right) - F\left(-\arcsin(1/\sqrt{6})+\pi\right)$

$\approx 0,942 - 0,755 = 0,187$

для конца отрезка x_{11}: $p_{11} = P\left(\frac{3\pi}{2} \le x < \frac{3\pi}{2}+a\right) =$

$= F\left(\frac{3\pi}{2} + \arccos\left(\sqrt{2/3}\right)\right) - F\left(\frac{3\pi}{2}\right) = 1 - 0,942 = 0,058$

Таким образом:

Ответ: 1. Алгоритм будет останавливаться

вблизи точек: $x_2 = -\frac{3\pi}{2}$; $f(x_2) = -1$

$x_7 = \arcsin(1/\sqrt{6}) - \pi$, $f(x_7) = -\frac{\sqrt{2/3}}{3}$

$x_8 = -\arcsin(1/\sqrt{6})$, $f(x_8) = -\frac{\sqrt{2/3}}{3}$

$x_1 = \frac{\pi}{2}$, $f(x_1) = 1$

$x_6 = \arcsin(1/\sqrt{6}) + \pi$, $f(x_6) = -\frac{\sqrt{2/3}}{3}$

и конец отрезка $x_{11} = \frac{3\pi}{2} + \arccos\left(\sqrt{2/3}\right)$; $f(x_{11}) = \frac{\sqrt{2/3}}{3}$

2. Доля случаев, которая сойдётся к

каждой из точек:

$$\left.\begin{array}{lll}
\text{глобальный min} & x_2: & p_2 = 0,166 \\
\text{локальный min} & x_7: & p_7 = 0,187 \\
\text{локальный min} & x_8: & p_8 = 0,186 \\
\text{глобальный min} & x_1: & p_1 = 0,216 \\
\text{локальный min} & x_6: & p_6 = 0,187 \\
\text{конец отрезка} & x_{11}: & p_{11} = 0,058
\end{array}\right\} \Sigma = 1$$

Чаще всего алгоритм будет равновероятно (практически из-за округления) попадать в локальные минимумы x_7, x_8, x_6.

Ozon Masters ML1 — Арешин Станислав Олегович

Теор. задание 2

стр 1

Задание №3

a) $f(x) = \max(0, 1 - ax)$, $a = const$.

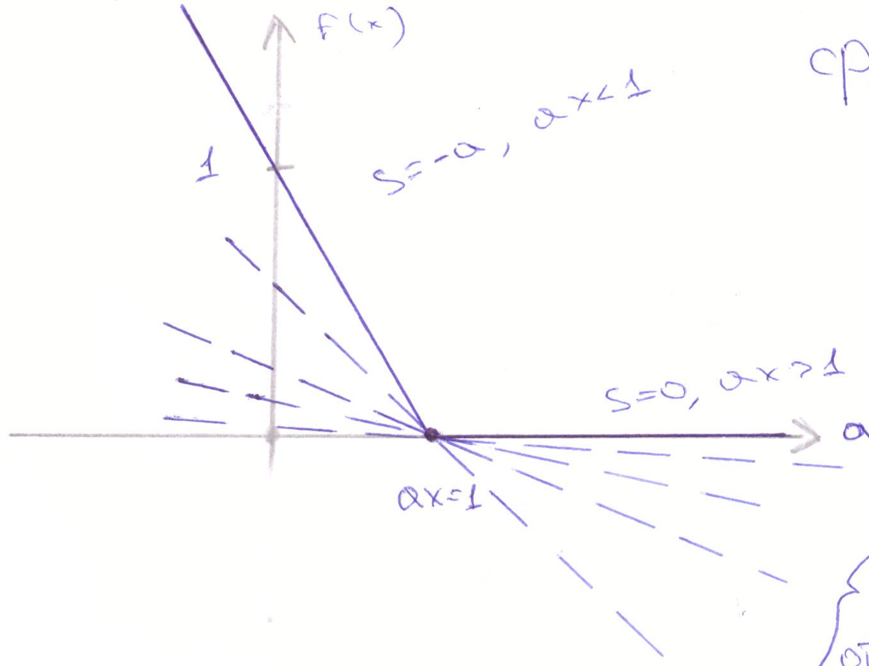
Похоже на Hinge Loss

$$\partial f(x) = \begin{cases} \frac{\partial(1-ax)}{\partial x}, & 1 - ax > 0 \\ ?, & 1 - ax = 0 \\ \{0\}, & 1 - ax < 0 \end{cases}$$

$$\partial f(x) = \begin{cases} \{-a\}, & ax < 1 \\ ?, & ax = 1 \\ \{0\}, & ax > 1. \end{cases}$$

Построим график: Пусть s - наклон прямой

Функция выпуклая.



субградиенты от $s=-a$ до $s=0$

Тогда в точке $ax = 1$ субдифференциал функции это все касательные с наклоном $[-a; 0]$

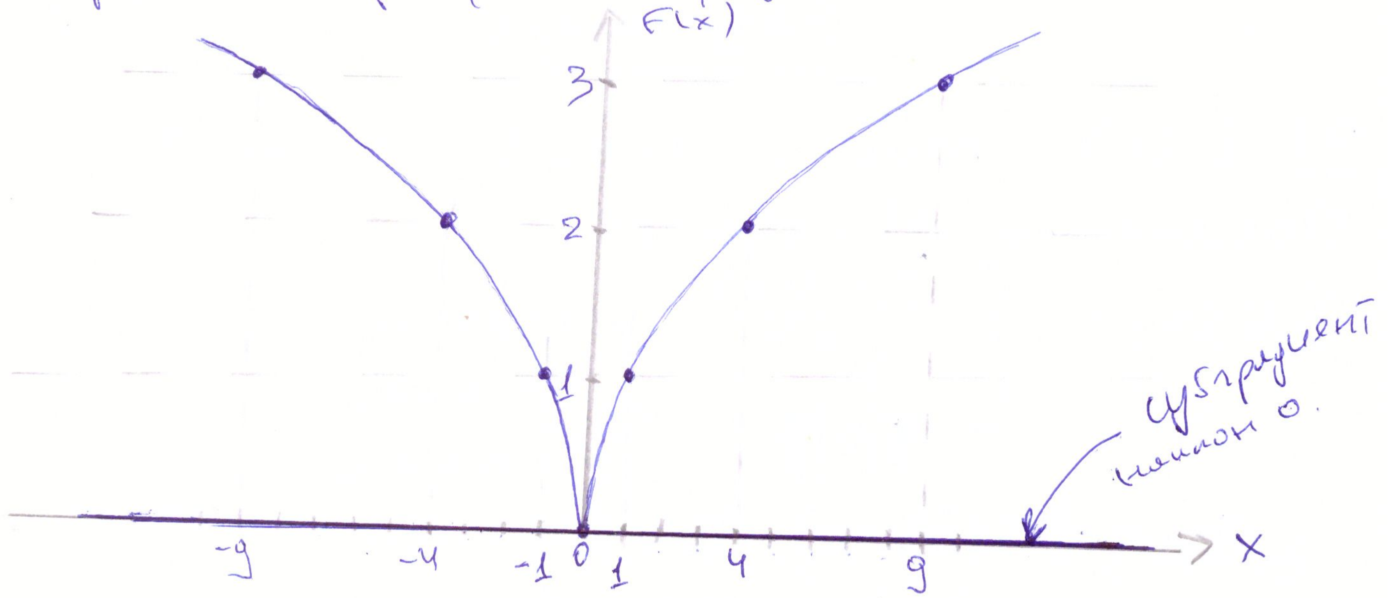
Таким образом,

$$\partial f(x) = \begin{cases} \{-a\}, & ax < 1 \\ [-a; 0], & ax = 1 \\ \{0\}, & ax > 1 \end{cases}$$

см. след лист №2

в) $f(x)=\sqrt{|x|}$

Построим график функции:



Функция $f(x)=\sqrt{|x|}$ не выпуклая. Единственный субградиент, существующий для этой функции в точке $x=0$ это 0 $\Rightarrow$ субдифференциал функции $f(x)$ в точке $x=0$ будет $\{0\}$

Таким образом,

$$\partial f(x)=\begin{cases}\dfrac{1}{2\sqrt{x}}, & x>0\\ \{0\}, & x=0\\ \dfrac{-1}{2\sqrt{-x}}, & x<0\end{cases}$$

Ozon Masters ML 1
Теор. задание 2.

Арешин Станислав Олегович

Задание №6

Док-во

1) Выпишем функцию правдоподобия:

$$p(x) = \frac{1}{1+e^{-z}}, \quad z = w^T x = w_0 + w_1 \underline{X}_1 + \dots + w_n \underline{X}_n$$

$$L(w_0, \dots, w_n) = \prod_{i: y_i = 1} p(x_i) \prod_{i: y_i = 0} (1 - p(x_i)) \to \max$$

$$\log L(w_0, \dots, w_n) = \log \prod_{i: y_i = 1} p(x_i) + \log \prod_{i: y_i = 0} (1 - p(x_i)) \to \max$$

$$\log L(w_0, \dots, w_n) = \sum_{i: y_i = 1} \log(p(x_i)) + \sum_{i: y_i = 0} \log(1 - p(x_i)) \to \max$$

Так как $1 - \frac{1}{1+e^{-z}} = \frac{1}{1+e^{z}}$

$$\log L(w_0, \dots, w_n) = \sum_{i: y_i = 1} \log\left(\frac{1}{1+e^{-z}}\right) + \sum_{i: y_i = 0} \log\left(\frac{1}{1+e^{z}}\right) \to \max$$

Пусть $y_i' = 2y_i - 1$

Тогда $\log L(w_0, \dots, w_n) = -\sum_i \log(1 + e^{-y_i' z_i}) \to \max$

2) Если w идеально разделяет выборку то, масштабируя w на константу, мы будем получать мн-во векторов, тоже идеально разделяющих выборку, то есть: $c(w_0 + w_1 \underline{X}_1 + \dots + w_n \underline{X}_n) = 0$,

для $\forall\, c > 1 \Rightarrow$ ~~если~~

если $w_0 + w_1 \underline{X}_1 + \dots w_n \underline{X}_n = 0$.

см. след лист №2

3) Подставим c в функцию правдоподобия: $cz_i = c(w_0 + w_1 X_1 + \dots + w_n X_n)$

$$\log L_c = -\sum_i \log(1+e^{-y_i' c z_i}) \to \max$$

Сравним две полученные функции правдоподобия $\log L$ и $\log L_c$:

$$e^{-y_i' z_i} > e^{-y_i' c z_i}$$

$$\Rightarrow \quad 1+e^{-y_i' z_i} > 1+e^{-y_i' c z_i}$$

$$\Rightarrow \quad \log(1+e^{-y_i' z_i}) > \log(1+e^{-y_i' c z_i})$$

$$\Rightarrow \quad \sum_i \log(1+e^{-y_i' z_i}) > \sum_i \log(1+e^{-y_i' c z_i})$$

$$\Rightarrow \quad -\sum_i \log(1+e^{-y_i' z_i}) < -\sum_i \log(1+e^{-y_i' c z_i})$$

$$\Rightarrow \quad \log L < \log L_c, \text{ для } \forall c > 1$$

То есть правдоподобие после увеличения вектора весов на константу увеличилось, следовательно, мы можем бесконечно увеличивать правдоподобие, масштабируя вектор весов на константу $c > 1$.

Таким образом, не существует вектора весов, который максимизировал бы правдоподобие ($\max \log L_c \to \infty$ при $c \to \infty$).

$cw = (cw_0, cw_1, \dots, cw_n)$

$c \to \infty \Rightarrow w \to \infty$

см. след лист №3

№6 (продолжение) стр 3

4) Проблема модели в том, что для минимизации функции правдоподобия достаточно устремить веса в бесконечность. Для борьбы с большими весами применяется регуляризация.

Если добавить регуляризацию в модель логистической регрессии, эта проблема будет решена