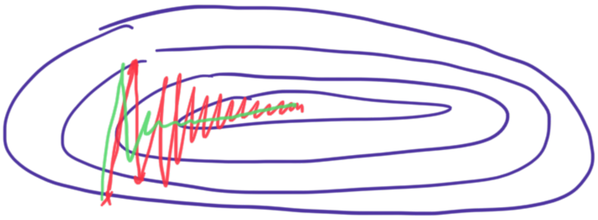
Модификации GD

оценки градиента

изменение формулы шага

$$w^{(k)} = w^{(k-1)} - \eta_k \underbrace{\nabla Q(w^{(k-1)})}_{\text{оценка вместо FG}}$$

①



Momentum (метод импульса)

$h_0 = 0$

$$h_k = \alpha h_{k-1} + \eta_k \underbrace{\nabla Q(w^{(k-1)})}_{\text{можно заменить на оценку}}$$

$\alpha \in (0, 1)$

$$w^{(k)} = w^{(k-1)} - h_k$$

② Пусть имеется разреженный признак x_j - почти на всех объектах $= 0$

$a(x) = \ldots + w_j x_j + \ldots$

$x_j = 0 \Rightarrow \frac{\partial Q}{\partial \ldots} = 0 \Rightarrow$ на большинстве ...

∂w_j → на большинстве итераций не будем менять w_j

(проблема только у стох. методов)

AdaGrad

как сильно мы сейчас изменим w_j

как сильно мы уже обучили w_j

$$G_{kj} = G_{k-1,j} + \left(\nabla_w Q(w^{(k-1)})\right)_j^2$$

$$w_j^{(k)} = w_j^{(k-1)} - \frac{\overset{=\text{const}}{\eta_k}}{\sqrt{G_{kj}+\varepsilon}}\left(\nabla_w Q(w^{(k-1)})\right)_j$$

G_{kj} только растет, убывание LR может оказаться слишком быстрым

RMSProp

Как AdaGrad, только

$$G_{kj} = \alpha G_{k-1,j} + (1-\alpha)\left(\nabla_w Q(w^{(k-1)})\right)_j^2$$

Adam

Разреженные модели

$$Q(w, X) + \alpha \|w\|_1 \to \min_w$$

$$\|w\|_1 = \sum |w_j|$$

Зачем разреживать модель?

- ускорение
- знаем, что не все призн. полезны
- $\ell < d$

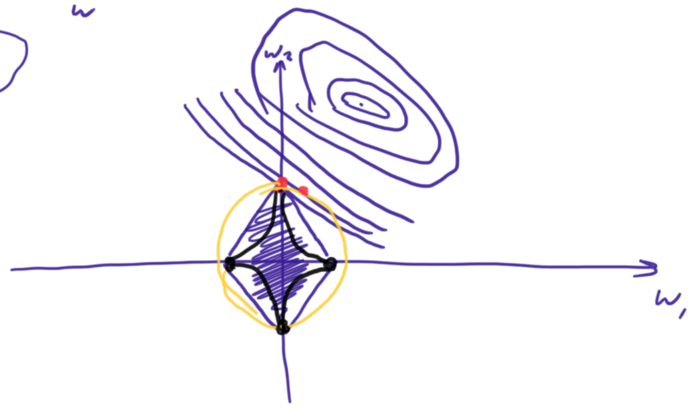
Почему L_1-рег. отбирает признаки?

① $Q(w, X) + \lambda \|w\|_1 \to \min_w$

$\Updownarrow$

$$\begin{cases} Q(w, X) \to \min_w \\ \|w\|_1 \le C \end{cases}$$

$L_{0,5}$



② $w = (1, \varepsilon)$ $\quad 0 < \delta < \varepsilon < 1$

$\|w - (\delta, 0)\|_2^2 = 1 - 2\delta + \delta^2 + \varepsilon^2 \leftarrow$

$\|w - (\delta, 0)\|_1 = 1 - \delta + \varepsilon$

$\|w - (0, \delta)\|_2^2 = 1 - 2\varepsilon\delta + \delta^2 + \varepsilon^2 \leftarrow$

$\|w - (0, \delta)\|_1 = 1 - \delta + \varepsilon$

⑤ молодец

③ Проксимальные методы

$$S_{\eta\alpha}(w_i) = \begin{cases} w_i - \eta\alpha, & w_i > \eta\alpha \\ 0, & |w_i| \le \eta\alpha \\ w_i + \eta\alpha, & w_i < -\eta\alpha \end{cases}$$

$$w^{(k)} = S_{\eta\alpha}\left(\underbrace{w^{(k-1)} - \eta \nabla Q(w^{(k-1)})}\right)$$

Линейная классификация

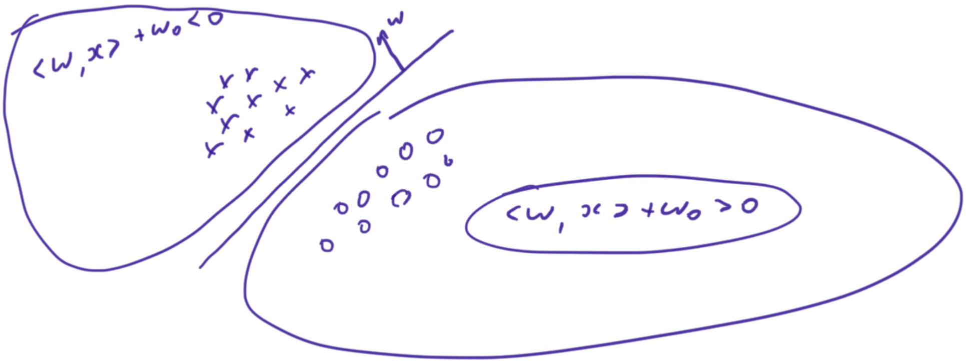
$\mathbb{X} = \mathbb{R}^d$

$X = \dots$ $\quad y = -1 \Rightarrow$ отриц. объект

$Y = \{-1, +1\}$ $\quad y = +1 \Rightarrow$ полож. объект.

$a(x) = \mathrm{sign}(\langle w, x\rangle + w_0)$ – линейный классификатор

$\langle w, x\rangle + w_0 = 0$ – гиперплоскость



① Обучение лин. класиф.

$$\underbrace{Q(a, X) = \frac{1}{l}\sum_{i=1}^{l}[a(x_i) \ne y_i]}_{\text{доля ошибок, error rate}} \to \min_a$$

$$\frac{1}{l}\sum_{i=1}^{l}[\underline{\mathrm{sign}(\langle w, x_i\rangle) \ne y_i}] \to \min_w$$

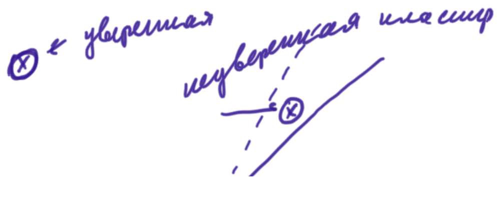
$$\frac{1}{l}\sum_{i=1}^{l}[\underline{y_i\langle w, x_i\rangle < 0}] \to \min$$

$y_i\langle w, x_i\rangle > 0 \Rightarrow$ ответ верный

$y_i\langle w, x_i\rangle < 0 \Rightarrow$ ответ неверный

$M_i = y_i\langle w, x_i\rangle$ – отступ (margin)

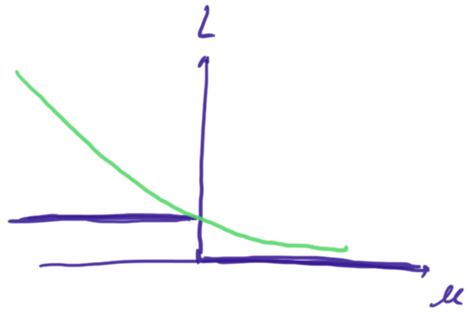
$|M_i|$ – расстояние от x_i до разделяющей гиперплоскости



$|M_i|$ – уверенность прогноза

$M_i \ll 0$ – обычно у выбросов

$$0 \le \frac{1}{\ell}\sum_{i=1}^{\ell}[y_i\langle w, x_i\rangle < 0] \le$$



$$L(M) = \underset{\text{пороговая ф-ция потерь}}{[M < 0]} \le \underbrace{\tilde{L}(M)}_{\text{гладкая}}$$

$$\le \frac{1}{\ell}\sum_{i=1}^{\ell}\tilde{L}(y_i\langle w, x_i\rangle) \to \min_w$$

$$\tilde{L}(M) = e^{-M}$$

$$\tilde{L}(M) = \frac{2}{1+e^{M}}$$

$\tilde{L}(M) = \log(1+e^{-M})$ – логистич. ф-ция потерь

$\tilde{L}(M) = \max(0, 1-M)$ – SVM

---

② <u>Метрики качества</u> классификации

1) Доля верных ответов

$$\text{accuracy}(a, X) = \frac{1}{\ell}\sum_{i=1}^{\ell}[a(x_i) = y_i]$$

НЕ ТОЧНОСТЬ!!!

$y = -1$: 950

$y = +1$: 50

$a(x) = -1$

accuracy: 0.95

- разные цены ошибки
- баланс классов может меняться

r_1	r_2	$r_1 - r_2$	$\frac{r_1 - r_2}{r_1}$
20%	10%	10%	50%
50%	25%	25%	50%
0.1%	0.01%	0.09%	90%