① Как выглядит бинарный линейный классификатор?

$$a(x) = \begin{cases} 1, & \text{если } f(x) > 0 \\ -1, & \text{если } f(x) \leq 0 \end{cases}$$

где $f(x) = w_0 + w_1 x_1 + \ldots + w_n x_n = w_0 + \langle w, x \rangle$,

где w_i - веса признаков x_i, $i = \overline{1, n}$

w_0 - порог

② Что такое отступ? Какие выводы можно сделать из знака отступа?

Отступом алгоритма $a(x) = \text{sign}\{f(x)\}$ на объекте x_i наз-ся $M_i = y_i f(x_i)$ (y_i - класс, к которому относится x_i)

$M_i \leq 0 \Leftrightarrow y_i \neq a(x_i)$ - алгоритм допустил ошибку

$M_i > 0 \Leftrightarrow y_i = a(x_i)$ - классифиц. алгоритм верно

③ Считают, что один из признаков x равен 1 и порог не записыв, а запис. как скалярное произв.

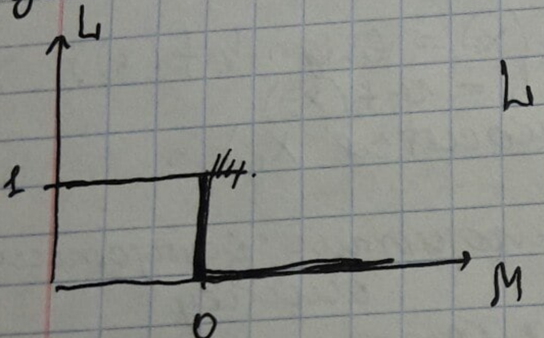
④ $$\tilde{Q}(w) = \sum_{i=1}^{\ell} L(M_i(w)) \longrightarrow \min_w$$

ф-ция эмпирического риска

ф-ция потерь

⑦ $L(M_i(w))$ – функция потерь – заменяют индикатор $I(M_i(x) < 0)$ – более удобные для дифф-ия или для применения каких-либо других методов оптимизации, не исп-щих градиента. Функция потерь в (.) 0 равна 1, после этого её зануляют (за искл-ем квадратичной ф-ции потерь)

⑧

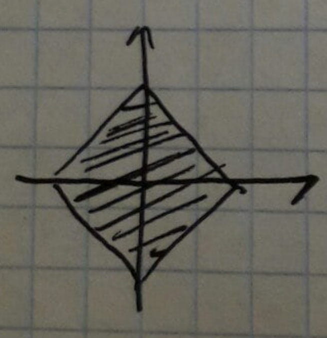


$L(M) = \max\{0, 1 - M\}$

⑨ Регуляризация – ограничение значений весов модели.

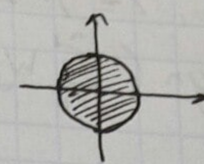
l_1 – регуляризация

$$\begin{cases} \tilde{Q} = \sum_{i=1}^{l} L(M_i) \to \min \\ \sum_{k=1}^{m} |w_k| \le \tau \end{cases}$$



l_2 - регуляризация

$$\begin{cases} \tilde{Q} = \sum_{i=1}^{\ell} L(M_i) \longrightarrow \min \\ \sum_{k=1}^{m} w_k^2 \leq \tau \end{cases}$$



(10) Если неограничивать степень полинома, веса должны быть достаточно большими, но тогда модель очень неустойчива, у неё есть резкие перебаты, т.е модель может быть очень переобучена. => Надо штрафовать модель за какие-то веса, например, оптимизируемую ф-л добавить сумму модулей весов или квадратов, т.е применить регуляр-ю l_1 или l_2, то степень полинома не будет очень большой и не будет переобучения.

Введение регуляризатора повышает устойчивость решения w. В случаях, когда минимум эмпирического риска достигается на мн-ве векторов w, регуляризация выбирает из них вектор с мин. нормой. Тем самым повышается устойчивость алгоритма, улучшается его обобщающая способность.

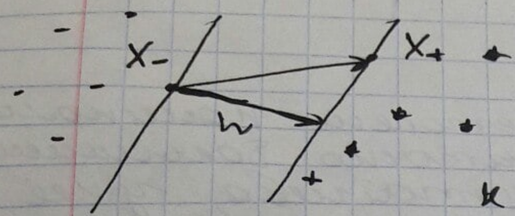
(3.2) Пусть $w \in \mathbb{R}^n$ имеет нормальное распред-ие, все его компоненты независ. и имеют равные дисперсии σ.

$$\ln p(w, \sigma) = \ln \left(\frac{1}{(2\pi\sigma)^{n/2}} \exp\left(- \frac{\|w\|^2}{2\sigma} \right) \right) = -\frac{1}{2\sigma}\|w\|^2 + const(w)$$

вероятностный смысл параметра регуляризации: он обратно пропорционален дисперсии вектора параметров $\tau = 1/\sigma$. Увеличив. параметр τ, уменьшаем дисперсию параметров, следовательно, запрещаем коэф-м w_i принимать слишком большие значения.

(3.3) $X = \mathbb{R}^n$, $Y = \{-1; +1\}$

(*) $a(x) = \mathrm{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0\right) = \mathrm{sign}(\langle w, x\rangle - w_0)$



Требуем: раздел-ая пл-ть максим. далеко отстояла от ближайших к ней (.) обоих классов.

Если выборка линейно неразделима, будем допускать ошибки.

$\xi_i \geqslant 0$ – ошибки на x_i, $i = 1 \dots \ell$.

Введем штраф за суммарную ошибку

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell} \xi_i \to \min \\ y_i(\langle w, x_i\rangle - w_0) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell \\ \xi_i \geqslant 0, \quad i = 1, \dots, \ell \end{cases}$$

условная задача оптимизации.

$M_i(w, w_0) = y_i(\langle w, x_i\rangle - w_0)$ – отступ объекта x_i от границы классов

Алгоритм (*) делает ошибку на объекте x_i $\Leftrightarrow$ когда $M_i < 0$.

Если $M_i \in (-1, +1)$, то $x_i \in$ раздел. полоса

Если $M_i > 1$, то x_i классиф-ся правильно и нах-ся на некот. расст-ии от полосы

Из усл. задачи оптимиз $\Rightarrow$ $\xi_i \geq 0$

$\xi_i \geq 1 - M_i$

$\sum_i \xi_i \to \min$

Т.е. $\xi_i = (1 - M_i)_+$

$\Downarrow$

Безусл. задача оптимизации

$$Q(w, w_0) = \sum_{i=1}^{\ell} (1 - M_i(w, w_0))_+ + \frac{1}{2C}\|w\|^2 \to \min$$