3.1 1 Бинарный линей. классиф

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

$f = \langle w, x \rangle$

2 Отступ алгоритма $a(x) = \text{sign} f(x)$ на объекте x_i наз-ся

$M_i = y_i f(x_i)$ $\quad M_i \le 0 \Leftrightarrow y_i \ne a(x_i)$

$M_i > 0 \Leftrightarrow y_i = a(x_i)$

3

$\text{sign}(\langle w, x \rangle - w_0) \longrightarrow \text{sign}(\langle w, x \rangle)$

надо добавить к векторам x (-1)

4 $\cancel{Q(w) = \sum_{i=1}^{l} (M_i(w) < 0)}$

~~наилучший~~ $\cancel{0}$

$$\tilde{Q}(w) = \sum_{i=1}^{l} \underbrace{L(M_i(w))}_{\text{ф-ия потерь}} \longrightarrow \min_w$$

Должен принимать наим. знач. (0).

5 $\tilde{Q}(w) = \sum_{i=1}^{l} (M_i(w) < 0)$

~~Если взять $w = x$, то~~

$\cancel{a(x) = \text{sign}(\langle x, x \rangle) = \text{sign}\, |x|^2 > 0}$

~~Тогда~~

Возьмем $\omega = 0$ (ну все нули)

Тогда $a(x) = \text{sign}(\langle x, 0\rangle) = \text{sign}\, 0 = 0$

$\Rightarrow M_i(\omega) = 0 \Rightarrow \widetilde{Q}(\omega) = 0$ – это min

(6) $\widetilde{Q}(\omega) = \sum_{i=1}^{l} \log_2 (1 + e^{-M_i})$

(7) Функция потерь – функция от y настоящих и предсказанных. Нужна, чтобы понимать качество модели. Обычно должна уменьшаться при приближении параметров друг к другу.

(8) Негладкая ф-ия потер. В данном случае с лин регрес. $\text{sign}(M)$ – негладкая ф-ия потерь

(9) Регуляризация – техника, направленная на ограничение коэффициентов, чтобы избежать переобучение.

$l_1: \sum_{k=1}^{m} |\omega_k| \leq \tau \quad l_2: \sum_{k=1}^{m} \omega_k^2 \leq \tau$

(10) Переобучение уменьшает обобщающую способность алгор. Регуляризация уменьш. переобучение, следовательно, увеличивает обобщ. способность.

(11) Если мы в остром минимуме, то скорее всего мы переобучились

(12) Делает более простой.

(13) Большее значение функционала аппрокс. риска будет принимать при наличии регуляризации.

~~14~~ Т.к. функционал будет проще, то есть без лишних минимумов.

(14) Если регуляризация оправдывает себя, значит избавляет от переобучения, то есть функционал меньше. Итого: большее значение без регуляризации.

(15) $Accuracy = \frac{tp + fp}{N}$ $\quad N = tp + fp + tn + fp$

$Precision = \frac{tp}{tp + fp}$

$Recall = \frac{tp}{tp + fn}$

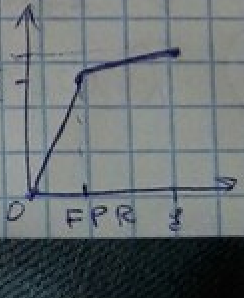
(16) ROC-кривая для бинар. классиф. это зависимость TPR от FPR

AUC - площадь под ROC-кривой

(17) ~~TPR = Recall~~ $FPR = \frac{fp}{fp + tn}$

$TPR = \frac{tp}{tp + fn}$

Нужно посчитать tp, tn, fp, fn и далее построить по ним ~~кривые~~ FPR и TPR и построить по ним:



$\boxed{3.2}$ $P(X, y | \omega)$ — вероятностная модель данных

$p(\omega)$ — априорное распределение

$$P(X^l, \omega) = P(X^l | \omega)\, p(\omega) \text{ — правдопод}$$

логарифм правдопод:

$$L(X^l, \omega) = \sum_{i=1}^{l} \ln p(x_i, y_i | \omega) + \ln p(\omega)$$

т.е. $\ln p(\omega)$ можно рассматривать как регуляризатор

l_2: $p(\omega)$ — n-мерное гауссовское

$$-\ln p(\omega, \sigma) = \frac{1}{2\sigma} \|\omega\|^2 + const(\omega)$$

l_1: $p(\omega)$ — n-мерное лапласа

$$-\ln p(\omega, C) = \frac{1}{C} \sum_{j=1}^{n} |\omega_j| + const(\omega)$$

$\boxed{3.3}$ $\min\limits_{i=1..l} y_i(\langle w, x_i \rangle - w_0) = 1$

$$\langle (x_+ - x_-), \frac{\omega}{\|\omega\|} \rangle = \frac{\langle \omega, x_+ \rangle - \langle \omega, x_- \rangle}{\|\omega\|} =$$

$$= \frac{(\omega_0 + 1) - (\omega_0 - 1)}{\|\omega\|} = \frac{2}{\|\omega\|} = \frac{2}{\langle \omega, \omega \rangle}$$

Итого:
$$\begin{cases} \langle \omega, \omega \rangle \to \min \\ y_i(\langle \omega, x_i \rangle - \omega_0) \geq 1 \end{cases}$$

Если выборка линейно неразделима!

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

$\xi_i \geq 0 \qquad \xi_i \geq 1 - M_i \qquad \sum_{i=1}^{\ell} \xi_i \to \min$

$$\Rightarrow \xi_i = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

Итого: $Q(w, w_0) = \sum_{i=1}^{\ell} (1 - M_i(w, w_0))_+$

$$+ \frac{1}{2C}\|w\|^2 \to \min_{w, w_0}$$

$\boxed{3.4}$ $\quad x_1^2 + 2x_2^2 = 3$

$x_1' = x_1^2 \quad x_2' = x_2^2 \quad \swarrow \varphi \quad w' = w$

Тогда разд пов-ть $\quad x_1' + 2x_2' - 3 = 0$

Размерность спрямляющего пр-ва $- 2$.

Ядро: $\langle \varphi(x), \varphi(w)\rangle$

$\boxed{3.6}$ Смотри $\boxed{3.1}$ 15, 16, 17

$\boxed{3.5}$ $\begin{cases} \sum_{i=1}^{l} L(M_i) \to \min \ (1) \\ \sum_{k=1}^{m} |\omega_k| \leq B \end{cases}$ (4)

(3) $\left\{ \sum_{i=1}^{l} L(M_i) \cancel{\to \min} + \lambda \sum_{k=1}^{m} |\omega_k| \to \min_{\omega} \right.$ $\lambda \geq 0$

$$L(\omega, \lambda) = \sum L(M_i) + \lambda \left(\sum |\omega_k| - b \right)$$

По т. Куна-Таккера

для существования решения $\hat{\omega}$ в (3)

необход и достаточно

$\exists \lambda \geq 0$: 1. $L(\hat{\omega}, \lambda) = \min_{\omega} L(\omega, \lambda)$

2. $\lambda \left(\sum |\omega_k| - b \right) = 0$

При $b > 0$ $\exists \omega$ $\sum |\omega_k| < b$, т.е условие Слейтера выполняется (для достаточн)

(4) $\Leftrightarrow$ 1

Осталось условие 2.

2 $\Leftrightarrow$ $\lambda = 0$ или $\sum |\omega_k| = b$

Если $\lambda = 0$, то $\hat{\omega}$ в $\sum |\omega_k| < b$

$\Rightarrow \sum |\omega_k| \leq b$ - упущение

и (3) $\Leftrightarrow$ (4) при $\lambda = 0$

Если $\lambda > 0$, то $\sum |\omega_k| = b$

выполнения этого условия можно добиться в (4) ограничив в значение портфеля весов оптимальной в (4).