№ 3. Серенко П. 493

1.) $f(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n = [x_0 \equiv 1] = \sum_{i=0}^{N} w_i x_i = \langle w, x \rangle$

$a(x) = \text{sign}(f(x)) = \text{sign}(\langle w, x \rangle)$

2.) $M_i = y_i f(x_i)$

Если $a(x_i) \neq y_i \Rightarrow M_i \leq 0$

Отступ показывает не только ошибается классификатор или нет, но и ~~показывает~~ насколько он уверен в принадлежности к данному классу.

3.) добавляется $x_0 \equiv -1 \Rightarrow \langle w, x \rangle$

4.) $Q = \frac{1}{\ell} \sum_{i=1}^{\ell} I[a(x_i) \neq y_i] = \frac{1}{\ell} \sum_{i=1}^{\ell} I[M_i \leq 0]$

Для наилучшего алгоритма $Q = 0$.

5.) [illegible]

3) добавляется $x_0 \equiv -1 \Rightarrow \langle w, x \rangle$

4) $Q = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) \neq y_i] = \frac{1}{\ell} \sum_{i=1}^{\ell} [M_i \leq 0]$

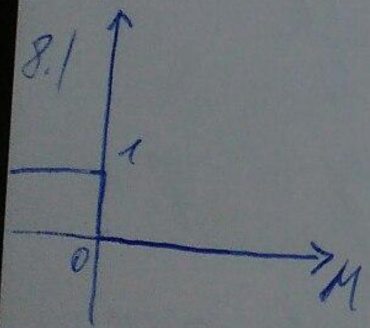
Для наилучшего алгоритма $Q = 0$.

5.) тривиально при $w = 0, \Rightarrow M_i = 0$ и $Q = 0$

6) $Q = \frac{1}{\ell} \sum_{i=1}^{\ell} L(M)$

7.) $L(a, x)$ - показывает ошибку алгоритма на объекте x. Функция неотрицательная и является невозрастающей для линейной классификации.

8.)



9) Регуляризатор штрафует за большие веса признаков, тем самым борясь с переобучением

$\ell_2: \tau \sum_{i=1}^{n} w_k^2$

$\ell_1: \tau \cdot \sum_{i=1}^{n} |w_k|$

τ - глобальный параметр, и неизменяется при минимизации функции потерь.

10.) Обобщ. сп-ть - сп-ть алгоритма не переучиться исключительно под имеющуюся выборку (переобучаться), а также адекватно отвечать на всем множестве значений объектов.
С этим можно бороться используя регуляризатора.

11.) ~~Видим~~ В окрестности этой точки идёт сильное возрастание значения функции аппроксимированного эмпирического риска. Данный параметр будет подавлять остальные => переобучение.

12.) Увеличивается при приближении к границам регуляризатора.

13.) С регуляризатором, т.к. он не даст переучиться под обучающ.

14.) Возможны оба варианта.

~~1.2~~) Регуляризатор дал ожидаемые результаты, не дав переобучиться на обучающей выборке.

2.) Обучающая выборка вытянута в ...

... регуляри-

13.) L регуляризатором, т.к. он не даст переобучиться под обучающ.

14.) Возможны оба варианта.

1.) Регуляризатор дал ухудшение результата, не дав переобучиться на обучающей выборке.

2.) Обучающая выборка включала в себя (или была близка) тестовую. и регуляризация только помешала.

15.) Для бинарной классификации ими являются метриками качества.

$accuracy = \frac{\text{верных ответов}}{\text{кол-во объектов}}$; ~~recall~~ $precision = \frac{tp}{tp+fp}$; $recall = \frac{tp}{tp+fn}$

16.) AUC - площадь $\in [0, 1]$ под ROC-кривой

ROC-кривая $TPR(FPR)$ где $TPR = \frac{TP}{P}$, $FPR = \frac{TN}{N}$

17.) Сортируем выборку по возрастанию по значениям дискр-й ф-и $f(x_i, w)$. Начинаем из точки $(0,0) \Rightarrow$ for i in $(1, \ell)$:

if $(y_i == -1)$

$FPR_i = FPR_{i-1} + \frac{1}{\ell_-}$

else:

$FPR_i = FPR_{i-1} + \frac{1}{\ell_+}$

№ 3.2

$p(x, y | w)$ - совместная пл-ть распределения объектов с классами.

Введем $p(w, \lambda)$ - параметрическое семейство априорных распределений

λ - гиперпараметр.

Пусть выборка может быть порождена каждой из плотностей $p(x, y | w)$ с вер-тью $p(w, \lambda)$

Используем метод максимального правдоподобия.

$$L_\lambda(w, x, y) = \ln p(x^\ell, y^\ell, w, \lambda) = \sum_{i=1}^{\ell} \ln p(x_i, y_i | w) + \ln p(w, \lambda) \to \max_w$$

Положим $\mathcal{L}(y_i, f(x_i, w)) = \ln p(x_i, y_i | w)$

$\lambda v(w) = \ln p(w, \lambda)$

$$\Rightarrow Q(w, x^\ell) = \sum_{i=1}^{\ell} \mathcal{L}(y_i, f(x_i, w)) + \lambda v(w) \to \min_w$$

$\Rightarrow v(w)$ - имеет смысл вероятностного априорного рас-ия параметров модели.

L_1 - п - мажное распределение ...

Используем метод максимального правдоподобия.

$$L_\lambda(w, X, Y) = \ln p(X^l, Y^l, w, \lambda) = \sum_{i=1}^{l} \ln p(x_i, y_i | w) + \ln p(w, \lambda) \to \max_w$$

положим $\mathcal{L}(y_i, f(x_i, w)) = \ln p(x_i, y_i | w)$

$$\lambda V(w) = \ln p(w, \lambda)$$

$$\Rightarrow Q(w, X^l) = \sum_{i=1}^{l} \mathcal{L}(y_i, f(x_i, w)) + \lambda V(w) \to \min_w$$

$\Rightarrow V(w)$ — имеет смысл вероятностного априорного рас-ия параметров модели.

l_1 — n-мерное распределение Лапласа

$$p(w, C) = \frac{1}{(2C)^n} \exp\left(-\frac{\|w\|_1}{C}\right)$$

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

l_2 — n-мерное распределение Гаусса

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

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

№ 3.3.

$X = \mathbb{R}^n, \; Y = \{-1, 1\}$

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

Предположим, что выборка линейно разделима: $\exists\, w, w_0$ т.ч.

$$Q(w, w_0) = \sum_{i=1}^{\ell} [y_i(\langle w, x_i\rangle - w_0) \le 0] = 0$$

Ищем оптимальное расположение гиперплоскости. Сделав нормировку параметров и домножив веса на const, получим $\min\limits_i y_i(\langle w, x_i\rangle - w_0) = 1$.

При макс. ширине полосы на ее краях будут лежать точки разных классов.

$$\Rightarrow \text{ширина полосы: } \left\langle x_+ - x_-, \frac{w}{\|w\|}\right\rangle = \frac{\langle w, x_+\rangle - \langle w, x_-\rangle}{\|w\|} = \frac{(w_0+1) - (w_0-1)}{\|w\|} = \frac{2}{\|w\|}$$

$\Rightarrow$ ширина полосы максимальна при минимальной норме w

$\Rightarrow \{ \; \frac{1}{2}\|w\|^2 \to \min$

$$\ldots \frac{\ldots}{\|w\|} = \frac{(w_0+1)-(w_0-1)}{\|w\|} = \frac{2}{\|w\|}$$

⇒ ширина полосы максимальна при минимальной норме w.

$$\Rightarrow \begin{cases} \frac{1}{2}\|w\|^2 \to \min\limits_{w} \\ y_i(\langle w, x\rangle - w_0) \geq 1, \; i = 1..l \end{cases}$$

В случае линейно разделимой выборки, $y_i(\langle w, x_i\rangle - w_0)$ не обязательно меньше. Ослабляем это ограничение и вводим в минимиз-ч функцию штраф за суммарную ошибку.

$$\begin{cases} \frac{1}{2}\|w\|^2 + C\sum\limits_{i=1}^{l} \varepsilon_i \to \min\limits_{w, w_0, \varepsilon} \\ y_i(\langle w, x_i\rangle - w_0) \geq 1 - \varepsilon_i, \; i = 1..l \\ \varepsilon_i \geq 0 \end{cases}$$

преобразуем $\begin{cases} \varepsilon_i \geq 0 \\ \varepsilon_i \geq 1 - y_i(\langle w, x_i\rangle - w_0) = 1 - M_i(w, w_0) \end{cases}$ minimum

⇒ Получили относительно безусловную задачу оптимизации

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\left(1 - M_i(w, w_0)\right)_+ \to \min_{w, w_0}$$

№3.4.

$$R(x,y) = \langle x, y\rangle^2 = (x_1 y_1 + x_2 y_2)^2 = x_1^2 y_1^2 + x_2^2 y_2^2 + 2 x_1 y_1 x_2 y_2 =$$

$$= \langle (x_1^2, x_2^2, \sqrt{2} x_1 x_2), (y_1^2, y_2^2, \sqrt{2} y_1 y_2)\rangle$$

Получим отображение в спрямляющее пр-во размерности 3.

$$\psi(x_1, x_2) \to (x_1^2, x_2^2, \sqrt{2} x_1 x_2)$$

$$\Rightarrow \underbrace{w_1}_{=1} x_1^2 + \underbrace{w_2}_{=2} x_2^2 + \underbrace{w_3}_{=0} \sqrt{2} x_1 x_2 + \underbrace{w_0}_{=-3} = x_1^2 + 2x_2^2 - 3 = 0$$

№3.5.

$$\begin{cases} f(x) \to \min \\ g(x) < 0 \end{cases} \Leftrightarrow \begin{cases} \nabla_x L(x, \mu) = \partial_x (f(x) + \mu g(x)) = 0; \\ \mu g(x) = 0; \\ \forall i \ \mu_i \geq 0 \end{cases}$$

$$\partial_x (f(x) + \mu g(x)) \Rightarrow f(x) + g(x)\mu \to \min_x$$

$$l_1: \begin{cases} \sum_j^n |w_j| \leq \tau \\ Q_n(w) = \|Fw - y\|^2 \to \min_w \end{cases} \quad \text{по т. КТ} \Rightarrow \begin{cases} Q(w) = \|Fw - y\|^2 + \lambda \sum_{j=1}^n |w_j| + const \to \min_w \\ \lambda \geq 0 \\ \lambda \left(\sum |w_j| - \tau\right) = 0 \end{cases}$$

№3.5.

$$\begin{cases} f(x) \to \min \\ g(x) < 0 \end{cases} \Leftrightarrow \begin{cases} \partial_x L(x, \mu) = \partial_x (f(x) + \mu g(x)) = 0; \\ \mu g(x) = 0; \\ \forall i \ \mu_i \geq 0 \end{cases}$$

$$\partial_x (f(x) + \mu g(x)) \Rightarrow f(x) + g(x)\mu \to \min_x$$

$$l_1: \begin{cases} \sum_j^n |w_j| \leq \tau \\ Q_n(w) = \|Fw - y\|^2 \to \min_w \end{cases} \Rightarrow \text{по т. КТ} \begin{cases} Q(w) = \|Fw - y\|^2 + \lambda \sum_{j=1}^n |w_j| + const \to \min_w \\ \lambda \geq 0 \\ \lambda \left(\sum_{j=1}^n |w_j| - \tau \right) = 0 \end{cases}$$

$$\Rightarrow \begin{cases} Q(w) = \|Fw - y\|^2 + \lambda \sum_{j=1}^n |w_j| + const \to \min_w \\ \lambda = 0 \text{ или } \sum_{j=1}^n |w_j| = \tau \end{cases}$$

$\Rightarrow$ пришли к задаче с добавлением штрафа по l_1-регул.