Чудукова Ирина, 491

Теоретические задачи

$\boxed{31}$

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

$a(x) = sign(f(x))$

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

2) $M_i = y_i f(x_i)$

$a(x_i) \neq y_i \Longleftrightarrow M_i \leq 0$

Если отступ ≤ 0, то это знач. ошибку классификатора

3) Добавляется $x_0 = 1$ фиктивный признак к вектору x, к вектору весов прибавляется w_0

$w_0 + \langle w, x \rangle \Rightarrow \langle w', x' \rangle$

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 $Q = 0$ (все предсказания верны)

5) Положим $w = 0$, тогда $M_i = 0$ и $Q = 0$

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

7) Ф-я потерь $L(a, x)$ сост. величине ошибки алгоритма a на объекте x

$L(a, x) \geq 0$ и явл. невозрастающей для задач классификации

8) $L(M) = [1 - M]_+$

9) Регуляризатор штрафует за большие веса признаков, штрафует за сложность.

L_1-рег-р $\tau \cdot \sum_{k=1}^{n} |w_k|$ — τ - пар-р рег-ии, чем он больше, тем сильнее влияние регуляризатора

L_2-рег-р $\tau \cdot \sum_{k=1}^{n} w_k^2$

10) $Q(a(X^l), X^k)$ - обобщающая способность (X^l и X^k - непер. выборки)

Если алг-м переобучен, то значение большое. Переобучение может произойти, если пар-ры алг-ма (веса признаков) выходят за определенные рамки. Регуляризация как раз борется с этим

11) Вне этой точки происходит сильное возрастание значения функции аппроксимированного эмпирического риска. Данный пар-р подавляет все остальные - переобучение

12) Она увеличивается при прибл. или выходе пар-ров (весов) за определенные границы

13) Для алгоритма с регуляризацией, т.к. регуляризация увеличивает её значение

14) Возможны два варианта

Первый – переобучение без рег-ии не превосходит доп-й вес, кот. вносит регуляризация

Второй – случай сильного переобучения

15) Термины вводятся для задачи бин-й классификации

Обозначения: TP – кол-во объектов, для кот. предск. +1 и ответ +1,

FN – предсказ. −1, ответ +1

TN – предсказ −1, ответ −1

FP – предсказ +1, ответ −1, тогда

$accuracy = \frac{TP+TN}{P+N}$, $precision = \frac{TP}{TP+FP}$; $recall = \frac{TP}{P}$

16) ROC-кривая – график зав-ти TPR от FPR, где $TPR = \frac{TP}{P}$; $FPR = \frac{FP}{N}$

Метрика AUC – площадь под ROC-кривой

17) ROC-кривая строится по мн-ву $\{(FPR_i, TPR_i)\}_{i=0}^{\ell}$, получаемого в рез-те след. алг-ма:

1 Сортируем выборку X^ℓ по зн-м дискр. ф-ции $h(x_i, w)$

2 $(FPR_0, TPR_0) = (0,0)$

for i in range(1, ℓ)

if (y_i == −1):

$(FPR_i, TPR_i) = (FPR_{i-1} + \frac{1}{\ell_-}, TPR_{i-1})$ ← кол-во негативных в выборке

else

$(FPR_i, TPR_i) = (FPR_{i-1}, TPR_{i-1} + \frac{1}{\ell_+})$ ← кол-во позитивных в выборке

$\boxed{3.2}$ $X \times Y$ – вер-ное пр-во

$p(x,y|w)$ – парам-ая модель совместной плотности распределения объектов и классов

Введем $p(w, \gamma)$ – парам-кое сем-во априорных распр-й, γ – гиперпараметр

Б/м считать, что выборка м/ы быть порождена каждой из плотностей $p(x,y|w)$ с вер-тью $p(w, \gamma)$

Приходим к принципу максимума правдоподобия:

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

Положим $L(y_i, f(x_i, w)) = -\ln p(x_i, y_i | w)$

$\gamma V(w) = -\ln p(w, \gamma)$

Приходим к $Q(w, X^\ell) = \sum_{i=1}^{\ell} L(y_i, f(x_i, w)) + \gamma V(w) \to \min_w$

принцип минимиз. ф-ции аппрокс. эмпир. риска

Таким образом, рег-р $V(w)$ имеет вер-й смысл априорного распр-я пар-ов модели

L_1-рег-р соотв-ет n-мерному распр. 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-рег-р соотв. n-мерн. распред. Гаусса

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

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

$\boxed{13.1}$ $X = \mathbb{R}^n$; $Y = \{-1, 1\}$

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

w - в-р весов

$x = (x^1 \dots x^n)$ - признаковое описание объекта

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 гиперпл-ти. Сделаем нормировку пар-ов алг-ма, умножим w и w_0 на const, т.ч.

$$\min_i y_i(\langle w, x_i \rangle - w_0) = 1$$

При макс. ширине разд. полосы, на границе д/ш лежать точки x_- и x_+ от классов -1 и $+1$ соотв.

Ширина полосы равна:

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

Таким образом, ширина полосы макс. при мин. норме w.

Получаем:
$$\begin{cases} \frac{1}{2}\|w\|^2 \to \min_{w, w_0} \\ y_i(\langle w, x_i \rangle - w_0) \ge 1, \quad i = 1, \dots, \ell \end{cases}$$

В случае лин. разделимости выборки $y_i(\langle w, x_i\rangle - w_0)$ не обязательно меньше. Ослабляем огранич-е и вводим в мин-й функц-л штраф за суммарную ошибку. Получаем:

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

Преобразуем усл-я на ε_i:

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

Минимизируем $\varepsilon_i = (1 - M_i(w, w_0))_+$

Получаем безусловную задачу опт-ии:

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

$\boxed{3.4}$ Рассм-м квадратичное ядро:

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

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

Получим отображ. в спрямл. пр-во $H = \mathbb{R}^3$

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

Лин. пов-ть в H будет иметь вид

$$\langle (x_1^2, x_2^2, \sqrt{2} x_1 x_2), (w_1, w_2, w_3)\rangle + w_0 = w_1 x_1^2 + w_2 x_2^2 + w_3 \sqrt{2} x_1 x_2 + w_0 = (1)$$

Положим $w = (1, 2, 0)$, $w_0 = -3$, тогда

$$(1) = x_1^2 + 2x_2^2 - 3 = 0$$

$\boxed{3.5}$ Т-ма Каруша-Куна-Таккера

$$\begin{cases} f(x) \to \min\limits_x \\ g(x) \leq 0 \end{cases} \Longleftrightarrow \begin{cases} \partial_x L(x, \mu) = \partial_x (f(x) + \mu g(x)) = 0 \\ \mu g(x) = 0, \ \mu \geq 0 \end{cases}$$

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

Огр-е l_1-нормы век-ра весов числа:

$$\begin{cases} Q_n(w) = \|Fw - y\|^2 \to \min\limits_w \\ \sum_{j=1}^{n} |w_j| \leq \tau \end{cases} \Longleftrightarrow \begin{cases} Q(w) = \|Fw - y\|^2 \to \min\limits_w \\ \sum_{j=1}^{n} |w_j| - \tau \leq 0 \end{cases}$$

По т-ме ККТ приходим к задаче:

$$\begin{cases} Q(\omega) = \|F\omega - y\|^2 + \lambda\left(\sum\limits_{j=1}^{n} |\omega_j| - \hat{t}\right) \\ \lambda \geqslant 0 \\ \lambda\left(\sum\limits_{j=1}^{n} |\omega_j| - \hat{t}\right) = 0 \end{cases} \Longleftrightarrow \begin{cases} Q(\omega) = \|F\omega - y\|^2 + \lambda \sum\limits_{j=1}^{n} |\omega_j| + const \to \min\limits_{\omega} \\ \lambda \geqslant 0 \\ \lambda = 0 \text{ или } \sum\limits_{j=1}^{n} |\omega_j| = \hat{t} \end{cases}$$

Т.е пришли к задаче с доб-ем штрафа с l_1-нормой.