№ 3.1. Горбунов Даниил 494

1) $a(x) = sign(f(x))$, $f(x)$ - дискриминантная ф-ия,

например: $a(x) = sign(\langle w, x \rangle + w_0)$

2) отступ: $y_i f(x_i)$, частный случай: $y_i (\langle w, x_i \rangle + w_0)$

$a(x) \neq y_i \Leftrightarrow$ отступ ≤ 0

3) достаточно просто добавить признак $x_{n+1} = 1$

и поставить $w_{n+1} = w_0$

4) $Q(a, X^\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) \neq y_i] \Leftrightarrow \frac{1}{\ell} \sum_{i=1}^{\ell} [M_i < 0]$ (из 2))

5) $w = 0$

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

7) $L(a, x)$ показывает ошибку алгоритма на объекте x, причём $L(a, x) \geq 0$. Для лёгкости оптимизации стараются выбирать L - выпуклыми.

(можно штрафовать только за отрицательный отступ в задаче классификации)

8) негладкая $L = I(M \leq 0)$

немонотонная $L = (1 - M)^2$

9) регуляризация - штраф за сложность модели в линейных моделях - это штраф за большие веса

l_1-регуляризация $\sum\limits_{i=1}^{n} |w_i|$

l_2-регуляризация $\sum\limits_{i=1}^{n} |w_i|^2$

смешанная $\alpha \sum\limits_{i=1}^{n} |w_i| + \beta \sum\limits_{i=1}^{n} |w_i|^2$, $\alpha + \beta = 1$

10) Обобщающая сп-ть оценивается ф-ом $Q(a(X^l), X^k)$ на train-выборке X^l и test-выборке X^k $X^l \cap X^k = \varnothing$. При переобучении функционал будет большим, регуляризация помогает это побороть.

11) Выход из минимума даёт сильное увеличение ф-ла.

12) Увеличивает функционал при выходе параметров алгоритма за допустимые границы.

13) Регуляризация увеличивает значение ф-ла риска, поэтому с ней.

15) реальный ответ

ответ алгоритма

	1	0
1	TP	FP
0	FN	TN

$$accuracy = \frac{TP + TN}{P + N}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{P}$$

16) $FPR = \frac{FP}{N}$

$TPR = \frac{TP}{P}$ roc-кривая - график зависимости $TPR(FPR)$ AUC - площадь под графиком

Корбут Даниил
494

№3.1.

18) сортируем выборку X^ℓ по значениям дискриминант. функции, $(FPR_0, TPR_0) = (0;0)$, а дальше пересчитываем FPR, TPR при добавлении нового объекта и ставим точку на графике.

№3.2.

Минимиз. логарифм. функцию потерь:

$$Q(x,w) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + F(w) = \frac{1}{n}\sum_{i=1}^{n} y_i \log \hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)$$
$$+ F(w) = \frac{1}{n}\sum_{i=1}^{n} y_i \log(f(w,x_i)) + (1-y_i)\log(1-f(w,x_i)) + F(w) \to \min$$

Это аналогично максимизации функции правдоподобия

$$\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i)^{y_i} \cdot (1-\hat{y}_i)^{1-y_i} \cdot e^{-F(w)} =$$
$$= \frac{1}{n}\sum_{i=1}^{n} (f(w,x_i))^{y_i} \cdot (1-f(w,x_i))^{1-y_i} e^{-F(w)} = \frac{1}{n}\sum_{i=1}^{n} p(x_i, y_i | w) \cdot e^{-F(w)}$$

Рассмотрим вер. пр-во $\overset{X \times Y}{\vee}$ с распр. $p(x,y|w)$

$p(w) = p(w, \lambda) = e^{-F(w)}$ – априорная вер.

λ – гиперпараметр

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} p(x_i, y_i | w) e^{-F(w)} = \frac{1}{n}\sum_{i=1}^{n} p(x_i, y_i | w)\, p(w) = p(x_i, y_i$$

$\Rightarrow$ регуляризация играет роль априорного распределения весов.

ℓ_1-регуляризация

$w \in \mathbb{R}^n$ имеет n-мерное распр. Лапласа

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

все веса независ., с нулевым мат. ож. и один. дисперсиями. C – гиперпараметр, тогда, логарифмируя:

$$-\ln p(w, C) = \frac{1}{C}\sum_{i=1}^{n} |w_i| + C(w) \text{ - регуляр. по } l_1 \text{ норме}$$

l_2-регуляризация

$w \in \mathbb{R}^n$ имеет норм. распр.

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

после регуляризации:

$$-\ln p(w, \Sigma) = \frac{1}{2\Sigma}\|w\|^2 + c(w) \text{ - } l_2 \text{ регуляриз.}$$

№3.4.

Kernel Trick

$x_1^2 + 2x_2^2 = 3$

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

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

отобр. $\varphi(x_1, x_2) \to (x_1^2, x_2^2, \sqrt{2}, x_1, x_2)$

Линейная поверхность в H:

$\langle (x_1^2, x_2^2, \sqrt{2}\, x_1 x_2), (w_1, w_2, w_3)\rangle + w_0 = w_1 x_1^2 + w_2 x_2^2 +$
$+ w_3 \sqrt{2} x_1 x_2 = \left| w = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} w_0 = -3 \right| = x_1^2 + 2x_2^2 - 3 = 0$

№ 3.5 Корбут Даниил 494

l_1-регуляризация

ограничение l_1-нормы весов числом
$\parallel$
добавление штрафа с его l_1-нормой

По Т. Куна-Таккера

Если $\exists$ решение $w = (\hat{w}_1, \dots \hat{w}_n)$ для задачи с ограничениями

$$\begin{cases} \| \langle w, x \rangle - y \|^2 \to \min \\ \sum_{i=1}^{n} |w_i| - w \le 0 \end{cases}, \text{ то}$$

$\exists$ вектор $\lambda = (\lambda_1, \dots, \lambda_m)$, такой, что выполнены условия:

1) $\min \left(\| \langle w, x \rangle - y \|^2 + \lambda \left(\sum |w_i| - a \right) \right) =$

$= \| \langle \hat{w}, x \rangle - y \|^2 + \lambda \left(\sum |\hat{w}_i| - a \right)$

2) $\lambda \left(\sum |w_i| - a \right) = a$

3) $\lambda_i \ge 0$

=> 2 задачи эквивалентны и приводят к одному решению.

№ 3.6.

$$\text{PRECISION} = \frac{\text{верно угаданные объекты 1-го класса}}{(\text{общее количество объектов, которым мы предсказали 1-ый класс})}$$

$$\text{RECALL} = \frac{\text{верно угаданные объекты 1-го класса}}{\text{количество объектов 1-го класса}}$$

$$\text{ACCURACY} = \frac{\text{количество верных предсказаний}}{\text{общее кол-во верных предсказ}}$$

ROC и AUC уже расписал в № 3.1

алгоритм тоже :)