Шеншиной Ю. 491

3.1.

1. $a(x) = \text{sign}\, f(x)$, где $f(x) = w_0 + \langle w, x \rangle$

2. $M_i = y_i f(x_i)$
Если $\text{sign}\, M_i > 0$, то класс был предсказан верно
Если $\text{sign}\, M_i < 0$, то неверно.

3. Было: $w_0 + \langle w, x \rangle$ — добавим фиктивную координату к X
Стало: $\langle \tilde{w}, \tilde{x} \rangle$
$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \Rightarrow \tilde{x} = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}, \tilde{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{pmatrix}$

4. $Q(x) = \frac{1}{m} \sum\limits_{i=1}^{m} I(M_i \leq 0)$
Чтобы алгоритм был оптимальным (все предсказания верны), нужно, чтобы $Q(X) = 0$.

5. Если все $M_i < 0$, то, положив $w = \vec{0}$, получим $\forall i\ M_i = 0 \Rightarrow Q(x) \equiv 0$.

6. $Q = \frac{1}{m} \sum\limits_{i=1}^{m} L(M)$

7. Ф-ия потерь $L(ax)$ - величина ошибки алгоритма a на объекте x.
Ф-ия $L(a, x) \geq 0$ и не возрастает.

8. $L(M) = (1 - M)_+$

9. Регуляризация - добавление нек-рой дополнительной информации к условию с целью решить некорректно поставленную задачу (например, разделение неперемасовостью выборки со смещением классов) или предотвратить переобучение.
Пример: l_1 - регуляризация: $\gamma \cdot \sum\limits_{k=1}^{n} |w_k|$
l_2 - регуляризация: $\gamma \cdot \sum\limits_{k=1}^{n} |w_k|^2$

10. Пусть X и Y - непересекающиеся выборки. Обобщающая способность алгоритма a определяется функционалом $Q(a(X), Y)$. Переобученный алгоритм приводит к большому значению Q. Добавление регуляризатора решает проблему переобучения, не позволяя какому-то из весов сильно превосходить остальные.

11. Т.к. минимум острый, то даже небольшое отклонение от точки минимума приведет к резкому увеличению значения функционала и переобучению.

12. Увеличивается по мере приближения весов к определённым границам и выходе за них.

13. Для построенного с регуляризацией, т.к. регуляризатор ≥ 0.

14. Рассмотрим случай обучение без регуляризатора:

Если при отсутствии регуляризатора алгоритм переобучился на обучающей выборке, то на тестовой выборке получим плохой результат
$\Rightarrow$ значение без регуляризатора может быть больше.

Если же при отсутствии регуляризатора алгоритм не переобучался, то значение функционала с регуляризатором может быть больше.

$\Rightarrow$ возможны оба случая

Вопросы 15, 16 и 17 совпадают с вопросами 1, 2 и 3 задачи 3.6 соответственно.

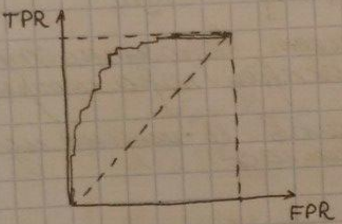
15. $\text{Accuracy} = \dfrac{\#\text{правильные ответы}}{\#\text{все ответы}}$ — доля правильных ответов

$$\text{Precision} = \frac{(\text{true positive})}{(\text{true positive}) + (\text{false positive})}$$

$$\text{Recall} = \frac{(\text{true positive})}{(\text{true positive}) + (\text{false negative})}$$, где

		Реальный класс	
		1	0
Спрогнозированный класс	1	true positive	false positive
	0	false negative	true negative

16. ROC-AUC — метрика, оценивающая качество ранжирования (в задачах бинарной классификации) по вероятности принадлежности к целевому классу.



ROC-AUC = площадь под графиком

$TPR = \dfrac{tp}{tp+fn}$ — true positive rate

$FPR = \dfrac{fp}{fp+tn}$ — false positive rate

17. Рассмотрим выборку X из m объектов.
Пусть $m_- = \sum_{i=1}^{m} I(y_i = -1)$, $m_+ = \sum_{i=1}^{m} I(y_i = 1)$

1) Упорядочим выборку X по убыванию значения $\langle x_i, w \rangle$
2) Начальная точка ROC-кривой:
   $(FPR_0, TPR_0) := (0, 0)$
   $AUC := 0$
3) for i in $(1, m)$:
   if $(y_i = -1)$:
   $$\left.\begin{aligned} FPR_i &= FPR_{i-1} + \frac{1}{m_-} \\ TPR_i &= TPR_{i-1} \\ AUC &= AUC + \frac{1}{m_-} TPR_i \end{aligned}\right\} \text{сместились на 1 шаг вправо}$$
   else:
   $$\left.\begin{aligned} FPR_i &= FPR_{i-1} \\ TPR_i &= TRR_{i-1} + \frac{1}{m_+} \end{aligned}\right\} \text{сместились на 1 шаг вверх}$$

— получим ROC-кривую.

3.2 Пусть $p(x_i, y_i | w)$ – совместное распред. x_i и y_i,
$\{ q_\ell(w) \mid \ell \in A \}$ – семейство априорных распред. w.

Будем считать, что выборка может иметь любую из $p(x, y | w)$ плотностей с вероятностью $q_\ell(w)$.

Будем искать w с помощью метода максимального правдоподобия:

$$f = \prod_{i=1}^{m} p(x_i y_i | w) \cdot q_\ell(w)$$

$$\tilde{L}(w) = \ln f = \sum_{i=1}^{m} \left[\ln p(x_i y_i | w) + q_\ell(w) \right] \longrightarrow \max_w$$

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

$$\alpha \xi(w) = -\ln \ (w)$$

Получаем задачу оптимизации:

$$Q(w) = \sum_{i=1}^{m} \left[L(x_i y_i W) + \alpha \xi(w) \right] \longrightarrow \min_w$$

Т.е. регуляризатор ξ имеет вероятностный смысл априорного распред. параметра w.

4 l_1-регуляризатор – соотв. n-мерному распределению Лапласа:

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

$$-\ln q(w) = \frac{1}{C} \sum_{i=1}^{m} |w_i| + \overset{const}{\tilde{C}(w)}$$

4 l_2-регуляризация – соотв. n-мерному распред. Гаусса:

$$q(w) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left(-\frac{\|w\|_2}{2\sigma}\right)$$

$$-\ln q(w) = \frac{1}{2\sigma} \|w\|_2 + \tilde{C}(w)$$

3.3 Пусть $X \in \mathbb{R}^n$, $Y = \{-1, 1\}$

$a(x) = sign(\langle w, x \rangle - w_0)$

Предположим, что выборка может быть линейно разделена, т.е

$\exists w, w_0:\ Q(w, w_0) = 0$, где $Q(w, w_0) = \sum_{i=1}^{m} I(y_i(\langle w, x_i \rangle - w_0) \le 0)$

Будем искать оптимальное положение гиперплоскости.

Отнормируем параметры w и w_0 так, чтобы

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

Строим разделяющую полосу.

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

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

=> для увеличения ширины разделяющей полосы нужно минимизировать $\|w\|$.

Формулируем задачу:

$$\begin{cases} \frac{1}{2}\|w\|^2 \longrightarrow \min\limits_{w, w_0} \\ y_i(\langle w, x \rangle - w_0) \ge 1 \end{cases}$$

Для случая линейно неразделимой выборки введём штраф за выбросы:

$$\begin{cases} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m} \xi_i \longrightarrow \min\limits_{w, w_0, \xi} \\ y_i(\langle w, x_i \rangle - w_0) \ge 1 - \xi_i \\ \xi_i \ge 0 \end{cases}$$

Сведём задачу к задаче безусловной оптимизации:

$$\begin{cases} \xi_i \ge 1 - y_i(\langle w, x_i \rangle - w_0) \\ \xi_i \ge 0 \end{cases}$$

Т.к. $M_i = y_i(\langle w, x_i \rangle - w_0)$, то имеем задачу

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

(3.4) $x_1^2 + 2x_2^2 = 3$

Рассмотрим квадратичное ядро:

$K(x,y) = \langle x, y\rangle^2 = (x_1 y_1 + x_2 y_2)^2 = x_1^2 y_1^2 + 2x_1x_2y_1y_2 + x_2^2 y_2^2 =$

$= \langle (x_1^2, x_2^2, \sqrt{2}x_1x_2), (y_1^2, y_2^2, \sqrt{2}y_1y_2)\rangle$

Получаем отображение $\varphi: \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{pmatrix} \in \mathbb{R}^3$

В спрямляющем пространстве поверхность будет иметь вид гиперплоскости

$\langle \tilde{x}, w\rangle - w_0 = 0$

$\langle (x_1^2, x_2^2, \sqrt{2}x_1x_2), (w_1, w_2, w_3)\rangle - w_0 = 0$

$\Rightarrow \quad w = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \quad w_0 = -3.$

(3.5) Из т. ККТ:

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

Рассмотрим задачу:

$\begin{cases} \sum\limits_{i=1}^{m} L(x_i, y_i, w) \longrightarrow \min\limits_w \\ \sum\limits_{j=1}^{m} |w_j| < \alpha \end{cases} \Longleftrightarrow \begin{cases} \sum\limits_{i=1}^{m} L(x_i, y_i, w) \longrightarrow \min\limits_w \\ \sum\limits_{i=1}^{m} |w_i| - \alpha < 0 \end{cases} \quad (*)$

Задача (*) эквивалентна задаче (по т. ККТ):

$(**) \quad \begin{cases} \sum\limits_{i=1}^{m} L(x_i, y_i, w) + \mu \sum\limits_{i=1}^{m} |w_i| - \alpha\mu \longrightarrow \min\limits_w \\ \mu\left(\sum\limits_{i=1}^{m} |w_i| - \alpha\right) = 0 \\ \mu \geq 0 \end{cases}$

Из (**) видно, что $\left[\begin{array}{l} \sum\limits_{i=1}^{m} |w_i| = \alpha \\ \mu = 0 \end{array} \right.$

$\Rightarrow$ имеем задачу с l_1-регуляризац. $\begin{cases} \sum\limits_{i=1}^{m} L(x_i, y_i, w) + \mu \sum\limits_{i=1}^{m} |w_i| \longrightarrow \min\limits_w \\ \mu \geq 0 \\ \left[\begin{array}{l} \mu = 0 \\ \sum\limits_{i=1}^{m} |w_i| = \alpha \end{array} \right. \end{cases}$