Машинное обучение, ФКН ВШЭ
Семинар №16

1 Метод опорных векторов

Задача 1.1. *Рассмотрим двухмерную задачу классификации ($d = 2$) с двумя классами $\mathbb{Y} = \{-1, +1\}$. В обучающей выборке 5 точек: три точки $\{(1,1),(1,2),(2,3)\}$ класса $+1$ и две точки $\{(3,1),(4,2)\}$ класса -1. Найдите линейный классификатор $a(x) = \operatorname{sign}(w_1 x_1 + w_2 x_2 + b)$, который будет получен в результате обучения методом опорных векторов.*

Решение. Выборка является линейно разделимой, поэтому можно решать следующую задачу:

$$\begin{cases} \frac{1}{2}(w_1^2 + w_2^2) \to \min_2 \\ y_i(\langle w, x_i \rangle + b) \geqslant 1. \end{cases}$$

Запишем лагранжиан:

$$L(w, b, \lambda) = \frac{1}{2}(w_1^2 + w_2^2) + \sum_{i=1}^{5} \lambda_i \left(1 - y_i(\langle w, x_i \rangle + b)\right)$$

и условия Куна–Таккера:

$$\begin{cases} w_1 - \lambda_1 - \lambda_2 - 2\lambda_3 + 3\lambda_4 + 4\lambda_5 = 0; \\ w_2 - \lambda_1 - 2\lambda_2 - 3\lambda_3 + \lambda_4 + 2\lambda_5 = 0; \\ \lambda_1 + \lambda_2 + \lambda_3 - \lambda_4 - \lambda_5 = 0; \\ \lambda_1 = 0 \quad \text{или} \quad 1 - w_1 - w_2 - b = 0; \\ \lambda_2 = 0 \quad \text{или} \quad 1 - w_1 - 2w_2 - b = 0; \\ \lambda_3 = 0 \quad \text{или} \quad 1 - 2w_1 - 3w_2 - b = 0; \\ \lambda_4 = 0 \quad \text{или} \quad 1 + 3w_1 + w_2 + b = 0; \\ \lambda_5 = 0 \quad \text{или} \quad 1 + 4w_1 + 2w_2 + b = 0; \end{cases}$$

(мы опустили условия о выполнении ограничений в прямой и двойственной задачах).

Нарисовав выборку на плоскости, можно заметить, что объекты $(1,2)$ и $(4,2)$ не являются опорными, и поэтому соответствующие им двойственные переменные будут равны нулю: $\lambda_2 = \lambda_5 = 0$. Остальные же объекты при оптимальной разделяющей полосе будут опорными и их отступы будут равны 1. Получаем систему

$$\begin{cases} w_1 - \lambda_1 - \lambda_2 - 2\lambda_3 + 3\lambda_4 + 4\lambda_5 = 0; \\ w_2 - \lambda_1 - 2\lambda_2 - 3\lambda_3 + \lambda_4 + 2\lambda_5 = 0; \\ \lambda_1 + \lambda_2 + \lambda_3 - \lambda_4 - \lambda_5 = 0; \\ 1 - w_1 - w_2 - b = 0; \\ \lambda_2 = 0; \\ 1 - 2w_1 - 3w_2 - b = 0; \\ 1 + 3w_1 + w_2 + b = 0; \\ \lambda_5 = 0. \end{cases}$$

Здесь мы опустили все неравенства. Решая систему, получим $b = 1.5$, $w_1 = -1$, $w_2 = 0.5$, $\lambda_1 = 0.375$, $\lambda_2 = 0$, $\lambda_3 = 0.25$, $\lambda_4 = 0.625$, $\lambda_5 = 0$.

■

Задача 1.2. *Рассмотрим задачу с линейно разделимой выборкой. Допустим, мы решили двойственную задачу SVM и нашли вектор двойственных переменных λ. Покажите, что половина ширины разделяющей полосы ρ может быть вычислена по следующей формуле:*

$$\frac{1}{\rho^2} = \sum_{i=1}^{\ell} \lambda_i.$$

Решение. Поскольку выборка линейно разделима, то все объекты, для которых $\lambda_i \neq 0$, окажутся на границе разделяющей полосы. Для них будет выполнено равенство

$$y_i \left(\langle w, x_i \rangle + b \right) = 1,$$

из которого можно выразить b:

$$b = y_i - \langle w, x_i \rangle.$$

Домножим обе стороны на $\lambda_i y_i$ и просуммируем по i (заметим, что для объектов не на границе разделяющей полосы выполняется $\lambda_i y_i = 0$):

$$b \sum_{i=1}^{\ell} \lambda_i y_i = \sum_{i=1}^{\ell} \lambda_i y_i^2 - \sum_{i=1}^{\ell} \lambda_i y_i \langle w, x_i \rangle.$$

Поскольку w, b и λ здесь — решения прямой и двойственной задач, то для них выполнены условия Куна-Таккера. В частности,

$$\sum_{i=1}^{\ell} \lambda_i y_i = 0,$$

$$w = \sum_{i=1}^{\ell} \lambda_i y_i x_i.$$

Заметим также, что $y_i^2 = 1$. Воспользовавшись этими тремя равенствами, получаем:

$$0 = \sum_{i=1}^{\ell} \lambda_i - \|w\|^2.$$

Ранее мы доказали, что в SVM ширина разделяющей полосы равна $\frac{2}{\|w\|}$, поэтому

$$0 = \sum_{i=1}^{\ell} \lambda_i - \frac{1}{\rho^2}.$$

Отсюда получаем требуемое равенство.

■

Задача 1.3. *Пусть $(w, b, \xi_1, \dots, \xi_\ell)$ — оптимальное решение прямой задачи SVM. Предположим, что $\xi_3 > 0$. Выразите отступ объекта x_3 для обученного линейного классификатора через значения $(\xi_1, \dots, \xi_\ell)$.*

Решение. Заметим, что, поскольку $\xi_i > 0$, то объект x_i является опорным нарушителем. Отсюда следует, что $\lambda_i = C$. Напомним, что для двойственной задачи можно записать условия дополняющей нежесткости:

$$\lambda_i[y_i\left(\langle w, x_i\rangle + b\right) - 1 + \xi_i] = 0,$$

откуда можно получить, что $y_i\left(\langle w, x_i\rangle + b\right) - 1 + \xi_i = 0 \Leftrightarrow M_i = y_i\left(\langle w, x_i\rangle + b\right) = 1 - \xi_i$.

■

Задача 1.4. *Пусть мы решили двойственную задачу SVM и получили решение $(\lambda_1, \dots, \lambda_\ell)$. Пусть мы также восстановили оптимальный порог b. Выразите:*

1. *Квадрат нормы $\|w\|^2$ оптимального вектора w для прямой задачи;*
2. *Сумму $\sum_{i=1}^{\ell} \xi_i$ оптимальных значений параметров $\xi_1, \dots, \xi_\ell$ для прямой задачи.*

Решение.

1. Напомним, что из условий Куна-Таккера для двойственной задачи имеем $w = \sum_{i=1}^{\ell} \lambda_i y_i x_i$. Отсюда

$$\|w\|^2 = \langle w, w\rangle = \left\langle \sum_{i=1}^{\ell} \lambda_i y_i x_i, \sum_{j=1}^{\ell} \lambda_j y_j x_j \right\rangle = \sum_{i,j=1}^{\ell} \lambda_i \lambda_j y_i y_j \langle x_i, x_j\rangle.$$

2. Напомним, что имеет место

$$\mu_i \xi_i = 0 \Leftrightarrow (\mu_i = 0) \text{ или } (\xi_i = 0),$$

поэтому имеет смысл рассматривать лишь те объекты, для которых $\mu_i = 0$. Из $\lambda_i + \mu_i = C$ имеем $\lambda_i = C \neq 0$. Отсюда и из $\lambda_i[y_i\left(\langle w, x_i\rangle + b\right) - 1 + \xi_i] = 0$ имеем

$$y_i\left(\langle w, x_i\rangle + b\right) - 1 + \xi_i = 0 \Leftrightarrow \xi_i = 1 - y_i\left(\langle w, x_i\rangle + b\right) =$$
$$= 1 - y_i\left(\left\langle \sum_{j=1}^{\ell} \lambda_j y_j x_j, x_i \right\rangle + b\right) = 1 - y_i\left(\sum_{j=1}^{\ell} \lambda_j y_j \langle x_i, x_j\rangle + b\right).$$

Отсюда имеем:

$$\sum_{[i=1]}^{\ell} \xi_i = \sum_{i=1}^{\ell} \left(1 - y_i \left(\sum_{j=1}^{\ell} \lambda_j y_j \langle x_i, x_j \rangle + b\right)\right) =$$

$$= \ell - \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} y_i y_j \lambda_j \langle x_i, x_j \rangle - b \sum_{i=1}^{\ell} y_i.$$

■