Семинары по линейным классификаторам

Евгений Соколов
sokolov.evg@gmail.com

5 ноября 2013 г.

4 Метод опорных векторов

Задача 4.1. *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

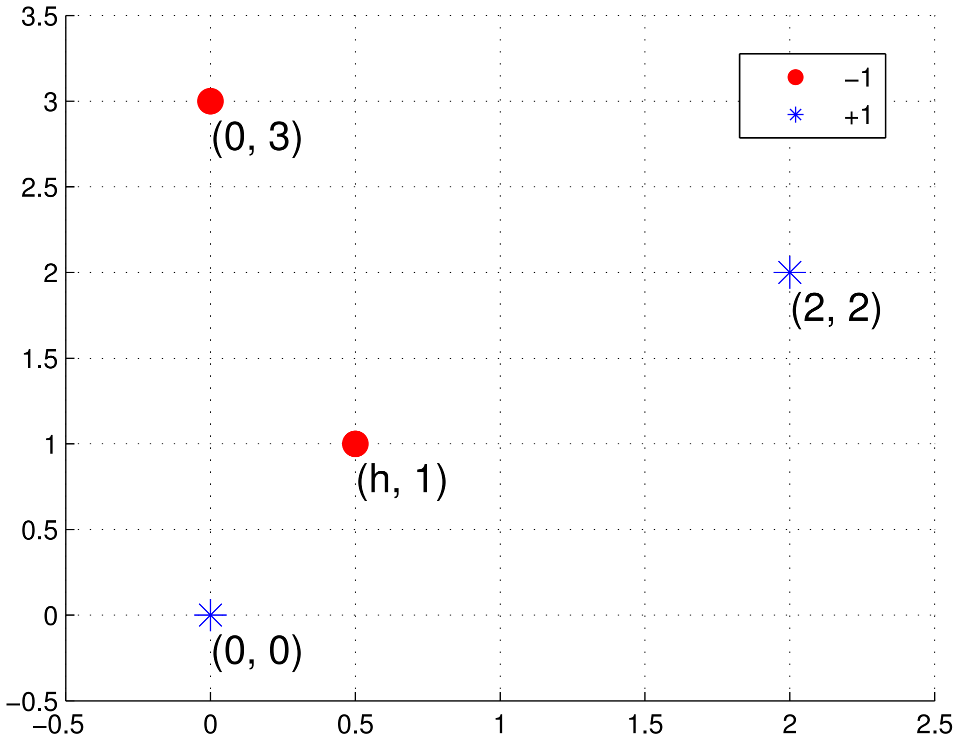

Рис. 1. К задаче 2.

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. Решаю 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 4.2. *На рисунке 1 изображена выборка из четырех объектов и двух классов. Величина $0 \leqslant h \leqslant 3$ — параметр. Будем строить разделяющую гиперплоскость методом опорных векторов:*

$$\begin{cases} \frac{1}{2}\|w\|^2 \to \min; \\ y_i(\langle w, x_i\rangle + \ ) \geqslant 1, \quad i = 1, \ldots, 4. \end{cases} \tag{4.1}$$

1. *При каких значениях параметра h оптимизационная задача (4.1) будет разрешима?*
2. *Будет ли меняться наклон оптимальной разделяющей гиперплоскости при изменении параметра h?*
3. *Как ширина разделяющей полосы, соответствующей оптимальной разделяющей гиперплоскости, выражается через параметр h?*

Решение. Заметим, что объекты класса $+1$ лежат на прямой $y = x$. Легко видеть, что объекты можно безошибочно разделить гиперплоскостью только если объект $(h, 1)$ лежит слева от прямой, соединяющей объекты класса $+1$, то есть при $h < 1$.

При $0 \leqslant h < 1$ оптимальная разделяющая полоса всегда будет проходить через объекты $(h, 1)$, $(0, 0)$ и $(2, 2)$, и угол ее наклона будет одинаков — отличаться будет лишь ширина.

Видно, что ширина разделяющей полосы равна расстоянию от объекта $(h, 1)$ до прямой $y = x$, которое можно вычислить по формуле

$$\frac{1}{\sqrt{2}}|h - 1|.$$

■

Задача 4.3. *Рассмотрим задачу с линейной 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:

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

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.

■