

Задача к собеседованию на ИППИ

задача предложена *Бурнаевым Е.В.*

23 апреля 2018

Задача

Задача классификации заключается в том, что по выборке данных

$$\{(\mathbf{x}_i, y_i = y(\mathbf{x}_i))\}_{i=1}^n, \quad \mathbf{x}_i \in \mathbb{R}^d, \quad y_i \in \{-1, 1\}$$

необходимо построить модель зависимости $\hat{y}(\mathbf{x})$ такую, что $\hat{y}(\mathbf{x}) \in \{-1, 1\}$ и для большинства значений \mathbf{x} прогноз метки класса $\hat{y}(\mathbf{x})$ совпадает с настоящей меткой класса $y(\mathbf{x})$.

Рассматривается модель линейной разделяющей гиперплоскости

$$\hat{y}(\mathbf{x}) = \text{sign}(\mathbf{x}_i^T \mathbf{w} + w_0)$$

Для оценки вектора параметров \mathbf{w} , w_0 максимизируется отступ разделяющей гиперплоскости от объектов обучающей выборки:

$$\begin{aligned} & \max_{\mathbf{w}, w_0, \|\mathbf{w}\|=1} M, \\ & s. t. \quad y_i(\mathbf{x}_i^T \mathbf{w} + w_0) \geq M, \quad i = \overline{1, n}. \end{aligned}$$

Нужно описать, как решать такую задачу оптимизации и какими свойствами обладает ее решение.

Решение

1. Классы разделимы

Вместо использования ограничений, поделим обе части неравенства на модуль вектора весов. Эквивалентное неравенство:

$$\frac{1}{||\mathbf{w}||} y_i (\mathbf{x}_i^T \mathbf{w} + w_0) \geq M,$$

или, взяв $||\mathbf{w}|| = \frac{1}{M}$,

$$\begin{aligned} & \min_{\mathbf{w}, w_0} ||\mathbf{w}||^2, \\ & s. t. \quad y_i (\mathbf{x}_i^T \mathbf{w} + w_0) \geq 1, \quad i = \overline{1, n}. \end{aligned}$$

Функция Лагранжа:

$$L = \frac{1}{2} ||\mathbf{w}||^2 - \sum_{i=1}^N \lambda_i (y_i (\mathbf{x}_i^T \mathbf{w} + w_0) - 1),$$

условия Каруша-Куна-Таккера:

- стационарности: $\max_{\mathbf{x}} L = L(\hat{\mathbf{x}})$
- дополняющей нежесткости: $\lambda_i (y_i (\mathbf{x}_i^T \mathbf{w} + w_0) - 1) = 0, \quad i = \overline{1, N}$
- неотрицательности: $\lambda_i \geq 0, \quad i = \overline{1, N}$

Из условия стационарности мы должны приравнять производные к нулю по \mathbf{w} и w_0 .

$$\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i$$

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

Подставив эти значения в функцию Лагранжа, получим

$$L = \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^N \lambda_i \lambda_k y_i y_k \mathbf{x}_i^T \mathbf{x}_k,$$

и теперь нужно минимизировать $-L$ при $\lambda_i \geq 0$.

Условие дополняющей нежесткости:

$$\lambda_i(y_i(\mathbf{x}_i^T \mathbf{w} + w_0) - 1) = 0, \quad i = \overline{1, N}$$

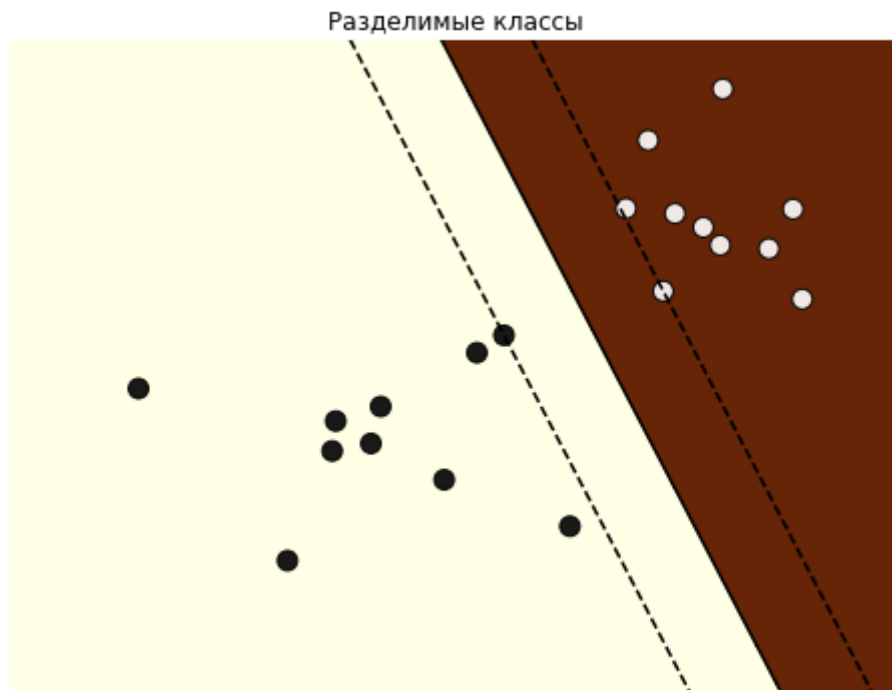
Либо $(y_i(\mathbf{x}_i^T \mathbf{w} + w_0) = 1$ при $\lambda_i > 0$, то есть \mathbf{x}_i лежит на *границе* разделяющей области, либо $(y_i(\mathbf{x}_i^T \mathbf{w} + w_0) > 1$ при $\lambda_i = 0$, и тогда объект на границе не лежит.

$$\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i$$

Вектор \mathbf{w} есть линейная комбинация так называемых **опорных объектов**, для которых выполнено условие $\lambda_i \neq 0$.

Чтобы найти w_0 , нужно решить условие дополнительной нежесткости для одного из опорных объектов.

```
In [42]: fig, axes = plt.subplots(1, 1, figsize=(8, 6))
plot_decision_function(clf_no_weights, sample_weight_constant, axes,
                        u"Разделимые классы")
plt.show()
```



2. Неразделимые классы

Разрешаем классификатору допускать ошибки, но будем за них ругать.

Аналогичным образом получаем следующую задачу с переменными \mathbf{w} , w_0 , ξ , а также ценой ошибки C :

$$\begin{aligned} \min_{\mathbf{w}, w_0} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i, \\ \text{s. t. } \quad & \xi_i \geq 0, \quad y_i(\mathbf{x}_i^T \mathbf{w} + w_0) \geq 1 - \xi_i, \quad i = \overline{1, n}. \end{aligned}$$

Функция Лагранжа для данной задачи примет вид

$$L = \frac{1}{2} ||\mathbf{w}||^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \lambda_i (y_i (\mathbf{x}_i \mathbf{w} + w_0) - (1 - \xi_i)) - \sum_{i=1}^N \mu_i \xi_i$$

Приравняв к нулю производные, получим

$$\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i$$

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

$$\lambda_i = C - \mu_i$$

Функция Лагранжа может быть записана в таком виде:

$$L = \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^N \lambda_i \lambda_k y_i y_k \mathbf{x}_i^T \mathbf{x}_k$$

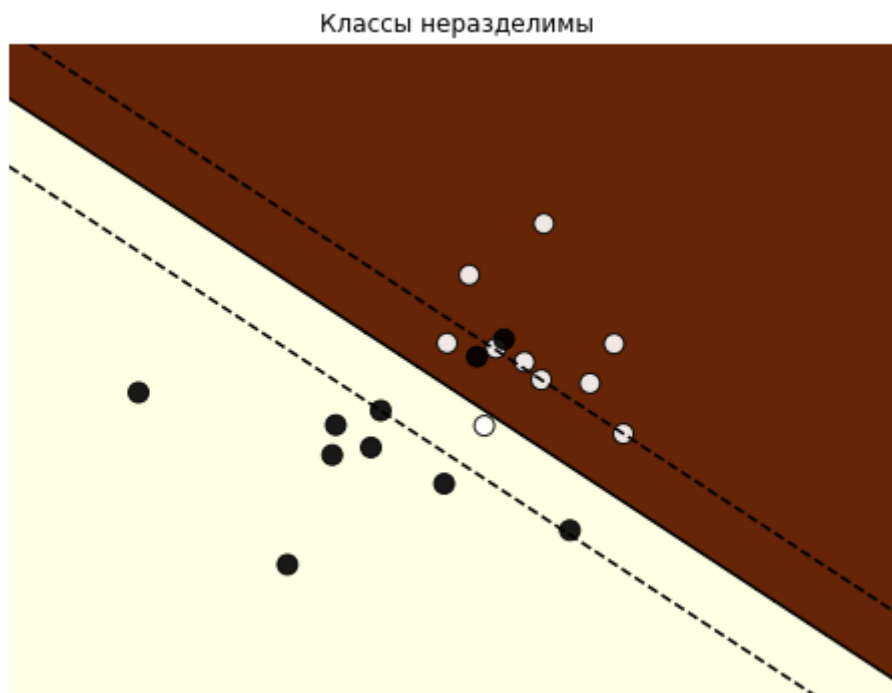
Вновь добавляем КТ-условия:

- дополняющей нежесткости:
 - $\lambda_i (y_i (\mathbf{x}_i^T \mathbf{w} + w_0) - (1 - \xi_i)) = 0, \quad i = \overline{1, N}$
 - $\mu_i \xi_i = 0$
- неотрицательности: $\lambda_i \geq 0, \quad i = \overline{1, N}$

Результат: \mathbf{w} есть линейная комбинация *опорных векторов*.

```
In [45]: fig, axes = plt.subplots(1, 1, figsize=(8, 6))
plot_decision_function(clf_no_weights, sample_weight_constant, axes,
                      u"Классы неразделимы")

plt.show()
```



3. У объектов есть веса

Мы продолжаем разрешать классификатору допускать ошибки, но теперь их стоимость различается для всех объектов выборки.

Запишем задачу:

$$\min_{\mathbf{w}, w_0} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N W_i \xi_i,$$

$$s. t. \xi_i \geq 0, \quad y_i(\mathbf{x}_i^T \mathbf{w} + w_0) \geq 1 - \xi_i, \quad i = \overline{1, n}.$$

За W_i обозначен вес i -го объекта.

Функция Лагранжа для данной задачи примет вид

$$L = \frac{1}{2} ||\mathbf{w}||^2 + C \sum_{i=1}^N W_i \xi_i - \sum_{i=1}^N \lambda_i (y_i (\mathbf{x}_i \mathbf{w} + w_0) - (1 - \xi_i)) - \sum_{i=1}^N \mu_i \xi_i$$

Приравняв к нулю производные, получим

$$\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i$$

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

$$\lambda_i = C W_i - \mu_i$$

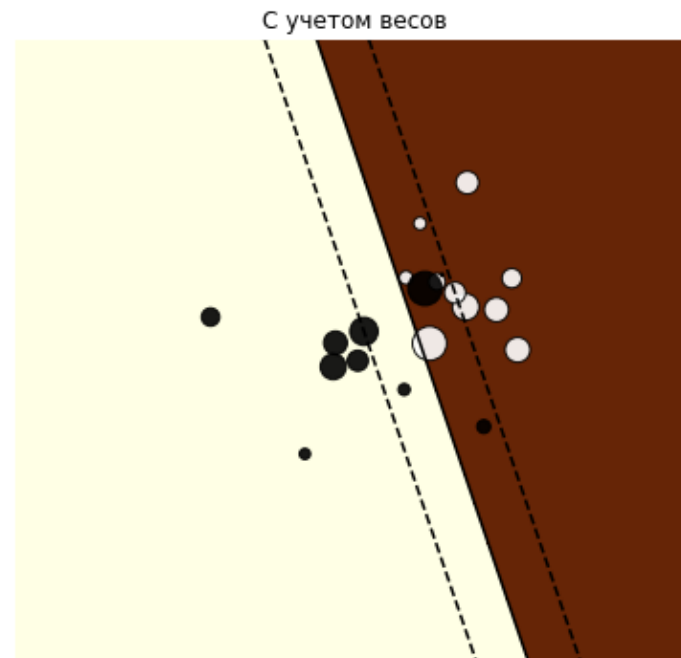
КТ-условия не изменятся.

Единственная разница по сравнению с предыдущим случаем - другое ограничение на λ_i

При $C \rightarrow 0$ или $C \rightarrow +\infty$ $CW_i = C \Rightarrow$ влияние весов учтено не будет.

```
In [39]: fig, axes = plt.subplots(1, 2, figsize=(14, 6))
plot_decision_function(clf_no_weights, sample_weight_constant, axes[0],
                        u"Без учета весов")
plot_decision_function(clf_weights, sample_weight_last_ten, axes[1],
                        u"С учетом весов")

plt.show()
```



Литература

- T. Hastie and others. *The elements of statistical learning*
- Mehryar Mohri and others. *Foundations of Machine Learning*
- Википедия
- Xulei Yang, Q. Song. *Weighted support vector machine for data classification*