Семинар 7.
ММП, осень 2012–2013
20 ноября

Темы семинара:

- Линейные методы классификации;
- Двойственная задача оптимизации;
- Метод опорных векторов (SVM), soft-margin;
- Ядра, ядровая функция.

1 Разбор домашнего задания

Задача 1. Рассмотрим двухмерную задачу классификации с двумя классами $\mathbb{Y} = \{-1, +1\}$. В обучающей выборке 5 точек: $\{(1,1),(1,2),(2,3)\}$ класса $+1$ и $\{(3,1),(4,2)\}$ класса -1. Постройте оптимальную разделяющую гиперплоскость $a(\boldsymbol{x}) = \operatorname{sgn}(w_1 x_1 + w_2 x_2 - w_0)$, выписав соответствующую задачу оптимизации и решив ее с помощью методов, описанных на прошлом семинаре.

Какие из точек обучающей выборки получились опорными? Какова ширина получившейся разделяющей полосы?

Подумайте, какие точки плоскости при добавлении в обучающую выборку не изменили бы решения? Какие бы из них стали опорными векторами? Какие бы не стали опорными?

Решение. Нам надо решить следующую задачу:

$$\begin{cases} \frac{1}{2}(w_1^2 + w_2^2) \to \min; \\ y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0) \geqslant 1, \end{cases}$$

где $\boldsymbol{w} = (w_1, w_2)$. Это выпуклая задача оптимизации. Несложно заметить, что обучающая выборка линейно разделима, а значит найдутся такие w_0, w_1, w_2, что все условия-неравенства выполнятся *строго*. Значит условия Куна–Таккера являются необходимым и достаточными условиями на оптимальное решение этой задачи. Запишем Лагранжиан:

$$L(\boldsymbol{w}, w_0, \lambda_1, \ldots, \lambda_5) = \frac{1}{2}(w_1^2 + w_2^2) + \sum_{i=1}^{5} \lambda_i \big(1 - y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0)\big)$$

и условия Куна–Таккера:

$$\begin{cases} w_1 - \lambda_1 - \lambda_2 - 2\lambda_3 + 3\lambda_4 + 4\lambda_5 = 0; \\ w_2 - \lambda_1 - 2\lambda_2 - 3\lambda_3 + 1\lambda_4 + 2\lambda_5 = 0; \\ \lambda_1 + \lambda_2 + \lambda_3 - \lambda_4 - \lambda_5 = 0; \\ \lambda_1(1 - w_1 - w_2 + w_0) = 0; \quad \lambda_1 \geqslant 0; \quad (1 - w_1 - w_2 + w_0) \leqslant 0; \\ \lambda_2(1 - w_1 - 2w_2 + w_0) = 0; \quad \lambda_2 \geqslant 0; \quad (1 - w_1 - 2w_2 + w_0) \leqslant 0; \\ \lambda_3(1 - 2w_1 - 3w_2 + w_0) = 0; \quad \lambda_3 \geqslant 0; \quad (1 - 2w_1 - 3w_2 + w_0) \leqslant 0; \\ \lambda_4(1 + 3w_1 + w_2 - w_0) = 0; \quad \lambda_4 \geqslant 0; \quad (1 + 3w_1 + w_2 - w_0) \leqslant 0; \\ \lambda_5(1 + 4w_1 + 2w_2 - w_0) = 0; \quad \lambda_5 \geqslant 0; \quad (1 + 4w_1 + 2w_2 - w_0) \leqslant 0. \end{cases}$$

Мы уже немного знаем о решении этой задачи. Нам остается решить эту систему для всевозможных комбинаций условий и найти решение. Чтобы не выписывать все вручную, можно написать для этого, например, простую процедуру на MATLAB: все, что нужно — решать линейные системы уравнений и для решений проверять соответствующие условия системы.

Сейчас мы не будем перебирать все решения и прибегнем к хитрости. Несложно догадаться, что объекты $(1, 2)$ и $(4, 2)$ опорными не будут. А раз так, то для них отступ будет строго больше одного, а значит, из-за условий дополняющей нежесткости соответствующие им множители Лагранжа будут нулевыми: $\lambda_2 = \lambda_5 = 0$. Остальные объекты будут опорными и их отступы будут равны 1. Отсюда записывается система

$$\begin{cases} w_1 - \lambda_1 - \lambda_2 - 2\lambda_3 + 3\lambda_4 + 4\lambda_5 = 0; \\ w_2 - \lambda_1 - 2\lambda_2 - 3\lambda_3 + 1\lambda_4 + 2\lambda_5 = 0; \\ \lambda_1 + \lambda_2 + \lambda_3 - \lambda_4 - \lambda_5 = 0; \\ \lambda_1 \geqslant 0; \quad (1 - w_1 - w_2 + w_0) = 0; \\ \lambda_2 = 0; \quad (1 - w_1 - 2w_2 + w_0) \leqslant 0; \\ \lambda_3 \geqslant 0; \quad (1 - 2w_1 - 3w_2 + w_0) = 0; \\ \lambda_4 \geqslant 0; \quad (1 + 3w_1 + w_2 - w_0) = 0; \\ \lambda_5 = 0; \quad (1 + 4w_1 + 2w_2 - w_0) \leqslant 0. \end{cases}$$

Ее решение — $(w_0 = -1.5, w_1 = -1, w_2 = 0.5, \lambda_1 = 0.375, \lambda_3 = 0.25, \lambda_4 = 0.625)$.

Задача. *На рисунке 1 вы видите выборку из четырех двумерных объектов и двух классов. Величина $0 \leqslant h \leqslant 3$ — параметр. Мы будем решать задачу методом оптимальной разделяющей гиперплоскости:*

$$\begin{cases} \frac{1}{2}\|\boldsymbol{w}\|^2 \to \min_{\boldsymbol{w}, w_0}; \\ y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0) \geqslant 1, \quad i = 1, \ldots, 4. \end{cases} \tag{ex.1}$$

а) Для каких допустимых значений параметра h условия-неравенства в (ex.1) будут выполнимы?

б) Будет ли меняться наклон оптимальной разделяющей гиперплоскости при изменении параметра h? Как?

в) Как ширина разделяющей полосы, соответствующей оптимальной разделяющей гиперплоскости, выражается через параметр h?

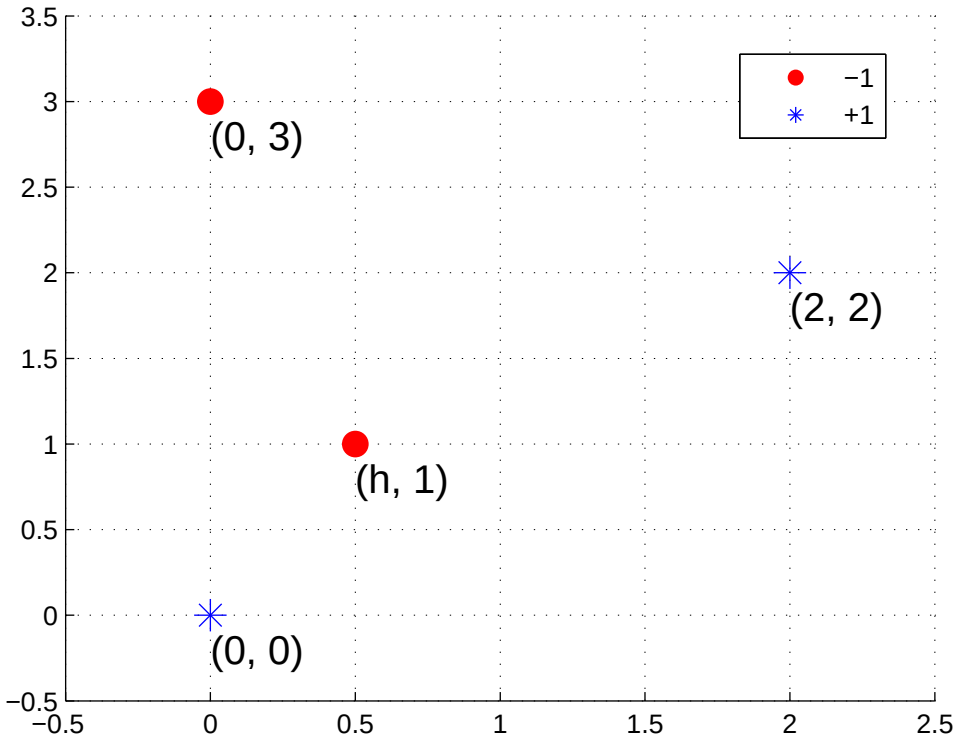

Рис. 1: Обучающие объекты.

2 Двойственная задача оптимизации.

Рассмотрим снова выпуклую задачу условной оптимизации:

$$\begin{cases} f(\boldsymbol{x}) \to \min_{\boldsymbol{x}}; \\ g_i(\boldsymbol{x}) \leqslant 0; \quad i = 1, \ldots, p; \qquad\qquad \boldsymbol{x} \in \mathbb{R}^n, \\ A\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}, \quad A \in \mathbb{R}^{q \times n}, \ \boldsymbol{b} \in \mathbb{R}^q, \end{cases} \tag{1}$$

где функции f, g_i выпуклы. Мы уже знаем, что любой локальный оптимум выпуклой задачи является одновременно ее глобальным оптимумом. В данном случае функция Лагранжа записывается в виде:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i g_i(\boldsymbol{x}) + \sum_{j=p+1}^{p+q} \lambda_j g_j(\boldsymbol{x}),$$

где $g_{p+1}, \ldots, g_{p+q}$ — линейные функции из системы (1).

Введем на основе функции Лагранжа *двойственную функцию*:

$$h(\boldsymbol{\lambda}) = \min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}).$$

Двойственной задачей к задаче оптимизации (1) называется следующая задача:

$$\begin{cases} h(\boldsymbol{\lambda}) \to \max_{\boldsymbol{\lambda}}; \\ \lambda_i \geqslant 0; \quad i = 1, \ldots, p. \end{cases} \tag{2}$$

Обозначим с помощью $\boldsymbol{x}^*$ оптимальное решение прямой задачи (1) и с помощью $\boldsymbol{\lambda}^*$ решение двойственной задачи (2). Поскольку для $\boldsymbol{\lambda}^*$ $\lambda_i^* \geqslant 0$, $i = 1, \ldots, p$, то справедливо следующее соотношение:

$$h(\boldsymbol{\lambda}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i^* g_i(\boldsymbol{x}) + \sum_{j=p+1}^{p+q} \lambda_j^* g_j(\boldsymbol{x}) \leqslant$$

$$\leqslant f(\boldsymbol{x}^*) + \sum_{i=1}^{p} \lambda_i^* \underbrace{g_i(\boldsymbol{x}^*)}_{\leqslant 0} + \sum_{j=p+1}^{p+q} \lambda_j^* \underbrace{g_j(\boldsymbol{x}^*)}_{=0} \leqslant f(\boldsymbol{x}^*).$$

В случае, когда f, g_i выпуклы, выполняется *условие сильной двойственности*, и оптимальные значения функционала в прямой и обратных задачах совпадают между собой:

$$f(\boldsymbol{x}^*) = h(\boldsymbol{\lambda}^*).$$

Еще одним важным свойством двойственной задачи является тот факт, что она всегда *выпукла* (двойственная функция h является всегда вогнутой), даже если прямая задача выпуклой не была.

Задача. *Рассмотрим следующую выпуклую задачу условной оптимизации:*

$$\begin{cases} \langle \boldsymbol{x}, \boldsymbol{x} \rangle \to \min_{\boldsymbol{x}}; \\ A\boldsymbol{x} = \boldsymbol{b}, \quad A \in \mathbb{R}^{p \times n},\ \boldsymbol{b} \in \mathbb{R}^p, \end{cases} \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

Выпишите двойственную функцию и двойственную задачу.

Решение. Запишем функцию Лагранжа для нашей задачи:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \langle \boldsymbol{x}, \boldsymbol{x} \rangle + \boldsymbol{\lambda}^\top (A\boldsymbol{x} - \boldsymbol{b}).$$

По определению двойственная функция

$$h(\boldsymbol{\lambda}) = \min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \min_{\boldsymbol{x}} \big\{ \langle \boldsymbol{x}, \boldsymbol{x} \rangle + \boldsymbol{\lambda}^\top (A\boldsymbol{x} - \boldsymbol{b}) \big\}.$$

Несложно заметить, что функция Лагранжа выпуклая, значит минимум достигается в точке, где градиент равен нулю:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{0}. \tag{3}$$

Выведите самостоятельно тождество $\nabla_{\boldsymbol{x}} \langle \boldsymbol{\lambda}, A\boldsymbol{x} \rangle = A^\top \boldsymbol{\lambda}$. Пользуясь им, получаем из (3) следующее уравнение:

$$2\boldsymbol{x} + A^\top \boldsymbol{\lambda} = 0,$$

которое дает $\boldsymbol{x} = -\frac{1}{2} A^\top \boldsymbol{\lambda}$. Таким образом, двойственная функция запишется в виде:

$$h(\boldsymbol{\lambda}) = -\frac{1}{4} \lambda^\top A A^\top \boldsymbol{\lambda} - \boldsymbol{\lambda}^\top \boldsymbol{b}.$$

Это вогнутая функция. Двойственная задача запишется в виде

$$h(\boldsymbol{\lambda}) = -\frac{1}{4} \lambda^\top A A^\top \boldsymbol{\lambda} - \boldsymbol{\lambda}^\top \boldsymbol{b} \to \max_{\boldsymbol{\lambda}}.$$

3 Линейно неразделимая выборка. “Soft-margin”.

На прошлом семинаре мы рассмотрели случай линейно разделимых классов в обучающей выборке и оптимальную разделяющую гиперплоскость. Мы рассматривали классификатор вида

$$a(\boldsymbol{x}) = \operatorname{sgn} y(\boldsymbol{x}) = \operatorname{sgn}(\langle \boldsymbol{w}, \boldsymbol{x} \rangle - w_0).$$

Что, если не удается построить гиперплоскость, которая без ошибок разделяет классы обучающей выборки? Один из подходов к решению такой сложности — смягчение ограничений $y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0) \geqslant 1$, поскольку они в этом случае не могут быть выполнены. Заменим их условиями $y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0) \geqslant 1 - \xi_i$, где мы ввели *ослабляющие коэффициенты* $\xi_i \geqslant 0$, $i = 1 \ldots, \ell$.

Задача. *Опишите точки, удовлетворяющие этим смягченным ограничениям, в трех следующих случаях: а) $\xi_i = 0$, б) $0 < \xi_i \leqslant 1$, в) $\xi_i > 1$.*

Точки, для которых смягченные условия выполнены с $\xi_i = 0$ имеют отступ не меньше 1, а значит, правильно классифицированы. Если $0 < \xi_i \leqslant 1$, то отступ точки изменяется в пределах от 0 до 1, и эти точки тоже классифицированы без ошибок. В случае $\xi_i > 1$ отступ точки отрицателен и на ней классификатор ошибается. Запишем новую задачу оптимизации:

$$\begin{cases} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{\ell} \xi_i \to \min_{\boldsymbol{w}, w_0, \xi_i}; \\ y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0) \geqslant 1 - \xi_i, \ i = 1, \ldots, \ell; \\ \xi_i \geqslant 0, \ i = 1, \ldots, \ell. \end{cases} \tag{4}$$

Мы ввели коэффициент регуляризации $C \geqslant 0$, который определяет компромисс между количеством ошибок на обучающей выборке и величиной нормы вектора $\boldsymbol{w}$. Коэффициент регуляризации задается до решения задачи и обучение *метода опорных векторов* (SVM) состоит в решении задачи (4).

Задача. *Покажите, что для задачи условной оптимизации (4) можно записать эквивалентную задачу безусловной оптимизации, воспользовавшись условиями $\xi_i \geqslant \max\big(0, 1 - y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0)\big)$, которые выполнены для всех $i = 1, \ldots, \ell$. Какую связь этой задачи безусловной оптимизации с минимизацией эмпирического риска вы видите? Какая функция потерь используется? Как эта задача связана с решением логистической регрессии?*

Задача. *Запишите двойственную функцию и двойственную задачу для (4).*

Задача (4) является выпуклой. Запишем для нее функцию Лагранжа:

$$L(\boldsymbol{w}, w_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{\ell} \xi_i + \sum_{i=1}^{\ell} \lambda_i\big(1 - \xi_i - y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0)\big) - \sum_{j=1}^{\ell} \mu_j \xi_j.$$

Построим двойственную функцию и запишем двойственную задачу для системы (4).

Запишем условия на минимум функции Лагранжа по $(\boldsymbol{w}, w_0, \boldsymbol{\xi})$:

$$\nabla_{\boldsymbol{w}} L = \boldsymbol{w} - \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i = 0 \qquad \Rightarrow \qquad \boldsymbol{w} = \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i; \tag{5}$$

$$\frac{\partial L}{\partial w_0} = 0 \qquad \Rightarrow \qquad \sum_{i=1}^{\ell} y_i \lambda_i = 0; \tag{6}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \qquad \Rightarrow \qquad C - \lambda_i - \mu_i = 0. \tag{7}$$

С учетом соотношений (5)–(6) двойственная функция запишется в виде:

$$h(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2} \left\langle \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i, \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i \right\rangle + \sum_{i=1}^{\ell} \lambda_i - \sum_{j=1}^{\ell} \lambda_j y_j \left\langle \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i, \boldsymbol{x}_j \right\rangle =$$

$$= \frac{1}{2} \left\langle \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i, \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i \right\rangle + \sum_{i=1}^{\ell} \lambda_i - \left\langle \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i, \sum_{j=1}^{\ell} \lambda_j y_j \boldsymbol{x}_j \right\rangle =$$

$$= \sum_{i=1}^{\ell} \lambda_i - \frac{1}{2} \left\langle \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i, \sum_{i=1}^{\ell} \lambda_i y_i \boldsymbol{x}_i \right\rangle = \sum_{i=1}^{\ell} \lambda_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \lambda_i y_i \lambda_j y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_i \rangle.$$

Запишем двойственную задачу:

$$\begin{cases} \sum_{i=1}^{\ell} \lambda_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \lambda_i y_i \lambda_j y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_i \rangle \to \max_{\boldsymbol{\lambda}}; \\ \sum_{i=1}^{\ell} y_i \lambda_i = 0; \\ 0 \leqslant \lambda_i \leqslant C, \ i = 1, \ldots, \ell. \end{cases} \tag{8}$$

Поскольку прямая задача (4) является выпуклой, то оптимальные значения функционалов прямой и двойственной задачи совпадают. Кроме того, оптимальное решение $(\boldsymbol{w}^*, w_0^*)$ прямой задачи может быть выражено через оптимальное решение двойственной задачи $\boldsymbol{\lambda}^*$ аналитически с помощью тождества (5) и условий дополняющей нежесткости (из условий Куна–Таккера) для задачи (4):

$$\lambda_i \big(1 - \xi_i - y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0)\big) = 0, \quad i = 1, \ldots, \ell.$$

Задача. *Опишите три категории точек обучающей выборки, которым соответствуют случаи а) $\lambda_i = 0$, б) $0 < \lambda_i < C$, в) $\lambda_i = C$.*

1. $\big(1 - \xi_i - y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0)\big) < 0$.
   В этом случае $\lambda_i = 0$. Из условия (7) получаем, что $\mu_i = C$. Условия дополняющей нежесткости для задачи (4) ведут к тому, что в этом случае $\xi_i = 0$. Таким образом, эти объекты классифицируются правильно и для них отступ больше 1.

2. $0 < \lambda_i < C$.
   В этом случае $\big(1 - \xi_i - y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - w_0)\big) = 0$. Похожие соображения указывают на то, что $\xi_i = 0$. Эти точки классифицируются правильно и их отступ равен 1.

3. $\xi_i > 0$.
   В этом случае $\lambda_i = C$, $\big(1 - \xi_i - y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle - w_0)\big) = 0$. Отступы этих точек меньше 1 и среди них могут быть точки, которые неправильно классифицируются.

Точки для которых $\lambda_i > 0$ называются опорными. Порог w_0 может быть найден с помощью опорных точек категории 2.

Обратим внимание, что обучение SVM сводится к решению прямой задачи (4) либо двойственной задачи (8), которая имеет простые граничные условия и содержит всего ℓ переменных, в то время как прямая — $\ell + n + 1$ переменных.

Полученный классификатор с учетом (5) имеет вид

$$a(\boldsymbol{x}) = \operatorname{sgn}\left\{\sum_{i=1}^{\ell} y_i \lambda_i \langle \boldsymbol{x}_i, \boldsymbol{x}\rangle - w_0\right\}.$$

Задача. *Предположим, мы решили двойственную задачу (8) и оказалось, что $\lambda_i^* \in \{0, C\}$, $i = 1, \ldots, \ell$. Как вычислить оптимальный порог w_0^*, входящий в оптимальное решение прямой задачи (4)?*

4 Ядровая функция.

Мы убедились, что двойственная задача (8) и итоговый классификатор зависят только от скалярных произведений объектов обучающих выборки. Этот факт может быть эффективно использован в еще одном способе борьбы со случаем линейно неразделимой выборки — так называемом *ядерном переходе*.

Представим, что существует отображение $\psi\colon \mathbb{X} \to \mathcal{H}$, такое что наша обучающая выборка в пространстве $\mathcal{H}$ линейно разделима. Пространство $\mathcal{H}$ будем называть *спрямляющим*. Описанный аппарат метода опорных векторов может быть применен в новом пространстве для построения в нем линейного классификатора, который не ошибается на объектах обучающей выборки. В исходном пространстве объектов полученный классификатор в общем случае соответствует нелинейной разделяющей поверхности. Таким образом мы описали эффективный способ работы в линейно неразделимом случае.

Поскольку двойственная задача (8) и решающий классификатор в SVM (как в случае soft-margin так и в случае hard-margin) выражается через скалярные произведения объектов, нам достаточно заменить $\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle$ на $\langle \psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)\rangle$. В остальном метод останется без изменений.

Введем для спрямляющего пространства $\mathcal{H}$, которому соответствует отображение $\psi = \psi_{\mathcal{H}}$, *ядровую функцию*: $K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \langle \psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)\rangle$. Итак, мы будем называть *ядрами* функции $K\colon \mathbb{X}^2 \to \mathbb{R}$, для которых существует такое отображение обучающей выборки в новое пространство $\mathcal{H}$ ($\psi\colon \mathbb{X} \to \mathcal{H}$), что скалярное произведение в новом пространстве выражается этой функцией ($K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)\rangle$).

Теорема Мерсера утверждает, что следующие условия являются необходимыми и достаточными для того, чтобы функция f являлась ядром:

1. Симметричность: $f(\boldsymbol{x}_i, \boldsymbol{x}_j) = f(\boldsymbol{x}_j, \boldsymbol{x}_i)$;
2. Неотрицательно определена: $\int_{\mathbb{X}}\int_{\mathbb{X}} f(\boldsymbol{x}, \boldsymbol{x}')g(\boldsymbol{x})g(\boldsymbol{x}')d\boldsymbol{x}d\boldsymbol{x}' \geqslant 0$ для всех функций g, имеющих конечный второй момент $\int_{\mathbb{X}} g^2(\boldsymbol{x})d\boldsymbol{x}$.

Итак, выбрав спрямляющее пространство $\mathcal{H}$ и соответствующую ему ядровую функцию $K_{\mathcal{H}}$, мы можем решать двойственную задачу (8), заменив в ней $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$ на $K_{\mathcal{H}}(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Отметим, что нам вовсе необязательно для этого в явном виде искать соответствующее преобразование признаков и само спрямляющее пространство. Нам достаточно задать функцию ядра. Это ведет к так называемым методам *беспризнакового распознавания*.