Машинное обучение

Лекция 7. Линейные модели, SVM

МФТИ 2018

Материалы: В. Кантор, К. Воронцов

План

1. Логистическая регрессия

2. Метод Опорных Векторов (SVM)

3. О применении линейных моделей

1. Логистическая регрессия

Линейная регрессия

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

Линейная регрессия

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

$$Q = \sum_{i=1}^{N} L\big(y_i, a(x_i)\big)$$

Линейная регрессия

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

$$Q = \sum_{i=1}^{N} L\big(y_i, a(x_i)\big)$$

$$L\big(y_i, a(x_i)\big) = (y_i - a(x_i))^2$$

Линейная регрессия

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

$$Q = \sum_{i=1}^{N} L\big(y_i, a(x_i)\big)$$

$$L\big(y_i, a(x_i)\big) = (y_i - a(x_i))^2$$

$$L\big(y_i, a(x_i)\big) = |y_i - a(x_i)|$$

Линейная классификация

$$a(x) = \begin{cases} 1, \text{если } f(x) > 0 \\ -1, \text{если } f(x) > 0 \end{cases}$$

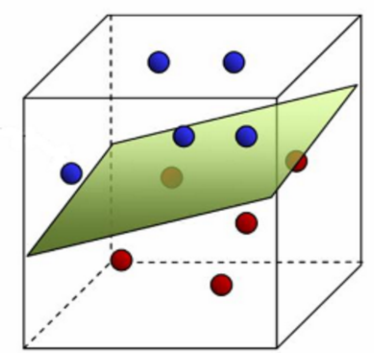
$$f(x) = w_0 + w_1 x_1 + \dots + w_n x_n$$

Линейная классификация

$$a(x) = \begin{cases} 1, \text{если } f(x) > 0 \\ -1, \text{если } f(x) > 0 \end{cases}$$

$$f(x) = w_0 + w_1 x_1 + \cdots + w_n x_n = w_0 + \langle w, x \rangle$$

Геометрическая интерпретация:
разделяем классы плоскостью

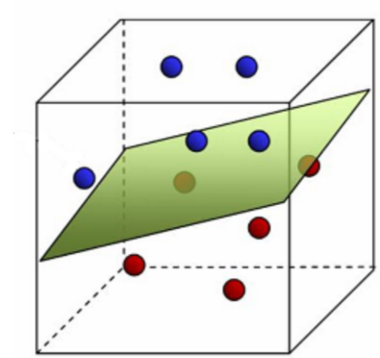


Линейная классификация

$$a(x) = \begin{cases} 1, \text{если } f(x) > 0 \\ -1, \text{если } f(x) > 0 \end{cases}$$

$$f(x) = \langle w, x \rangle$$

Геометрическая интерпретация:
разделяем классы плоскостью



Отступ (margin)

Отступом алгоритма $a(x) = sign\{f(x)\}$ на объекте x_i называется величина $M_i = y_i f(x_i)$

(y_i - класс, к которому относится x_i)

$$M_i \leq 0 \Leftrightarrow y_i \neq a(x_i)$$
$$M_i > 0 \Leftrightarrow y_i = a(x_i)$$

Функция потерь

$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0]$$

Функция потерь

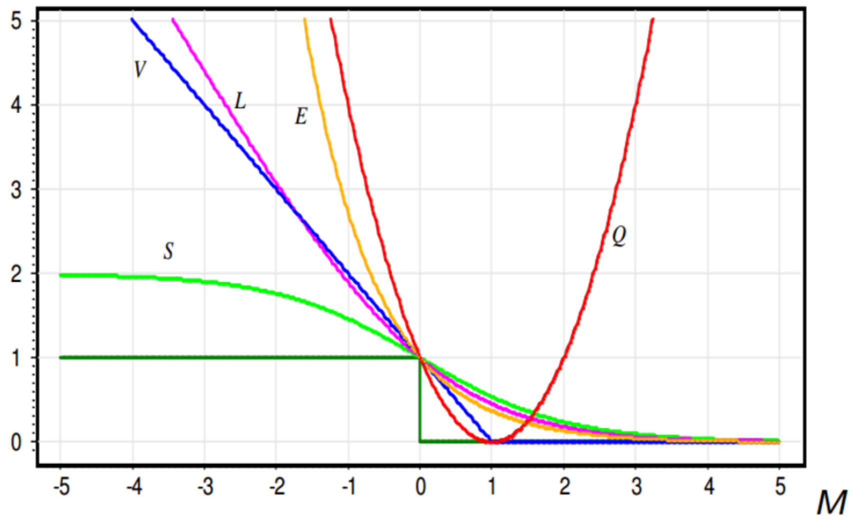
$$Q(w) = \sum_{i=1}^{\ell} \left[M_i(w) < 0 \right] \leqslant \widetilde{Q}(w) = \sum_{i=1}^{\ell} \mathscr{L}\big(M_i(w)\big) \to \min_{w};$$

Функция эмпирического риска

Функция потерь

Функция потерь

$$Q(w) = \sum_{i=1}^{\ell} \left[M_i(w) < 0 \right] \leqslant \widetilde{Q}(w) = \sum_{i=1}^{\ell} \mathscr{L}\left(M_i(w)\right) \to \min_{w};$$



$Q(M) = (1 - M)^2$
$V(M) = (1 - M)_+$
$S(M) = 2(1 + e^{M})^{-1}$
$L(M) = \log_2(1 + e^{-M})$
$E(M) = e^{-M}$

Логистическая регрессия

$y_i \in \{0, 1\}$

$$Q = -\sum_{i=1}^{\ell} y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \to \min_w$$

$$p_i = \sigma(-\langle w, x_i \rangle) = \frac{1}{1 + e^{-\langle w, x_i \rangle}}$$

Логистическая регрессия

$y_i \in \{0, 1\}$

$$Q = -\sum_{i=1}^{\ell} y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \to \min_w$$

$$p_i = \sigma(\langle w, x_i \rangle) = \frac{1}{1 + e^{-\langle w, x_i \rangle}} = P(y = 1 | x)$$

Логистическая регрессия

$y_i \in \{0, 1\}$

$$Q = -\sum_{i=1}^{\ell} y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \to \min_w$$

$$p_i = \sigma(\langle w, x_i \rangle) = \frac{1}{1 + e^{-\langle w, x_i \rangle}} = P(y = 1|x)$$

Как правило, добавляется ℓ_1 или ℓ_2-регуляризация, а оптимизационная задача решается с помощью SGD или метода Ньютона-Рафсона

Эквивалентность оптимизационных задач

$$Q = -\sum_{i=1}^{\ell} y_i \ln \frac{1}{1+e^{-\langle w,x_i\rangle}} + (1-y_i)\ln\frac{1}{1+e^{\langle w,x_i\rangle}} \to \min_w$$

$$-y_i \ln \frac{1}{1+e^{-\langle w,x_i\rangle}} - (1-y_i)\ln\frac{1}{1+e^{\langle w,x_i\rangle}} = \begin{cases} \ln\left(1+e^{-\langle w,x_i\rangle}\right), y_i = 1 \\ \ln\left(1+e^{\langle w,x_i\rangle}\right), y_i = 0 \end{cases}$$

$$Q = \sum_{i=1}^{\ell} \underbrace{\ln\left(1+e^{-y_i\langle w,x_i\rangle}\right)}_{L(M) = \ln(1+e^{-M_i})} \to \min_w \qquad y_i \in \{-1, 1\}$$

2. Метод Опорных Векторов (SVM)

Метод опорных векторов

Линейный классификатор:

$$a(x) = sign(< w, x > -w_0)$$

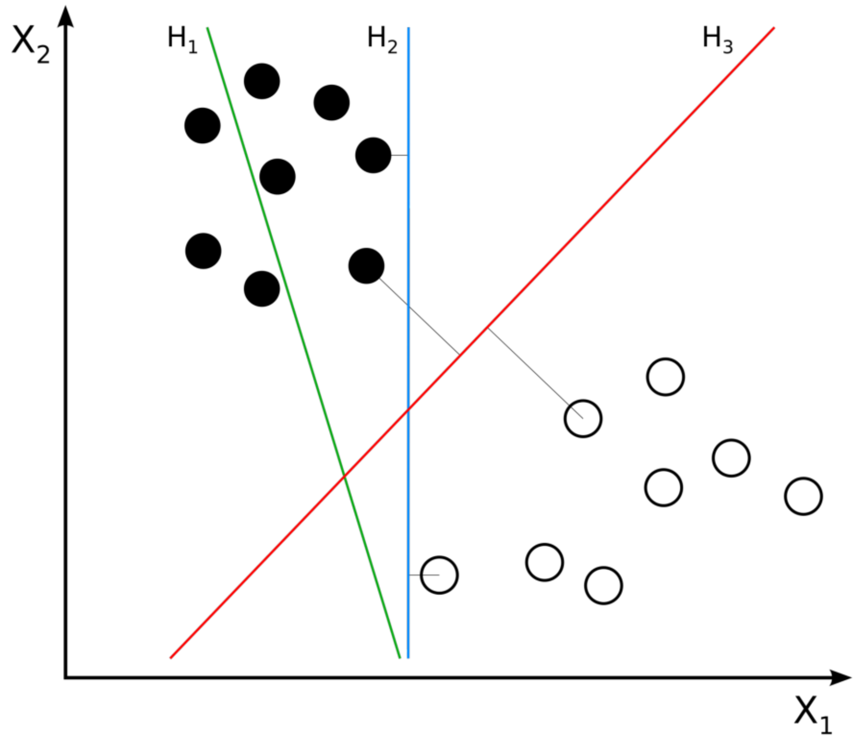
Использующий кусочно-линейную функцию потерь и L2-регуляризатор:

$$\sum_{i=1}^{l} \underbrace{\boxed{L(M_i)}}_{\text{функция потерь}} + \gamma \underbrace{\boxed{\|w\|^2}}_{\text{квадратичный регуляризатор}} \to \min_w$$

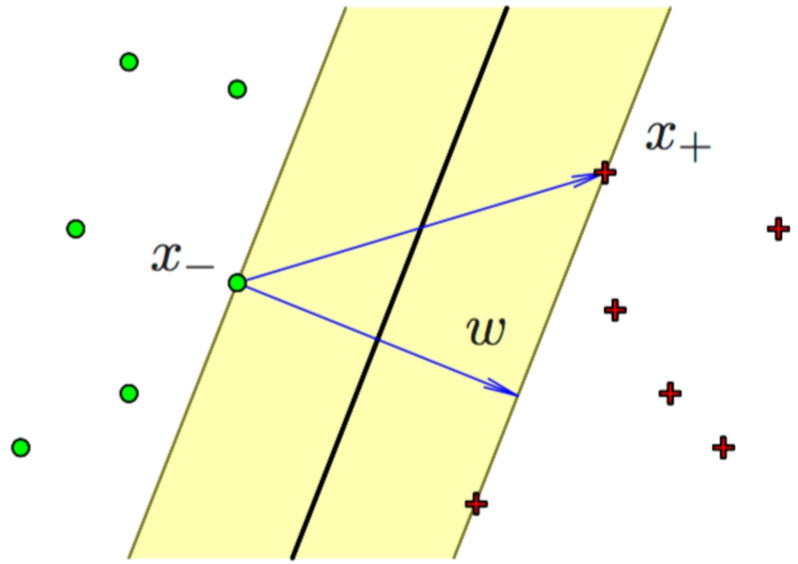
кусочно-линейная функция потерь:

$$L(M_i) = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

Построение разделяющей гиперплоскости

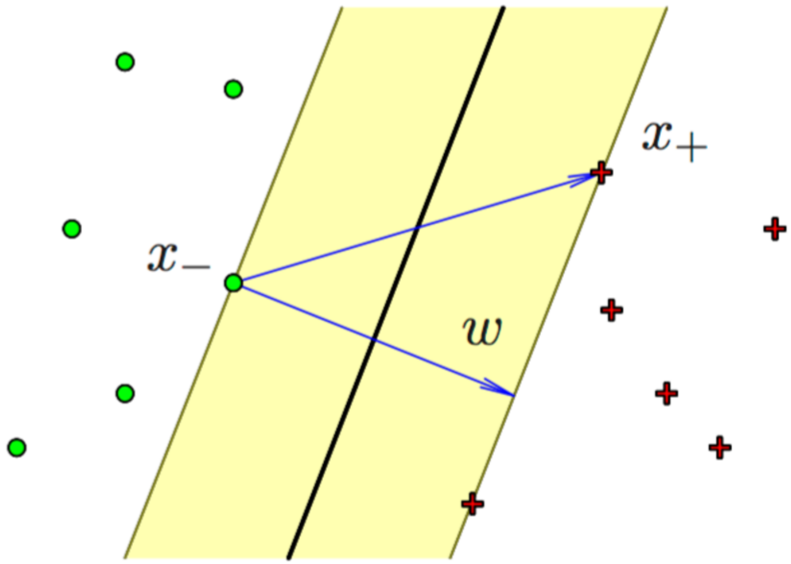


Разделяющая полоса



$$a(x) = \operatorname{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0\right) = \operatorname{sign}\big(\langle w, x\rangle - w_0\big)$$

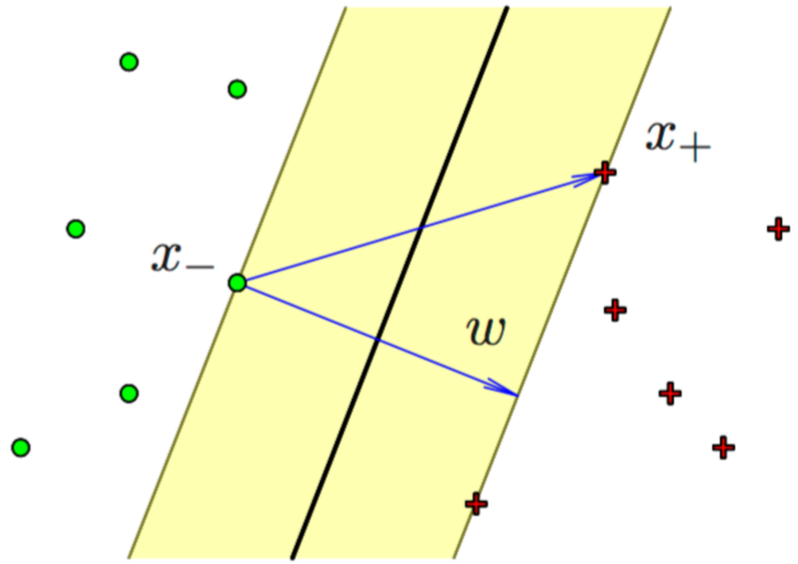
Разделяющая полоса



$$\min_{i=1,\dots,\ell} y_i\big(\langle w, x_i\rangle - w_0\big) = 1$$

$$a(x) = \mathrm{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0\right) = \mathrm{sign}\big(\langle w, x\rangle - w_0\big)$$

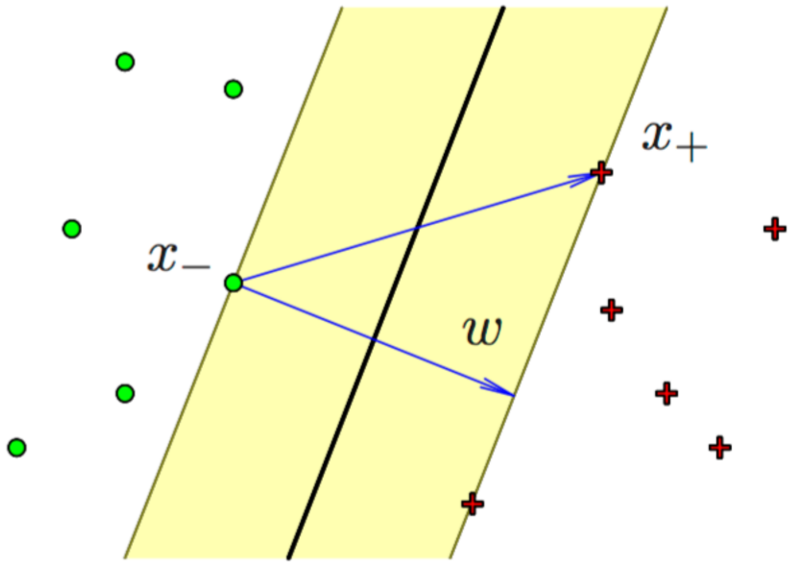
Ширина разделяющей полосы



$$\min_{i=1,\dots,\ell} y_i\big(\langle w, x_i\rangle - w_0\big) = 1$$

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

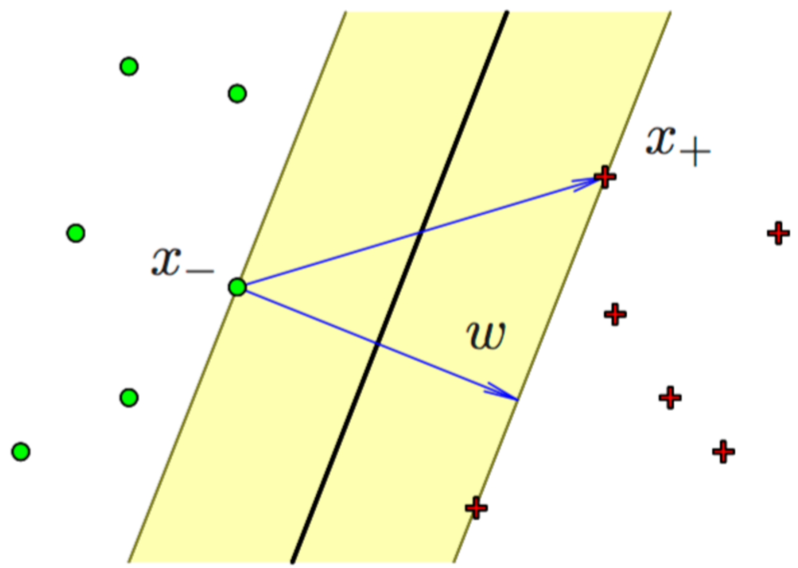
Ширина разделяющей полосы



$$\min_{i=1,\dots,\ell} y_i\bigl(\langle w, x_i\rangle - w_0\bigr) = 1.$$

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

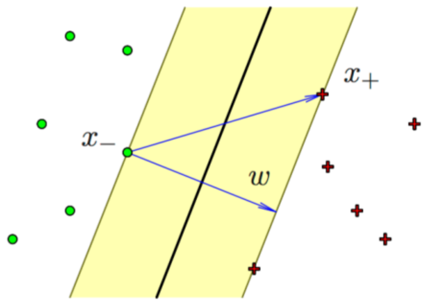
Ширина разделяющей полосы



$$\min_{i=1,\dots,\ell} y_i\big(\langle w, x_i\rangle - w_0\big) = 1$$

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+\rangle - \langle w, x_-\rangle}{\|w\|} = \frac{(w_0+1) - (w_0-1)}{\|w\|} = \frac{2}{\|w\|}$$

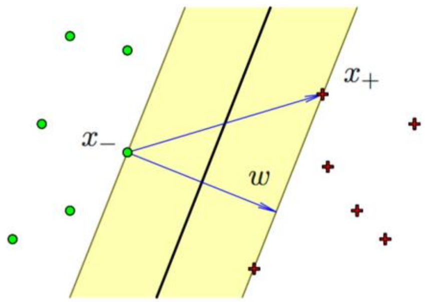
Ширина разделяющей полосы



$$\min_{i=1,\dots,\ell} y_i\big(\langle w, x_i\rangle - w_0\big) = 1$$

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+\rangle - \langle w, x_-\rangle}{\|w\|} = \frac{(w_0 + 1) - (w_0 - 1)}{\|w\|} = \frac{2}{\|w\|}$$

Максимизация зазора

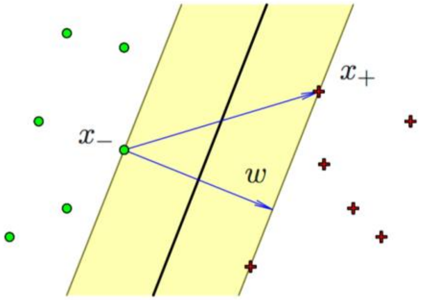


$$\min_{i=1,\dots,\ell} y_i\big(\langle w, x_i\rangle - w_0\big) = 1$$

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+\rangle - \langle w, x_-\rangle}{\|w\|} = \frac{(w_0+1)-(w_0-1)}{\|w\|} = \frac{2}{\|w\|}$$

$$\begin{cases} \langle w, w\rangle \to \min; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1, \quad i = 1, \dots, \ell \end{cases}$$

Случай линейно неразделимой выборки



$$\begin{cases} \langle w, w \rangle \to \min; \\ y_i\big(\langle w, x_i \rangle - w_0\big) \geqslant 1, \quad i = 1, \dots, \ell. \end{cases}$$

$$\begin{cases} \dfrac{1}{2}\langle w, w \rangle + C\displaystyle\sum_{i=1}^{\ell} \xi_i \to \min_{w, w_0, \xi}; \\ y_i\big(\langle w, x_i \rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geqslant 0, \quad i = 1, \dots, \ell. \end{cases}$$

Оптимизационная задача в SVM

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min\limits_{w, w_0, \xi}; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geqslant 0, \quad i = 1, \dots, \ell. \end{cases}$$

Причем здесь линейный классификатор
в привычном нам виде?

Безусловная оптимизационная задача в SVM

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min\limits_{w,w_0,\xi}; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1,\dots,\ell; \\ \xi_i \geqslant 0, \quad i = 1,\dots,\ell. \end{cases}$$

Напоминание:

$$M_i = y_i\big(\langle w, x_i\rangle - w_0\big)$$

отступ на i-том объекте

Безусловная оптимизационная задача в SVM

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min\limits_{w,w_0,\xi}; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geqslant 0, \quad i = 1, \dots, \ell. \end{cases}$$

Напоминание:

$$M_i = y_i\big(\langle w, x_i\rangle - w_0\big)$$

отступ на i-том объекте

$$\xi_i \geqslant 0$$

$$\xi_i \geqslant 1 - M_i$$

$$\sum_{i=1}^{l} \xi_i \to min$$

Безусловная оптимизационная задача в SVM

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min_{w,w_0,\xi}; \\ y_i\bigl(\langle w, x_i\rangle - w_0\bigr) \geqslant 1 - \xi_i, \quad i = 1,\dots,\ell; \\ \xi_i \geqslant 0, \quad i = 1,\dots,\ell. \end{cases}$$

Напоминание:

$$M_i = y_i\bigl(\langle w, x_i\rangle - w_0\bigr)$$

отступ на i-том объекте

$$\xi_i \geqslant 0$$

$$\xi_i \geqslant 1 - M_i \quad \Longrightarrow \xi_i = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

$$\sum_{i=1}^{l}\xi_i \to min$$

Безусловная оптимизационная задача в SVM

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min\limits_{w,w_0,\xi}; \\ y_i\bigl(\langle w, x_i\rangle - w_0\bigr) \geqslant 1 - \xi_i, \quad i = 1,\dots,\ell; \\ \xi_i \geqslant 0, \quad i = 1,\dots,\ell. \end{cases}$$

Напоминание:

$$M_i = y_i\bigl(\langle w, x_i\rangle - w_0\bigr)$$

отступ на i-том объекте

$$\xi_i \geqslant 0$$

$$\xi_i \geqslant 1 - M_i \quad \Longrightarrow \xi_i = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

$$\sum_{i=1}^{l} \xi_i \to min$$

$$Q(w, w_0) = \sum_{i=1}^{\ell}\bigl(1 - M_i(w, w_0)\bigr)_+ + \frac{1}{2C}\|w\|^2 \to \min_{w,w_0}$$

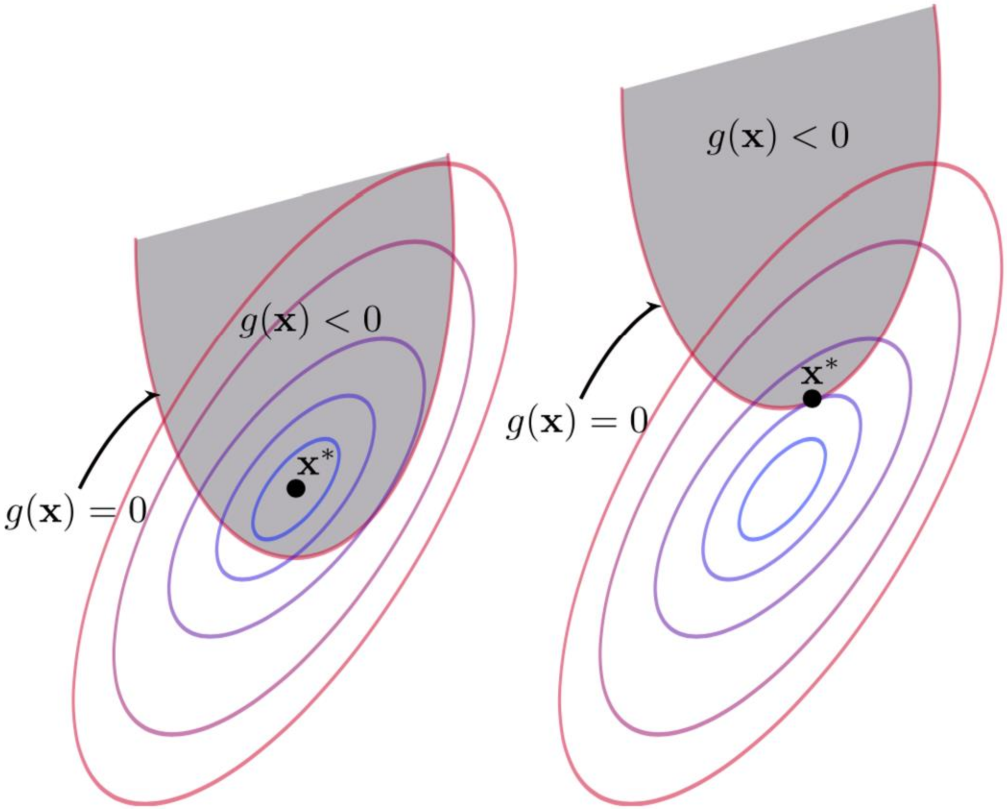
Безусловная оптимизационная задача в SVM

$$Q(w, w_0) = \sum_{i=1}^{\ell} \bigl(1 - M_i(w, w_0)\bigr)_+ + \frac{1}{2C}\|w\|^2 \;\to\; \min_{w, w_0}$$

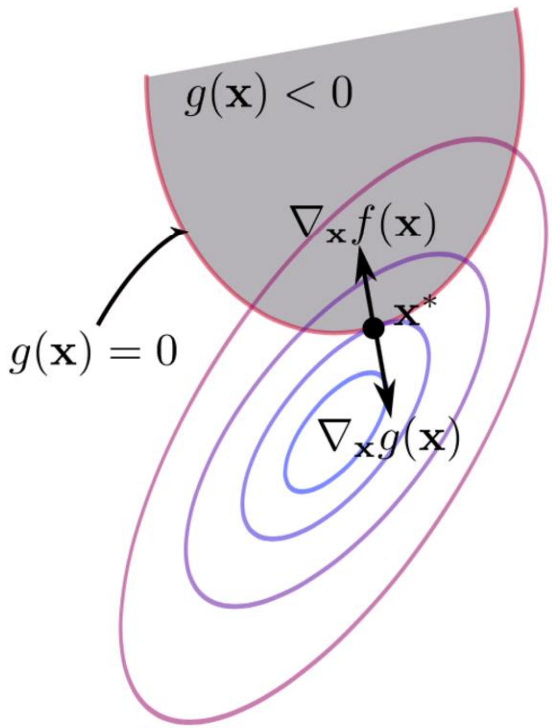
Напоминание: теорема Каруша-Куна-Таккера

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

Напоминание: теорема Каруша-Куна-Таккера



Напоминание: теорема Каруша-Куна-Таккера



Двойственная задача в SVM

$$
\begin{cases}
\frac{1}{2}\langle w, w\rangle + C\sum\limits_{i=1}^{\ell}\xi_i \to \min\limits_{w,w_0,\xi};\\
y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell;\\
\xi_i \geqslant 0, \quad i = 1, \dots, \ell.
\end{cases}
$$

Двойственная задача в SVM

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min\limits_{w, w_0, \xi}; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geqslant 0, \quad i = 1, \dots, \ell. \end{cases}$$

$$M_i(w, w_0) = y_i\big(\langle w, x_i\rangle - w_0\big)$$

Двойственная задача в SVM

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min_{w,w_0,\xi}; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geqslant 0, \quad i = 1, \dots, \ell. \end{cases}$$

$$M_i(w, w_0) = y_i\big(\langle w, x_i\rangle - w_0\big)$$

$$\mathscr{L}(w, w_0, \xi; \lambda, \eta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}\xi_i - \sum_{i=1}^{\ell}\lambda_i\big(M_i(w, w_0) - 1 + \xi_i\big) - \sum_{i=1}^{\ell}\xi_i\eta_i =$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{\ell}\lambda_i\big(M_i(w, w_0) - 1\big) - \sum_{i=1}^{\ell}\xi_i\big(\lambda_i + \eta_i - C\big),$$

Двойственная задача в SVM

$$\mathscr{L}(w, w_0, \xi; \lambda, \eta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}\xi_i - \sum_{i=1}^{\ell}\lambda_i\bigl(M_i(w, w_0) - 1 + \xi_i\bigr) - \sum_{i=1}^{\ell}\xi_i\eta_i =$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{\ell}\lambda_i\bigl(M_i(w, w_0) - 1\bigr) - \sum_{i=1}^{\ell}\xi_i\bigl(\lambda_i + \eta_i - C\bigr),$$

$$\begin{cases} \mathscr{L}(w, w_0, \xi; \lambda, \eta) \ \to \ \min\limits_{w, w_0, \xi} \ \max\limits_{\lambda, \eta}; \\ \xi_i \geqslant 0, \quad \lambda_i \geqslant 0, \quad \eta_i \geqslant 0, \quad i = 1, \dots, \ell; \\ \lambda_i = 0 \ \text{ либо } \ M_i(w, w_0) = 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \eta_i = 0 \ \text{ либо } \ \xi_i = 0, \quad i = 1, \dots, \ell; \end{cases}$$

Опорные векторы в SVM

$$\mathscr{L}(w, w_0, \xi; \lambda, \eta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}\xi_i - \sum_{i=1}^{\ell}\lambda_i\bigl(M_i(w, w_0) - 1 + \xi_i\bigr) - \sum_{i=1}^{\ell}\xi_i\eta_i =$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{\ell}\lambda_i\bigl(M_i(w, w_0) - 1\bigr) - \sum_{i=1}^{\ell}\xi_i\bigl(\lambda_i + \eta_i - C\bigr),$$

$$\frac{\partial\mathscr{L}}{\partial w} = w - \sum_{i=1}^{\ell}\lambda_i y_i x_i = 0 \qquad \Longrightarrow \qquad w = \sum_{i=1}^{\ell}\lambda_i y_i x_i;$$

Двойственная задача: итог

$$\frac{\partial \mathscr{L}}{\partial \xi_i} = -\lambda_i - \eta_i + C = 0 \qquad \Longrightarrow \qquad \eta_i + \lambda_i = C, \quad i = 1, \dots, \ell.$$

$$\begin{cases} -\mathscr{L}(\lambda) = -\sum\limits_{i=1}^{\ell} \lambda_i + \frac{1}{2} \sum\limits_{i=1}^{\ell} \sum\limits_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle \ \to \ \min\limits_{\lambda}; \\ 0 \leqslant \lambda_i \leqslant C, \quad i = 1, \dots, \ell; \\ \sum\limits_{i=1}^{\ell} \lambda_i y_i = 0. \end{cases}$$

Двойственная задача: итог

Типизация объектов:

1. $\lambda_i = 0;\ \ \eta_i = C;\ \ \xi_i = 0;\ \ M_i \geqslant 1.$
   — периферийные (неинформативные) объекты.
2. $0 < \lambda_i < C;\ \ 0 < \eta_i < C;\ \ \xi_i = 0;\ \ M_i = 1.$
   — опорные граничные объекты.
3. $\lambda_i = C;\ \ \eta_i = 0;\ \ \xi_i > 0;\ \ M_i < 1.$
   — опорные-нарушители.

Определение

Объект x_i называется *опорным*, если $\lambda_i \neq 0$.

Двойственная задача: итог

$$a(x) = sign\left(\sum_i \lambda_i y_i < x_i, x > -w_0\right)$$

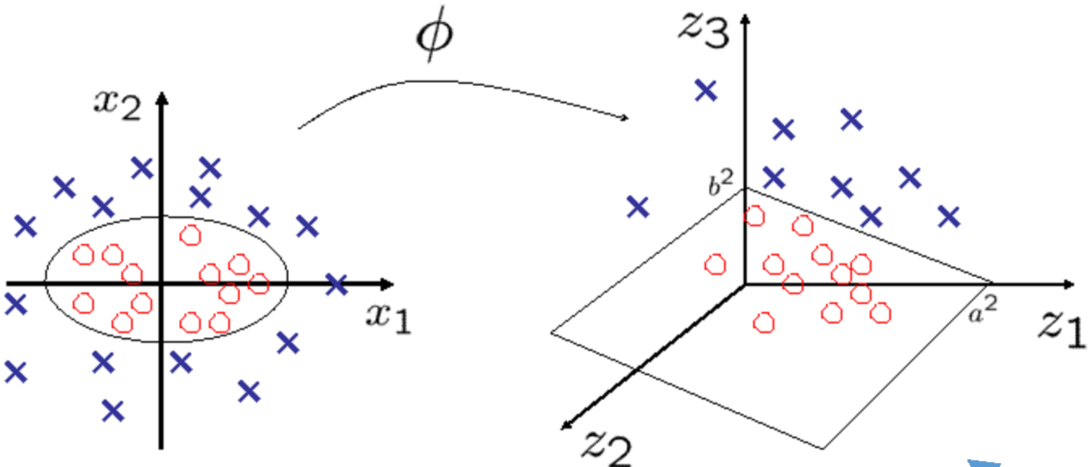
$$w = \sum_i \lambda_i y_i x_i$$

$$w_0 = Med(y_i - < w, x_i > | \lambda_i > 0)$$

Ключевые моменты

- Метод опорных векторов – линейный классификатор с кусочно-линейной функцией потерь (hinge loss) и L2-регуляризатором
- Придуман метод был из соображений максимизации зазора между классами
- В случае линейно разделимой выборки это означает просто максимизацию ширины разделяющей полосы
- А в случае линейно неразделимой выборки просто добавляется возможность попадания объектов в полосу и штрафы за эти попадания

Добавление новых признаков



$$\phi : (x_1, x_2) \longrightarrow (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

$$\left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 = 1 \longrightarrow \frac{z_1}{a^2} + \frac{z_3}{b^2} = 1$$

Спрямляющее пространство

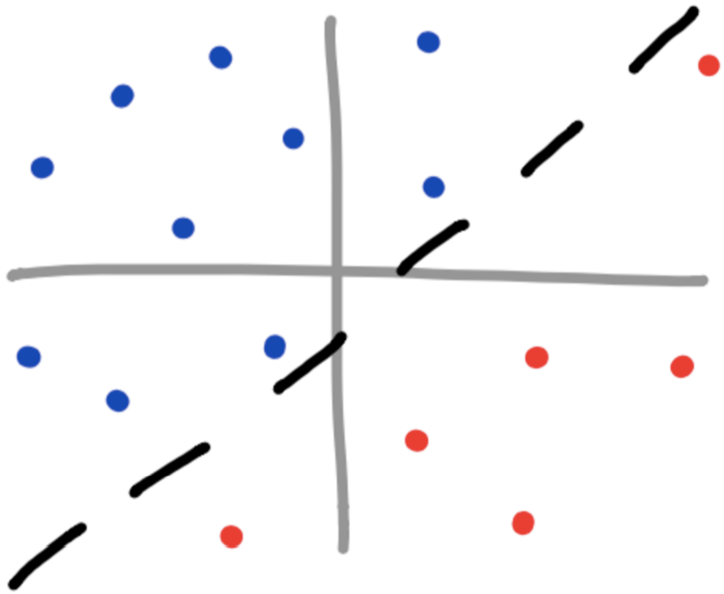
Kernel Trick

$$\begin{matrix} x \mapsto \phi(x) \\ w \mapsto \phi(w) \end{matrix} \implies < w, x > \mapsto < \phi(w), \phi(x) >$$

Можно не делать преобразование признаков явно, а вместо скалярного произведения $< w, x >$ использовать функцию K(w, x), представимую в виде:

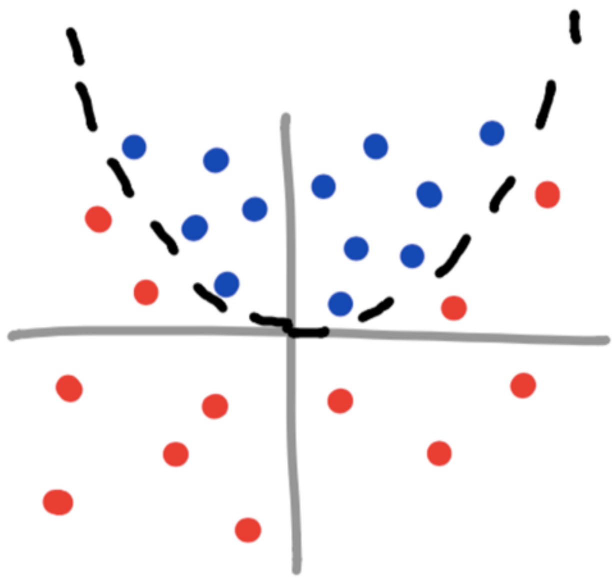
$$K(w, x) = < \phi(w), \phi(x) >$$

Линейное ядро



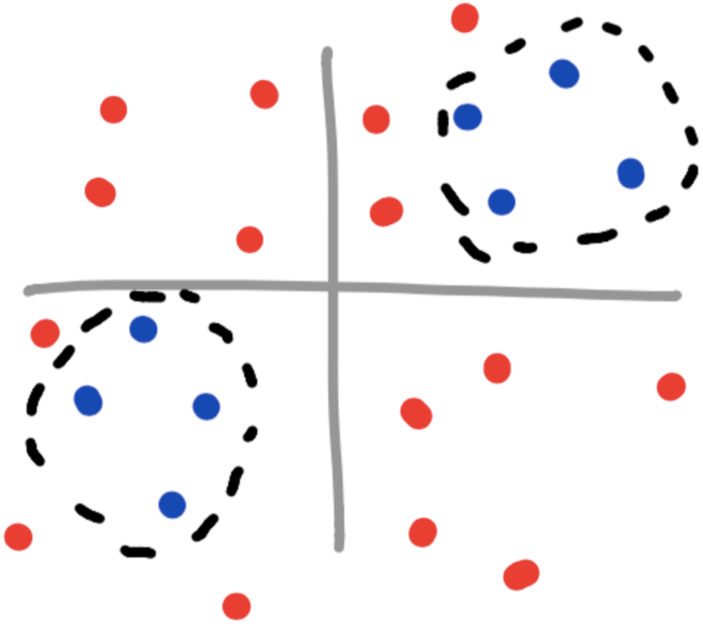
$$K(w, x) = < \mathrm{w}, x >$$

Полиномиальное ядро



$$K(w, x) = (\gamma < w, x > +r)^d$$

Радиальное ядро



$$K(w, x) = e^{-\gamma \|w - x\|^2}$$

Ядра и библиотеки

- LibSVM – можно выбирать ядра
- LibLinear – только линейное ядро
- Scikit-learn – обертка над LibSVM и LibLinear
- Vowpal Wabbit – только линейное ядро

Вид классификатора в SVM

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

Вид классификатора в SVM: теперь с ядром

$$a(x) = \operatorname{sign}\left(\sum_{i=1}^{h} \lambda_i y_i K(x_i, x) - w_0\right)$$

Мультиклассовая задача классификации

Мультиклассовая задача классификации

Как адаптировать SVM под задачу мультиклассовой классификации?

$$Y = \{1, \dots, K\}$$

Мультиклассовая задача классификации

Как адаптировать SVM под задачу мультиклассовой классификации?

$$Y = \{1, \ldots, K\}$$

- One-to-All
    - Строим K бинарных классификаторов

Мультиклассовая задача классификации

Как адаптировать SVM под задачу мультиклассовой классификации?

$$Y = \{1, \dots, K\}$$

- One-to-All
  - Строим K бинарных классификаторов
- One-to-One
  - Строим $K * (K-1)/2$ классификаторов

Мультиклассовая задача классификации

Как адаптировать SVM под задачу мультиклассовой классификации?

$$Y = \{1, \dots, K\}$$

- One-to-All
  - Строим K бинарных классификаторов
- One-to-One
  - Строим $K * (K-1)/2$ классификаторов
- Multiclass-SVM

Мультиклассовая задача: One-All Approach

In case of multilabel classification

$$y \in \{1, \dots, K\}$$

Lets train K classifiers

$$a_k(x) = \mathrm{sign}(\langle w_k, x \rangle + b_k),$$

k-th training sample is $(x_i, 2[y_i = k] - 1)_{i=1}^{l}$, $k = 1, 2, \dots, K$.

Final classifier:

$$a(x) = \underset{k \in \{1,\dots,K\}}{\arg\max} \Big\{ \langle w_k, x \rangle + b_k \Big\}.$$

Мультиклассовая задача: One-to-One

Lets train C_K^2 classifiers

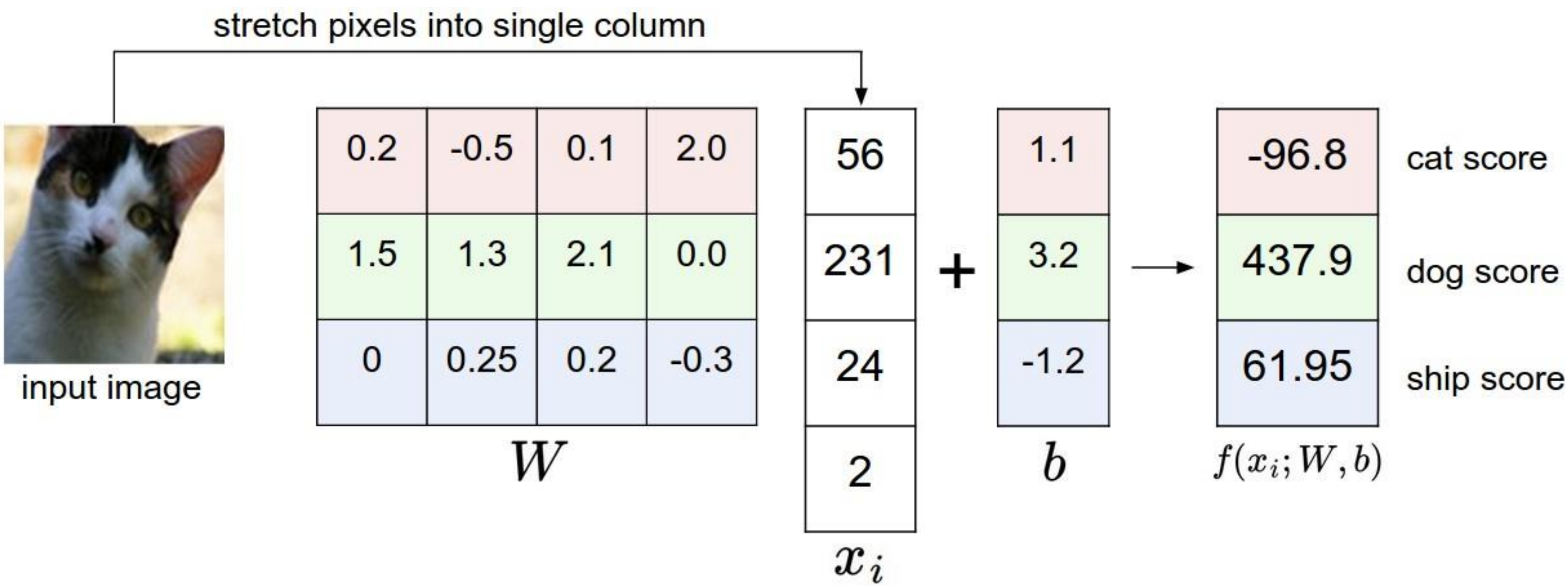
$$a_{ij}(x),\ k = 1, 2, \ldots, K,\ i \neq j$$

Training sample is set of objects with labels i and j.

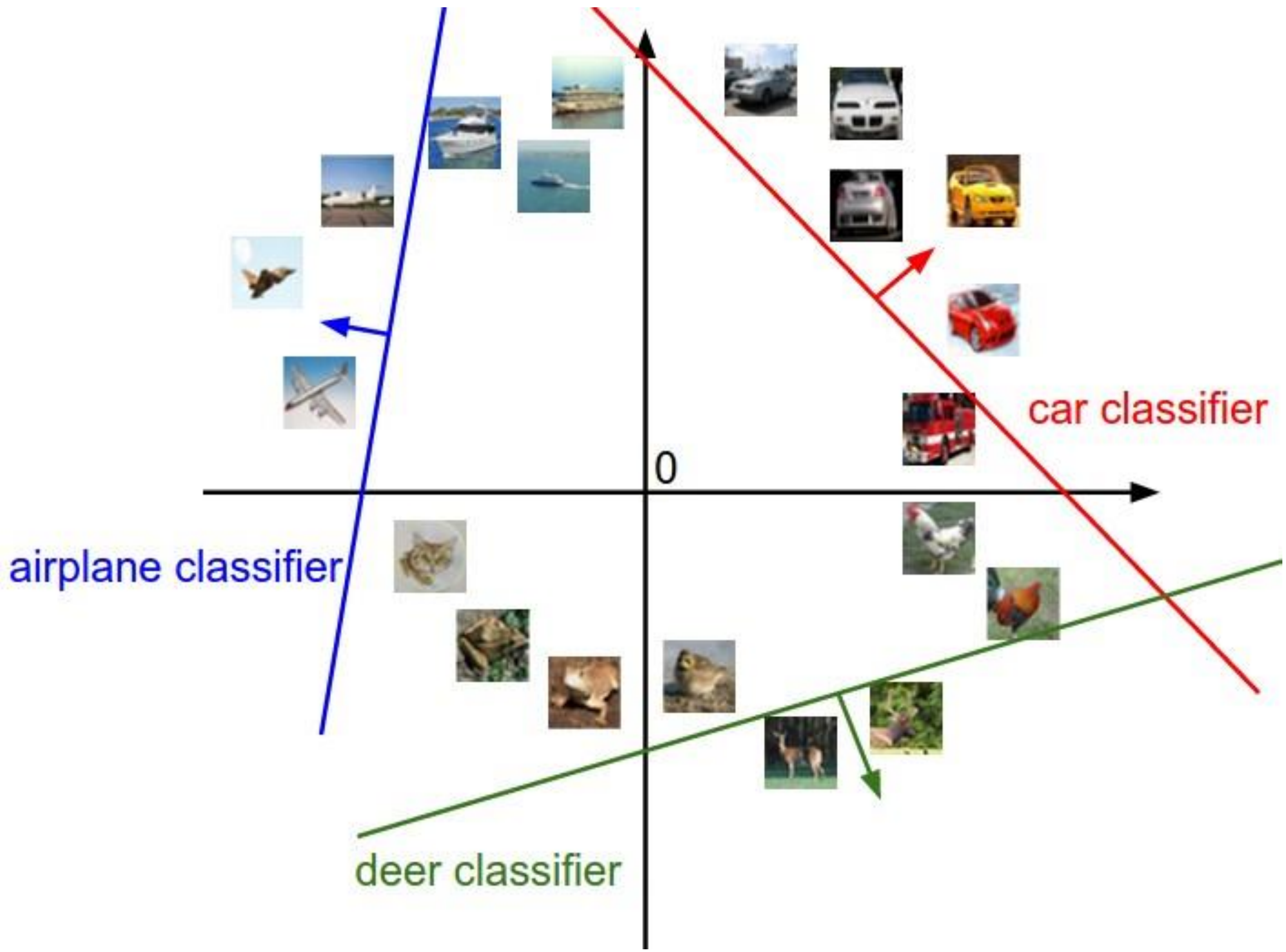
Final classifier:

$$a(x) = \underset{k \in \{1,\ldots,K\}}{\arg\max} \sum_{i=1}^{\ell} \sum_{j \neq i} [a_{ij}(x) = k].$$

Мультиклассовая задача: Multiclass-SVM



Мультиклассовая задача: Multiclass-SVM



Мультиклассовая задача: Multiclass-SVM

Lets fit K sets of parameters $w_1, \dots, w_K$.
Final classifier is defined as:

$$a(x) = \operatorname*{arg\,max}_{k \in \{1,\dots,K\}} \langle w_k, x \rangle.$$

Lets define loss function as:

$$\max_k \Big\{ \langle w_k, x \rangle + 1 - [k = y(x)] \Big\} - \langle w_{y(x)}, x \rangle.$$

Margin in multilabel case is Frobenius norm of $W \in \mathbb{R}^{K \times n}$:

$$\rho = \frac{1}{\|W\|^2} = \frac{1}{\sum_{k=1}^{K} \sum_{j=1}^{d} w_{kj}^2}.$$

Мультиклассовая задача: Multiclass-SVM

In linear separable case:

$$\begin{cases} \frac{1}{2}\|W\|^2 \to \min\limits_{W} \\ \langle w_{y_i}, x_i \rangle + [y_i = k] - \langle w_k, x_i \rangle \geqslant 1, \quad i = 1, \ldots, \ell; k = 1, \ldots, K. \end{cases}$$

In inseparable case:

$$\begin{cases} \frac{1}{2}\|W\|^2 + C \sum\limits_{i=1}^{\ell} \xi_i \to \min\limits_{W,\xi} \\ \langle w_{y_i}, x_i \rangle + [y_i = k] - \langle w_k, x_i \rangle \geqslant 1 - \xi_i, \quad i = 1, \ldots, \ell; k = 1, \ldots, K; \\ \xi_i \geqslant 0, \quad i = 1, \ldots, \ell. \end{cases}$$

SVM для регрессии

Линейная регрессия

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

Линейная регрессия

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

$$Q = \sum_{i=1}^{N} L\big(y_i, a(x_i)\big)$$

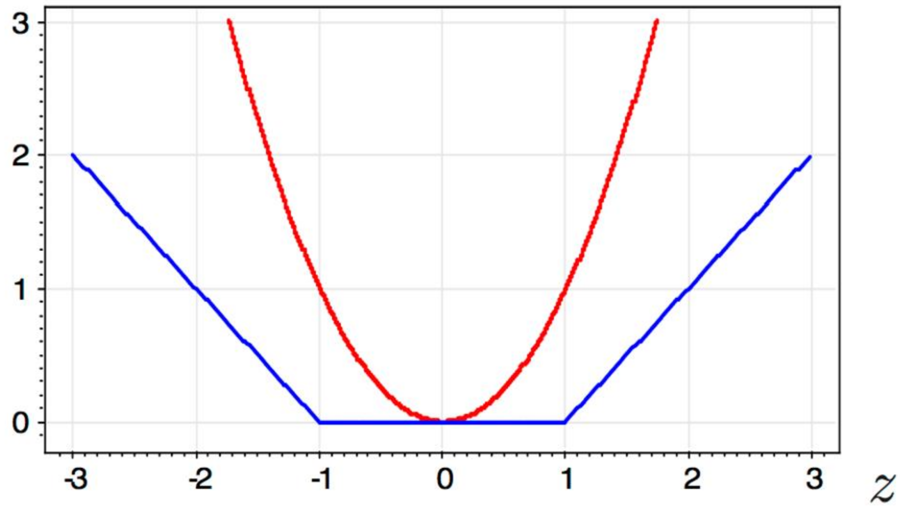
Линейная регрессия

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

$$Q = \sum_{i=1}^{N} L\big(y_i, a(x_i)\big)$$

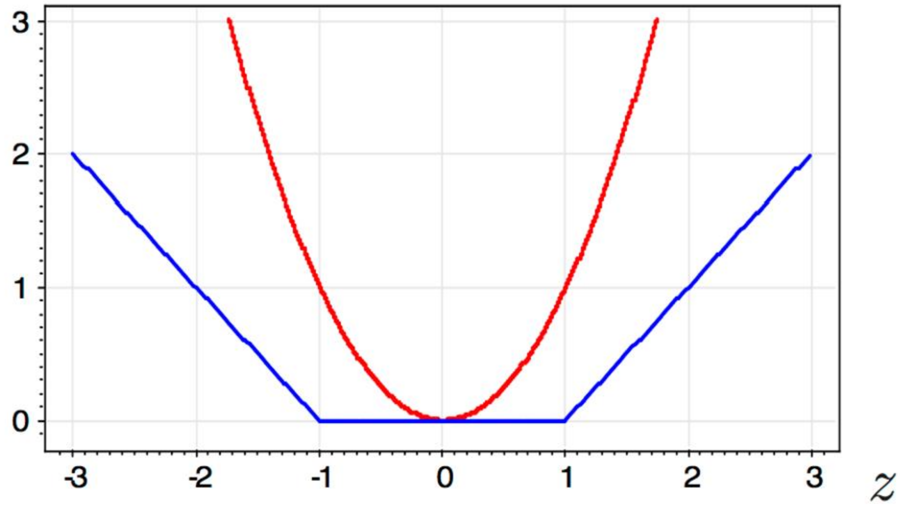
$$L\big(y_i, a(x_i)\big) = \left(y_i - a(x_i)\right)^2$$

Линейная регрессия



$$Q(a, X^\ell) = \sum_{i=1}^{\ell} \left(\langle w, x_i \rangle - w_0 - y_i \right)^2 + \tau \|w\|^2 \to \min_{w, w_0}$$

SVM для регрессии

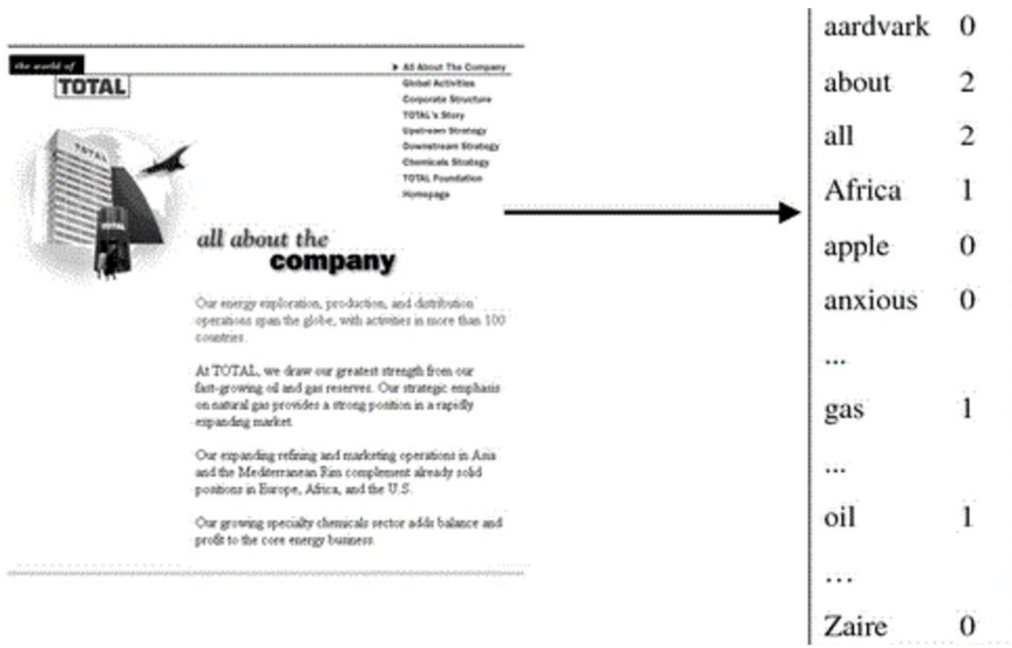


$$Q_\varepsilon(a, X^\ell) = \sum_{i=1}^{\ell} \left| \langle w, x_i \rangle - w_0 - y_i \right|_\varepsilon + \tau \langle w, w \rangle^2 \to \min_{w, w_0}$$

$$|M|_\varepsilon = (M - \varepsilon)_+ = \begin{cases} M, \text{если } M > \varepsilon, \\ 0, \text{если } M \leq \varepsilon \end{cases}$$

3. О применении линейных моделей

Классификация текстов



Обработка разреженных признаков

Часть признаков – разреженные, а часть – плотные, как быть?

- Строим на разреженных признаках линейную модель
- Ответ линейной модели добавляем как еще один плотный признак
- Либо строим на плотных признаках другую модель и усредняем с ней

Простые модели сложных систем

- Прогнозирование охвата целевой аудитории по параметрам рекламной кампании
- Прогнозирование дефектов в продукции на заводе
- Прогнозирование хим. состава в результате сложного процесса производства

Как правило – везде сложные зависимости от большого числа параметров, но не все параметры известны и мало точек

Резюме

1. Логистическая регрессия имеет яркую вероятностную интерпретацию
2. SVM – лучший метод линейной классификации
3. SVM обобщается для нелинейной классификации для линейной и нелинейной регрессии
4. Легко обобщается для задачи регрессии

5. Неясно, как подбираться ядро для реальных задача

Отзывы о лекции: https://goo.gl/forms/zeZiu1fSgrpPGp6T2