Метод с опорни вектори (Support Vector Machine – SVM). Невронни мрежи базирани на метода с опорните вектори.

- Дискриминационно обучение на класификатори
  - Намиране с обучение на разделяща хиперравнина (граница на решението); Лесно реализуемо програмно.
  - Проблем - липсва генерално решение, няма решение ако данните не са линейно разделими (перцептрон)
- Класификатори базирани на опорни вектори (SVM)
  - Задаване и получаване на граници и обобщение. Намиране на най-правилното, вместо на първото намерено решение.
  - Обучение на линейни SVM.

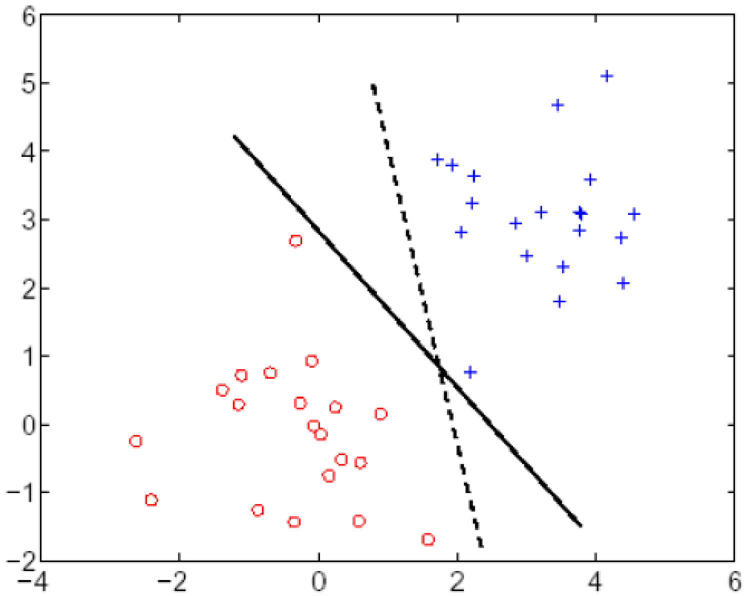
Линейна класификация

* Задача за бинарна класификация: нека зададем класова принадлежност на изходите y зададени в интервала {-1,1} към входни данни x

* Линеен класификатор: y=sign (w . x+ wo) като неговата 'решаваща повърхнина', т.е. разделящата хиперравнина е дефинирана с w.x+wo=0

* Линейна разделимост: Можем да намерим произволен линеен класификатор, при който да е удовлетворено условието всичките образци от обучаващата издава да бъдат правилно разпознати.

$$y_i[\mathbf{w} \cdot \mathbf{x}_i + w_0] > 0, \qquad \forall i = 1, \ldots, n$$



Перцептрон

- При Перцептрона целта е да се намери линия разделяща входните данни на +1 и -1 класове, които да съвпадат с желания изход y

$$o(\mathbf{x}) = \mathrm{sign}(\mathbf{w} \cdot \mathbf{x} + w_0) =? y(\mathbf{x})$$

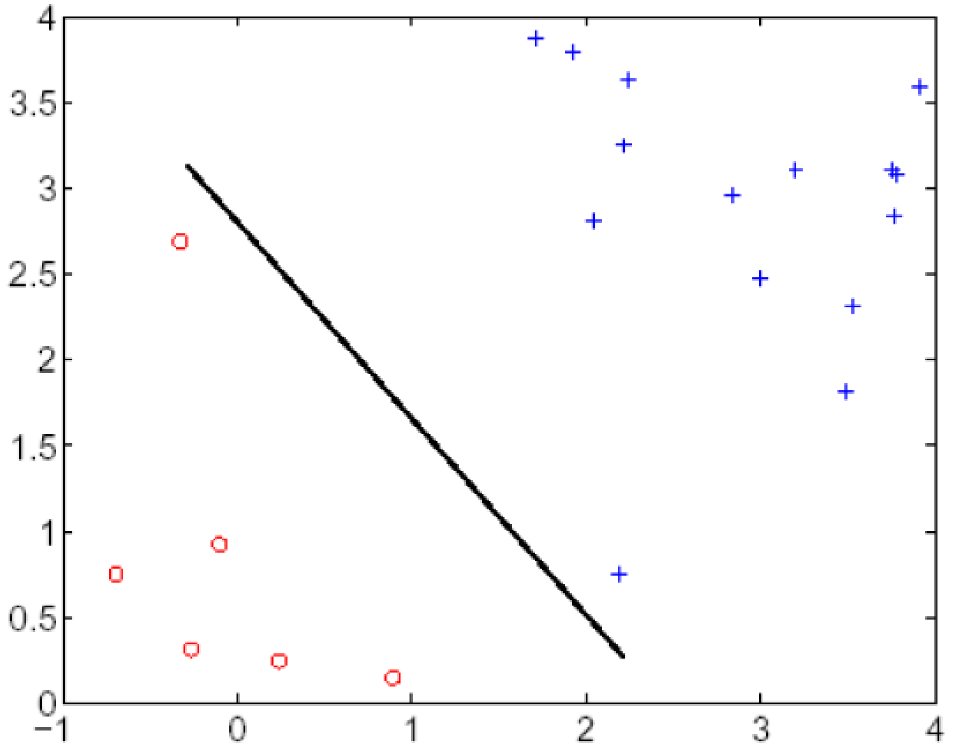
тогава за всеки елемент i трябва да е изпълнено

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + w_0) > 0$$

Можем да настроим теглата {w,wo} с правилото за обучение на перцептрона, което гарантира сходимост при коректно решение в случай, че имаме линейно разделими множества.

Проблем - Кое решение би ни дало най-добрата генерализация?

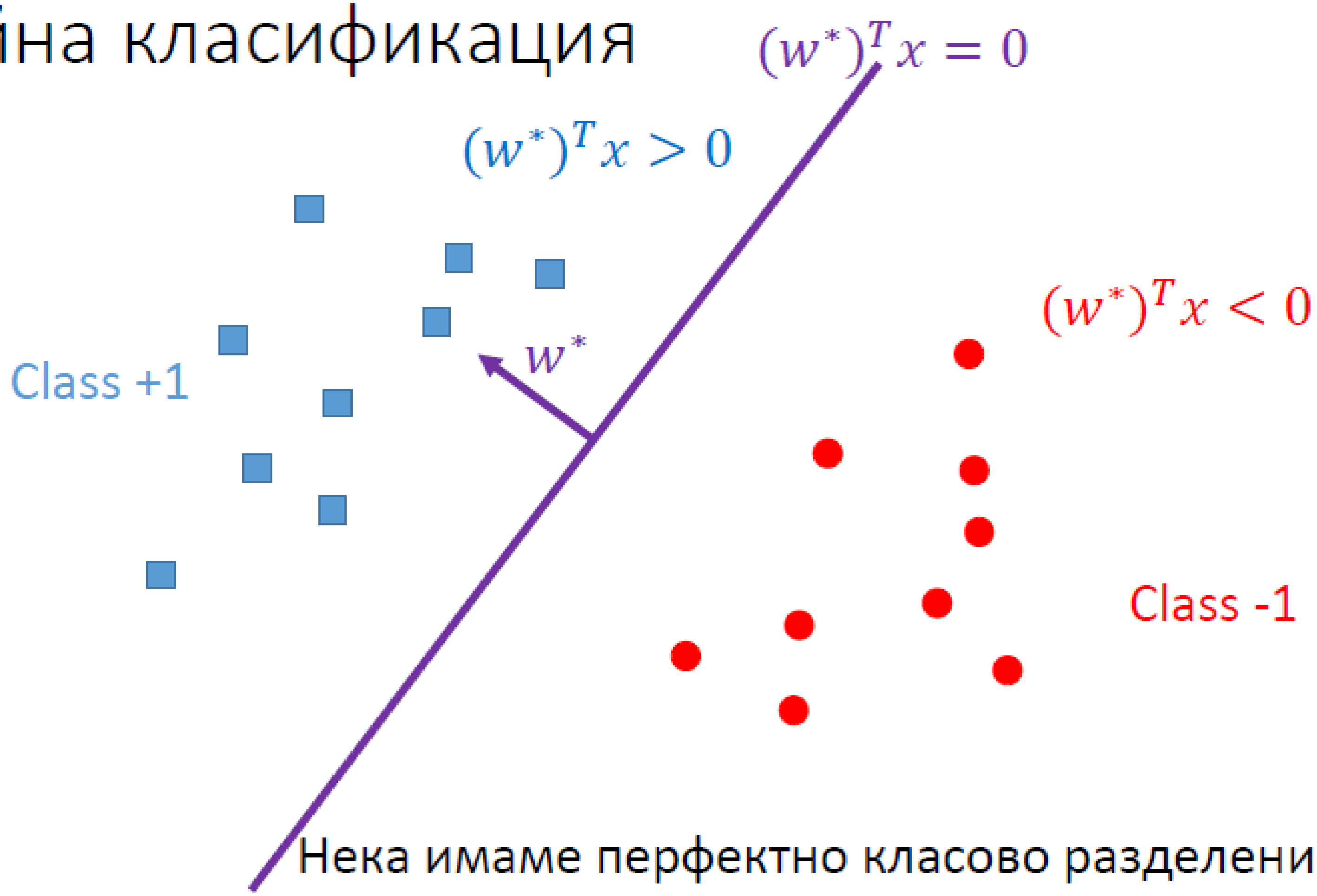
Геометрична интерпретация на линеен класификатор



- Марж: минимална разлика между класовете и границите на решението.

- Отговор: Линейната повърхност на решение с максимален марж.

Линейна класификация

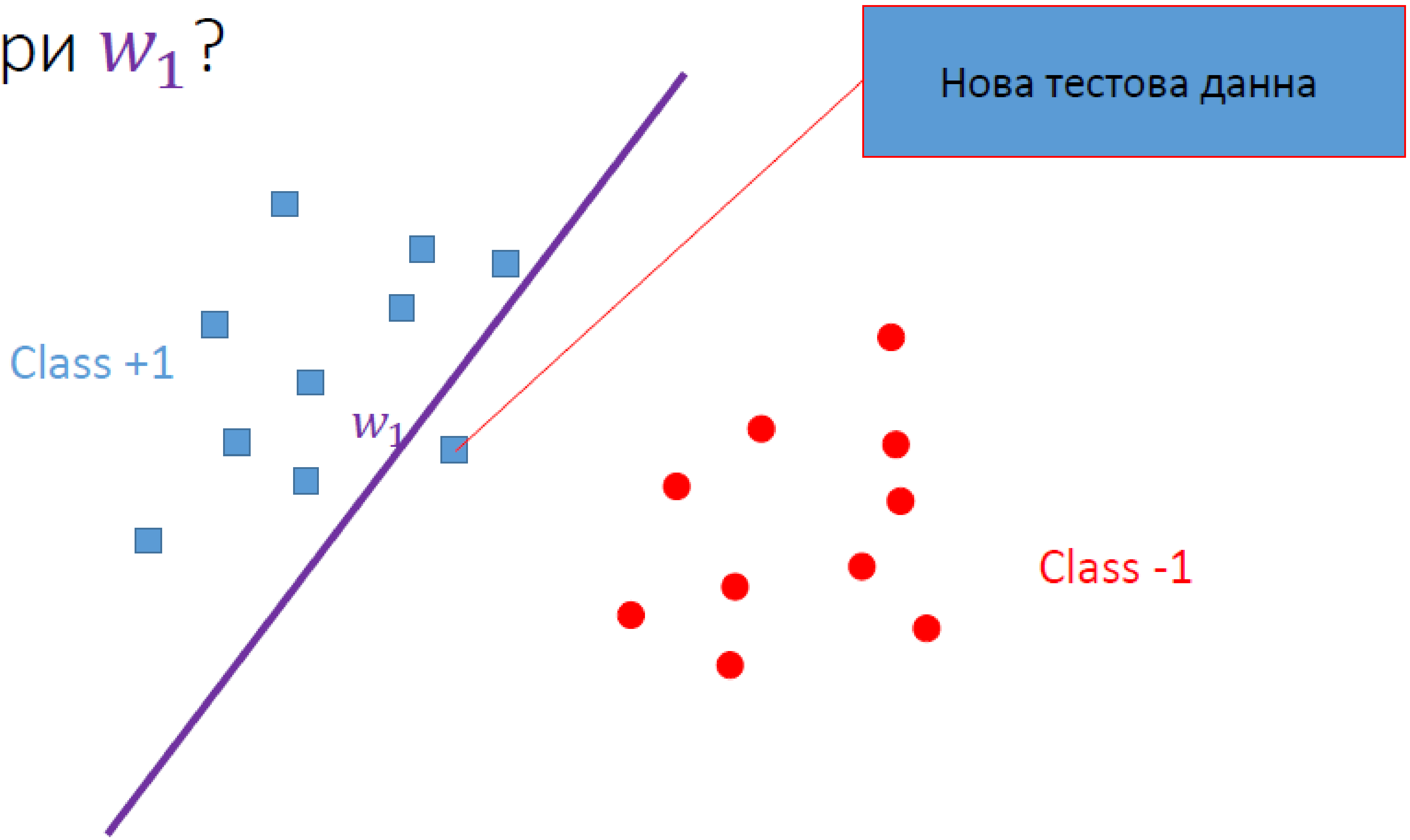


Нека имаме перфектно класово разделение

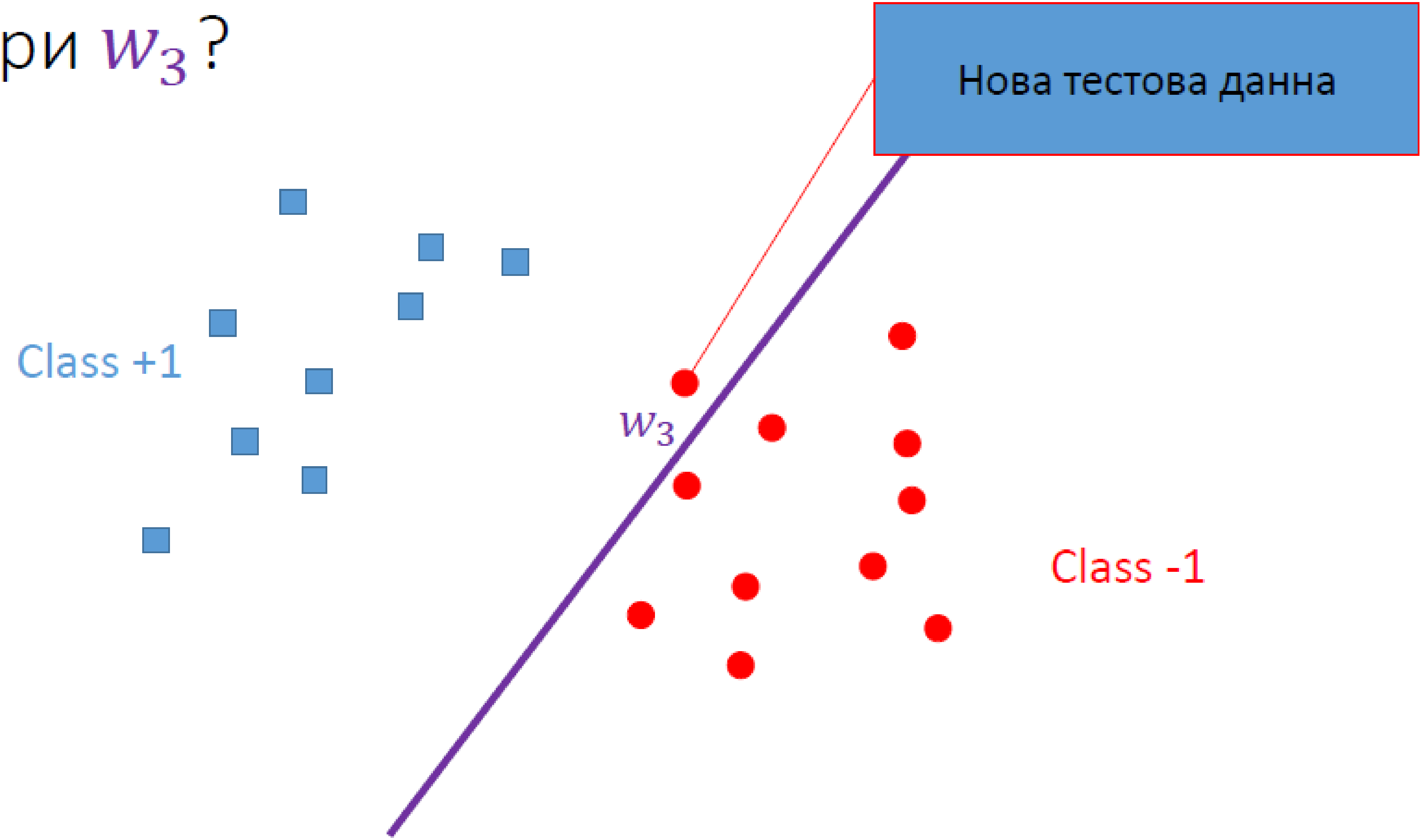
Множество решения?

Всички решения са с еднакви емпирични грешки, но са различаващи се при тестове с данни

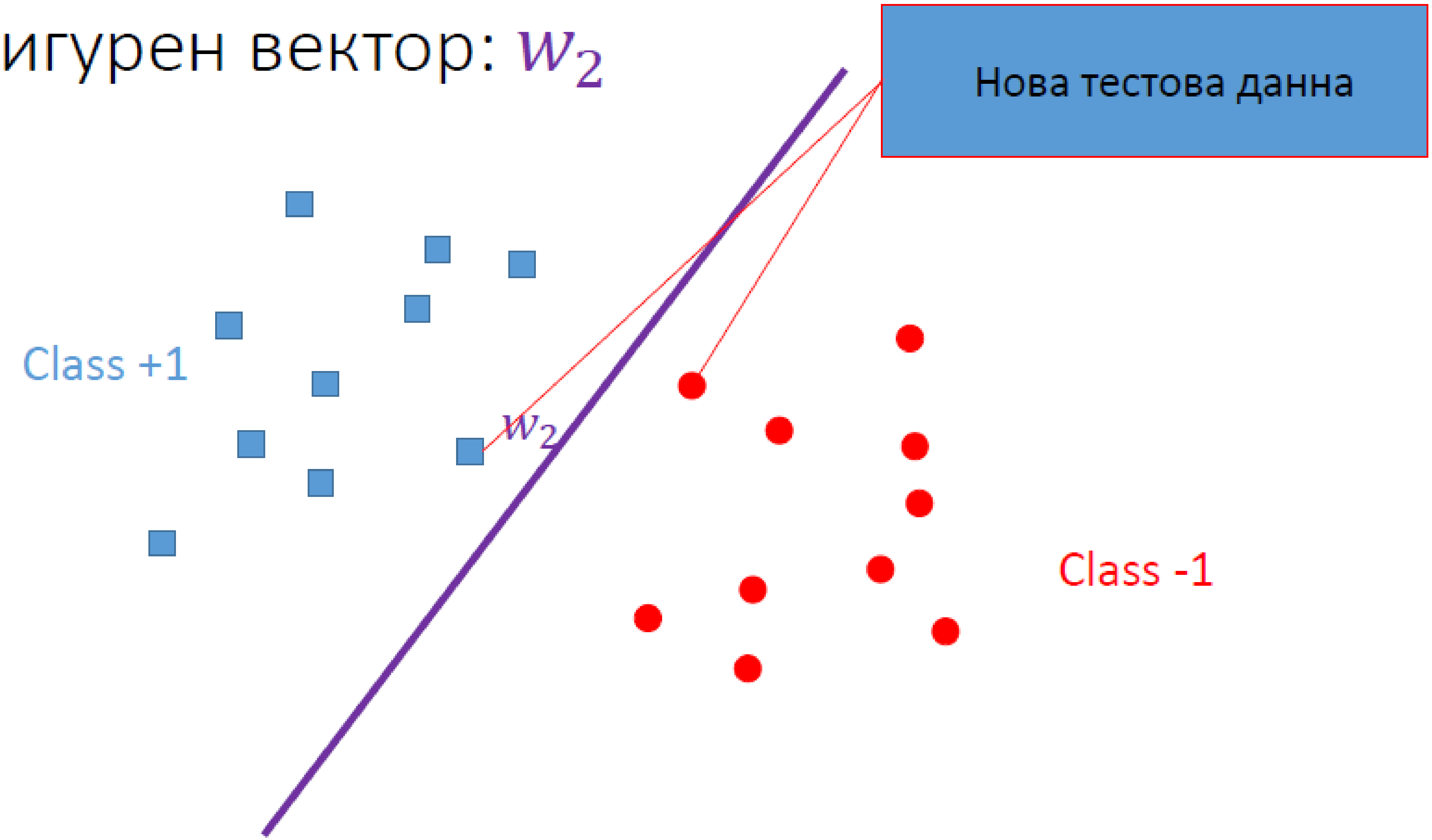
Ами при w_1?



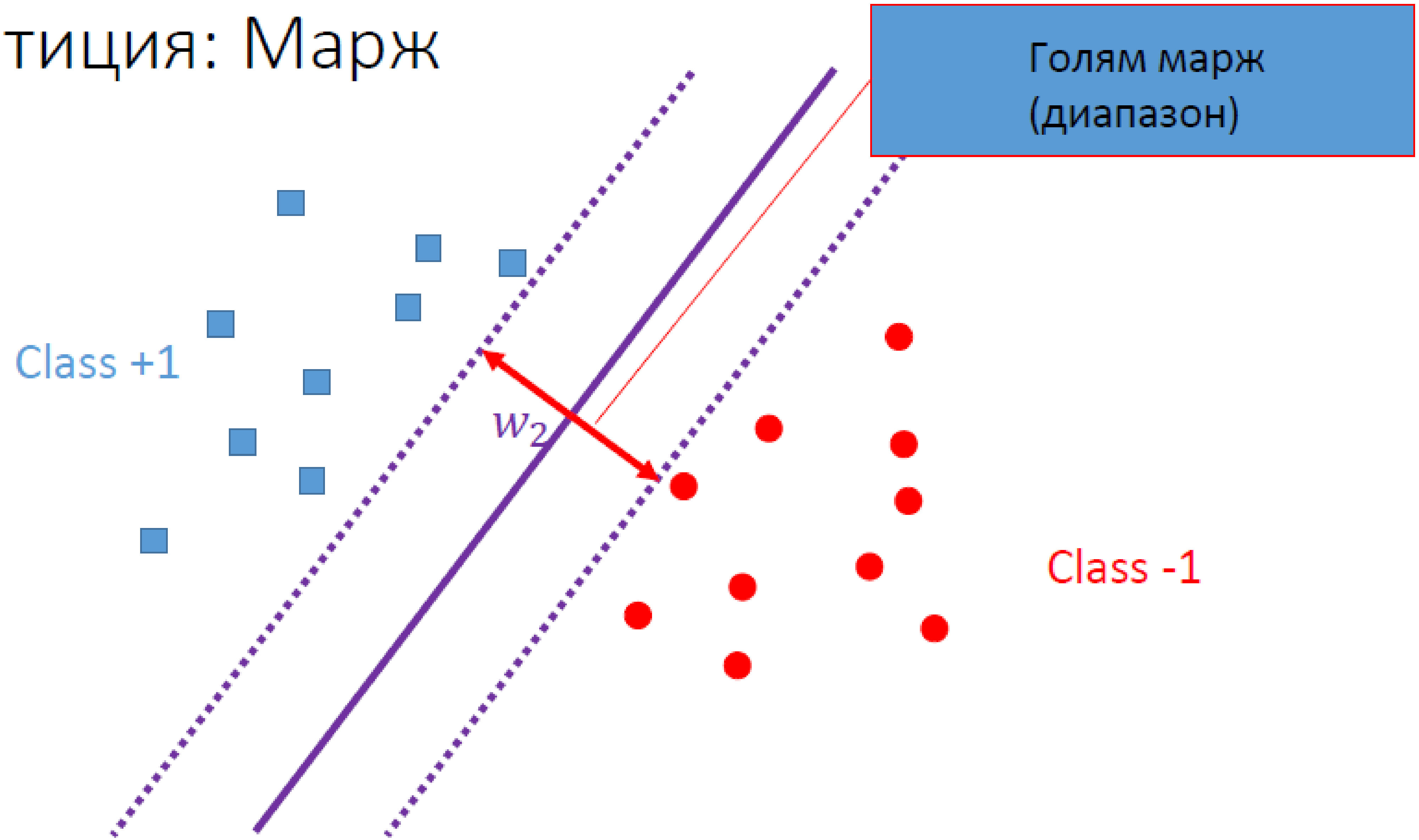
Или при w_3?

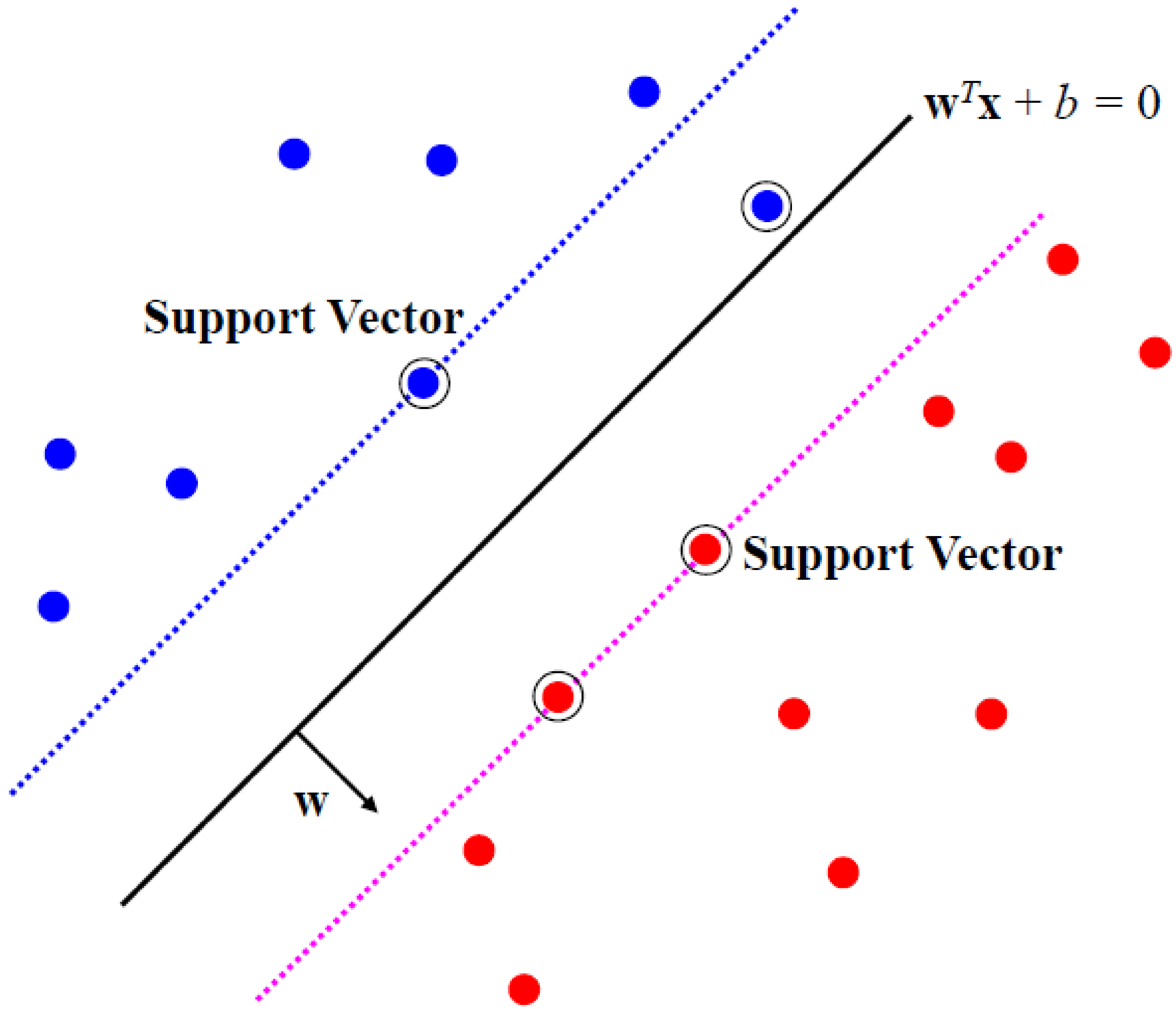


Най-сигурен вектор: w_2



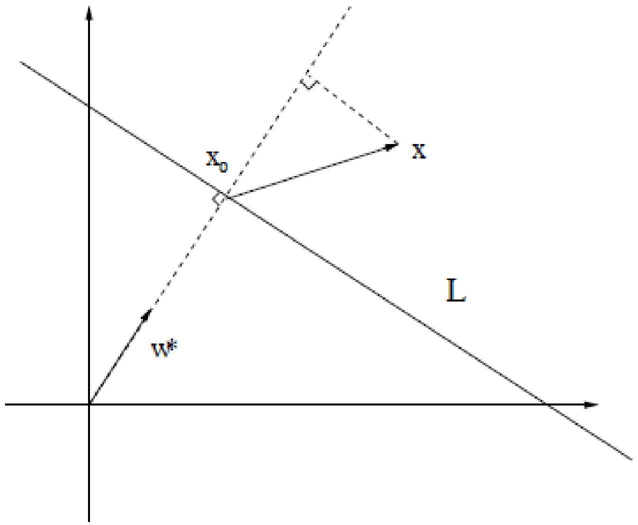
Интуитиция: Марж




$\mathbf{w}^T\mathbf{x} + b = 0$
Support Vector
Support Vector
w

Геометричен марж

- Представяне с векторна алгебра



- За всеки две точки x1 и x2 лежащи на L ние имаме w.(x1-x2)=0, което води до заключението, че w*=w/||w|| е нормалата към повърхнината L.
- За всяка точка xo в L, w.xo=-wo
- Разстоянието по знак между x и L се задава с $\mathbf{w}^* \cdot (\mathbf{x} - \mathbf{x}_0) = \frac{1}{||\mathbf{w}||}(\mathbf{w} \cdot \mathbf{x} + w_0)$
- Геометричния марж на $(\mathbf{x}_i, y_i)$ по отношение на L: $\gamma_i = y_i \frac{1}{||\mathbf{w}||}(\mathbf{w} \cdot \mathbf{x}_i + w_0)$.
- Геометричния марж на $\{(\mathbf{x}_i, y_i)_{i=1}^n\}$ по отношение на L: $\min_i \gamma_i$.

Линеен SVM Класификатор

- Линейният SVM увеличава геометричния марж на набора от данни за обучение:

$$\max_{\mathbf{w},w_0} \quad C \qquad (1)$$
$$s.t. \quad y_i \frac{1}{||\mathbf{w}||}(\mathbf{w}\cdot\mathbf{x}_i + w_0) \geq C, \quad i = 1, ..., n$$

За всяко решение, отговарящо на ограниченията, всяко положително мащабирано множествено число също ги удовлетворява. Така че с произволна настройка $||\mathbf{w}|| = 1/C$, е възможно да се формулира SVM като: $(\min ||x|| \Leftrightarrow \min 1/2||x||^2)$

$$\min_{\mathbf{w},w_0} \quad \frac{1}{2}||\mathbf{w}||^2 \qquad (2)$$
$$y_i(\mathbf{w}\cdot\mathbf{x}_i + w_0) \geq 1, \quad i = 1, ..., n$$

- С тази настройка дефинираме граница около линейната граница на решение с дебелина $1/||\mathbf{w}||$.

Решение за линейни SVM

- Можем да превърнем ограничената минимизация в неограничен проблем за оптимизация, като представим ограниченията като забранителни/наказателни членове (penality term)

$$\min_{\mathbf{w},w_0} \quad \frac{1}{2}||\mathbf{w}||^2 + \texttt{penality term}$$

За данни $(\mathbf{x}_i, y_i)$, може да се използва следния наказателен член.

$$\left\{ \begin{array}{l} 0, y_i(\mathbf{w}\cdot\mathbf{x}_i + w_0) \geq 1 \\ \infty, \texttt{otherwise} \end{array} \right\} = \max_{\alpha_i \geq 0} \alpha_i(1 - y_i[w_0 + \mathbf{w}\cdot\mathbf{x}_i])$$

Модифициране и пренаписване на процедурата за минимизация в:

$$\min_{\mathbf{w},w_0} \{\frac{1}{2}||\mathbf{w}||^2 + \sum_{i=1}^{n} \max_{\alpha_i \geq 0} \alpha_i(1 - y_i[w_0 + \mathbf{w}\cdot\mathbf{x}_i])\} \qquad (3)$$

$$= \min_{\mathbf{w},w_0} \max_{\{\alpha_i \geq 0\}} \{\frac{1}{2}||\mathbf{w}||^2 + \sum_{i=1}^{n} \alpha_i(1 - y_i[w_0 + \mathbf{w}\cdot\mathbf{x}_i])\}$$

- $\{\alpha_i\}$ се наричат мултипликатори на Лагранж

Решение за линейни SVM

Можем да разменим 'max' и 'min':

$$\min_{\mathbf{w},w_0} \max_{\{\alpha_i \geq 0\}} \{\frac{1}{2}||\mathbf{w}||^2 + \sum_{i=1}^{n} \alpha_i(1 - y_i[w_0 + \mathbf{w}\cdot\mathbf{x}_i])\} \tag{4}$$

$$= \max_{\{\alpha_i \geq 0\}} \min_{\mathbf{w},w_0} \underbrace{\{\frac{1}{2}||\mathbf{w}||^2 + \sum_{i=1}^{n} \alpha_i(1 - y_i[w_0 + \mathbf{w}\cdot\mathbf{x}_i])\}}_{J(\mathbf{w},w_0;\alpha)}$$

Първом минимизираме $J(\mathbf{w}, w_0; \alpha)$ по отношение на $\{\mathbf{w}, w_0\}$ за всички постоянни настройки на мултипликаторите на Лагранж:

$$\frac{\partial}{\partial \mathbf{w}} J(\mathbf{w}, w_0; \alpha) = \mathbf{w} - \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i = 0 \tag{5}$$

$$\frac{\partial}{\partial w_0} J(\mathbf{w}, w_0; \alpha) = -\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{6}$$

Решение за линейни SVM

Замествайки (5) и (6) обратно в $J(\mathbf{w}, w_0; \alpha)$:

$$\max_{\{\alpha_i \geq 0\}} \min_{\mathbf{w}, w_0} \underbrace{\{\frac{1}{2}||\mathbf{w}||^2 + \sum_{i=1}^{n} \alpha_i (1 - y_i[w_0 + \mathbf{w} \cdot \mathbf{x}_i])\}}_{J(\mathbf{w}, w_0; \alpha)} \qquad (7)$$

$$= \max_{\substack{\alpha_i \geq 0 \\ \sum_i \alpha_i y_i = 0}} \{\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j)\}$$

В крайна сметка, трансформираме оригиналното линейно SVM обучение в задача за квадратично програмиране (7), който има уникално и оптимално решение.

Можем да намерим оптималните настройки за мултипликаторите на Лагранж $\{\hat{\alpha}_i\}$ и след това да намерим оптималните тегла $\{\hat{\mathbf{w}}, \hat{w}_0\}$.

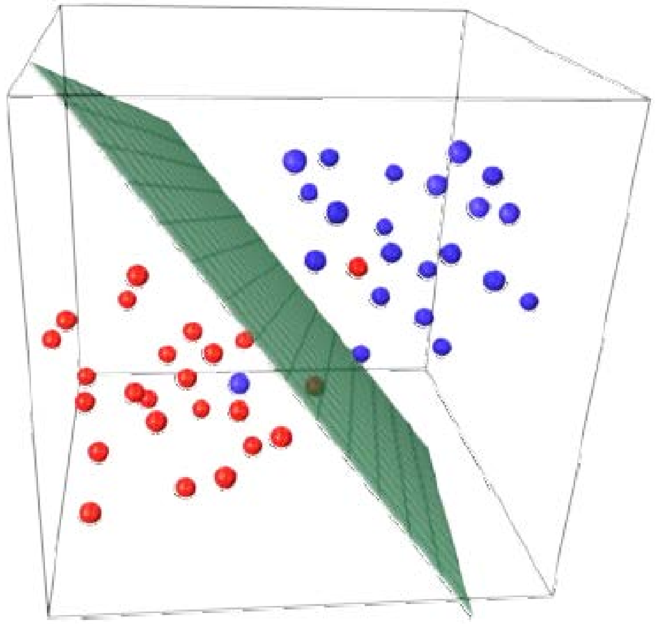
Обобщение за линейните SVM

Бинарна и линейна разделяща класификация

Линеен класификатор с максимален марж между данните от отделните класове.

Обучение на SVM с минимизация на:

$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$



По отношение на $\alpha_i \geq 0$ и $\sum_i \alpha_i y_i = 0$.

$$\hat{\mathbf{w}} = \sum_{i=1}^{n} \hat{\alpha}_i y_i \mathbf{x}_i.$$

Само една малка част от $\widehat{\alpha}_i$ ще бъдат различни от нула и съответстващите им $\mathbf{x}_i$ се наричат опорни вектори.

Прогнозата на нов образ x е знака на:

$$\widehat{w}_0 + \mathbf{x} \cdot \hat{\mathbf{w}} = \widehat{w}_0 + \mathbf{x} \cdot (\sum_{i=1}^{n} \widehat{\alpha}_i y_i \mathbf{x}_i) = \widehat{w}_0 + \sum_{i \in SV} \widehat{\alpha}_i y_i (\mathbf{x} \cdot \mathbf{x}_i)$$