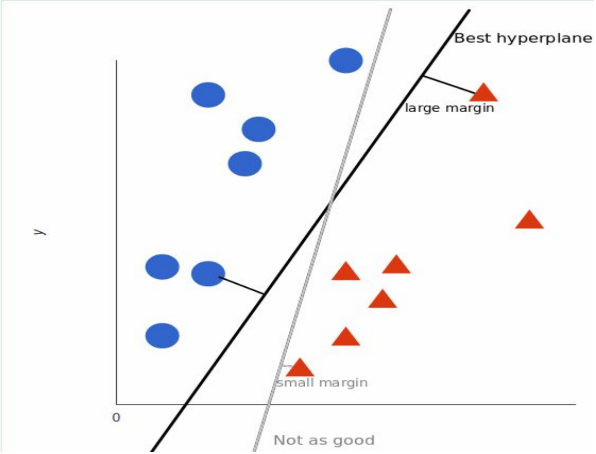
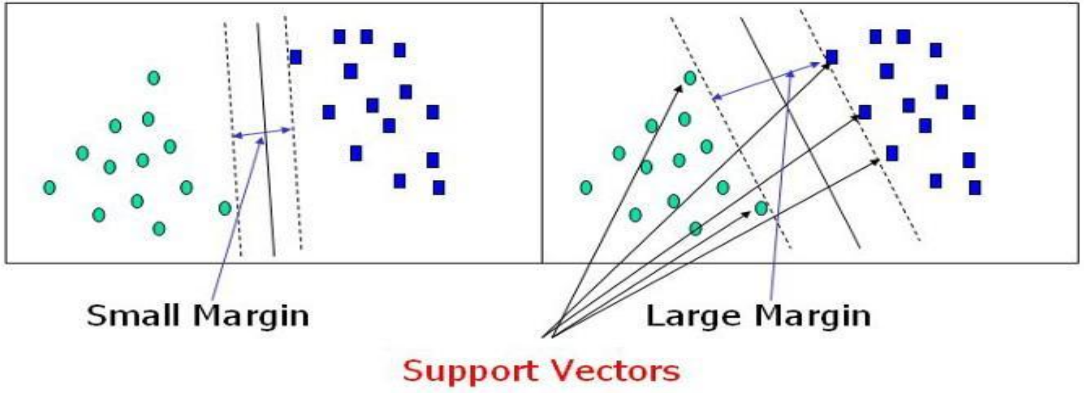
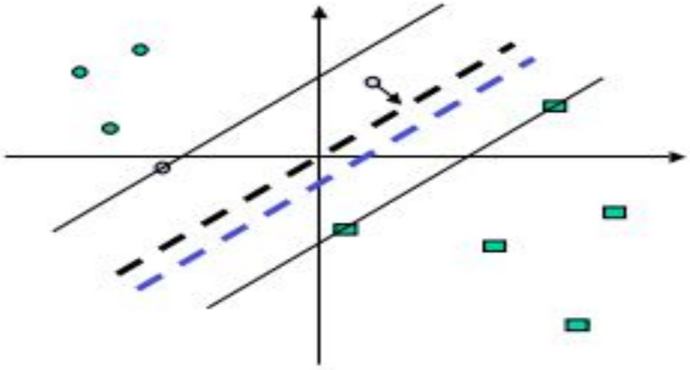
Ідея SVM

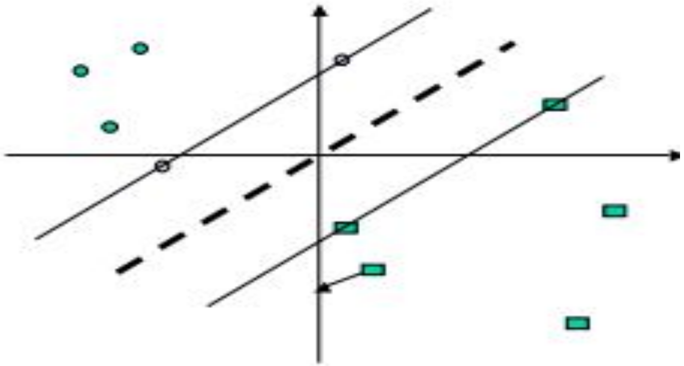

Best hyperplane
large margin
y
small margin
0
Not as good


Small Margin
Large Margin
Support Vectors



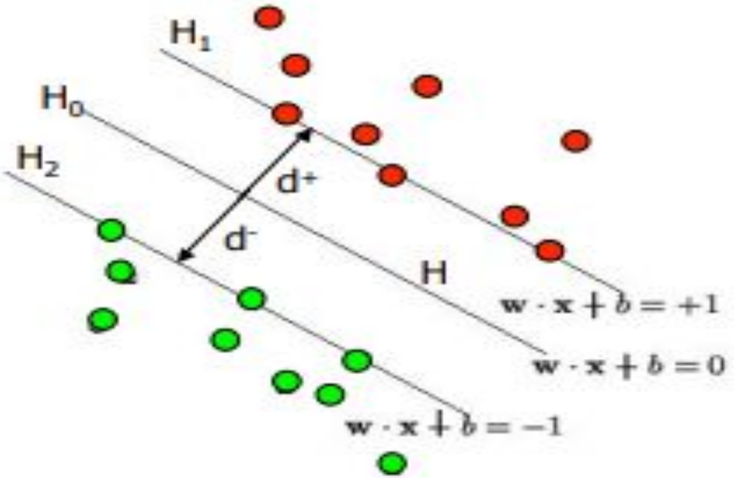
Зміщуємо опорні вектори - змущується вирішальна площина

Зміщюємо інші вектори - нічого не змінюється



$w \cdot x_i + b \geq +1$ when $y_i = +1$
$w \cdot x_i + b \leq -1$ when $y_i = -1$

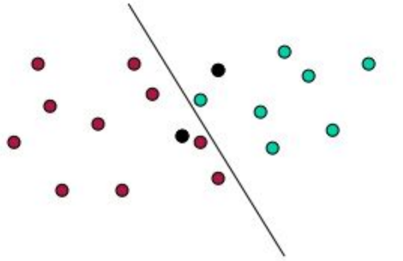
$\mathbf{w}^T\mathbf{x} + b \geq 0$ for $d_i = +1$
$\mathbf{w}^T\mathbf{x} + b < 0$ for $d_i = -1$

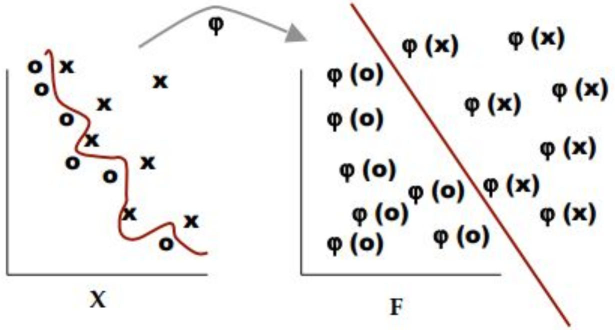


$$\frac{2}{\|w_0\|}$$

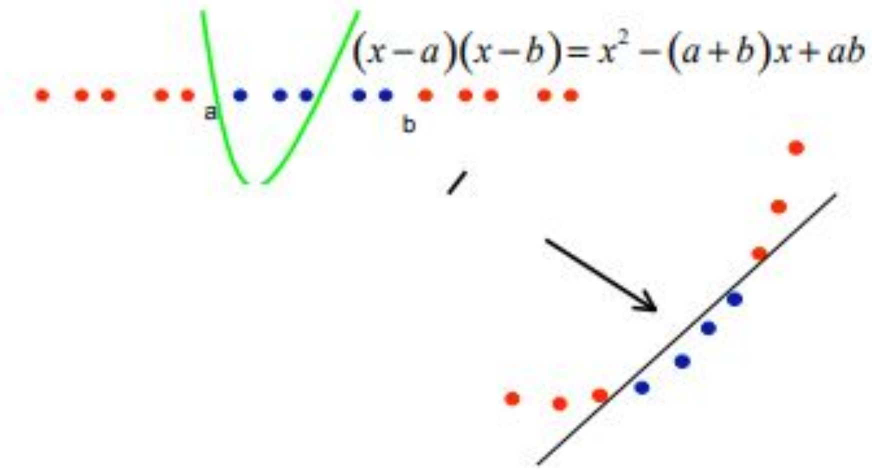
SVM для задачі розпізнавання образів

1. Нелінійне відображення вхідного вектора в простір ознак (feature space) більш високої розмірності.
2. Побудова оптимальної гіперплощини для поділу ознак





Операція 2. В цій операції використовується ідея побудови оптимальної розділяючої гіперплощини що визначається як лінійна функція від векторів, взятих з простору ознак, а не з початкового вхідного простору.



Ядро скалярного добутка

Нехай x - деякий вектор з вхідного простору, розмірність якого рівна m0. Нехай $\{\varphi_j(\mathbf{x})\}^{m1}{}_{j=1}$ - множина нелінійних перетворень з вхідного простору в простір ознак. Розмірність простору ознак позначимо через m1. Передбачається, що функції φj (x) визначені для всіх j. Маючи множина нелінійних перетворень, можна визначити гіперплощину, що визначає поверхню рішень

$$\sum_{j=1}^{m_1} w_j \varphi_j(x) + b = 0,$$

де $\left\{w_j\right\}^{m1}{}_{j=1} = i$— множина вагових коефіцієнтів, що пов'язують простір ознак з вихідним простором; b - поріг. Цей вираз можна спростити:

$$\sum_{j=0}^{m_1} w_j \varphi_j(x) = 0,$$

вважаючи, що φ0 (x) = 1 для всіх x. При цьому w0 є порогом b. Останнє рівняння визначає поверхню рішень, обчислену в просторі ознак в термінах лінійних вагових коефіцієнтів машини.

Величина φj (x) являє собою вхідний сигнал, який припадає на вагу wj в просторі ознак. Визначимо вектор

$$\varphi(x) = [\varphi_0(x), \varphi_1(x), \ldots, \varphi_{m1}(x)]^T,$$

де за визначенням $\varphi_0(x) = 1$ для всіх x.

В результаті вектор ознак φ (x) являє собою "образ", індукований в просторі ознак при пред'явленні вектора вхідного сигналу x. Таким чином, в термінах цього способу можна визначити поверхню рішень в компактній формі:

$$w^T \varphi(x) = 0.$$

Адаптуючи вираз до задачі лінійного поділу векторів в просторі ознак, можна записати:

$$w = \sum_{i=1}^{N} \alpha_i d_i \varphi(x_i),$$

де *вектор ознак* **φ**(x*i*) відповідає вхідному вектору **x***i* в *i*-му прикладі.

Агрегуючи ці два вирази можна визначити площину рішень у просторі ознак наступним чином :

$$\sum_{i=1}^{N}\alpha_i d_i \varphi^T(x_i)\varphi(x) = 0.$$

Терм **φ**т(х***і*)φ**(х**)** являє собою скалярний добуток двох векторів, індукованих в просторі ознак для вхідного вектора х і прикладу хі. Виходячи з цього, можна ввести поняття ядра скалярного добутку, що позначається як К {х, хі):

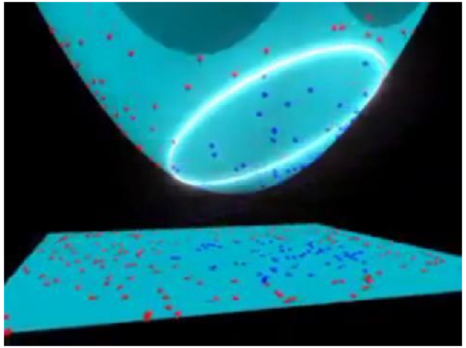
$$K(x,x_i) = \varphi^T(x)\varphi(x_i) = \sum_{j=0}^{m_1}\varphi_j(x)\varphi_j(x_i), \quad i=1,2,\ldots,N.$$

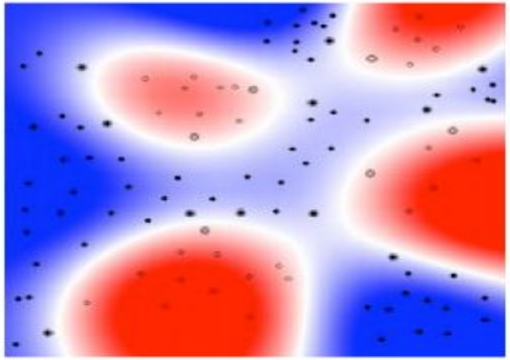
З цього визначення видно, що ядро скалярного добутку є симетричною функцією своїх аргументів, тобто для всіх і виконується співвідношення

$$K(x,x_i) = R(x_i,x).$$

Ядро скалярного добутку К (х, хі) можна використовувати для побудови оптимальної гіперплощини в просторі ознак, не задаючи його в явному вигляді. Оптимальна гіперплощина визначається як: $\sum_{i=1}^{N}\alpha_i d_i K(x,x_i) = 0.$

Нелінійна SVM





$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (1 + \boldsymbol{x}_i^T \boldsymbol{x}_j)^p; \quad K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(\gamma ||\boldsymbol{x}_i - \boldsymbol{x}_j||^2);$$