Lecture 26

# Kernels and Support Vector Machine with Quantum Circuits

*of the course "Fundamentals of Quantum Computing"*
*(by IAPS and QUANTERALL)*

Stoyan Mishev, Vesselin Gueorguiev and Vladimir Gerdjikov

INSTITUTE *for* ADVANCED PHYSICAL STUDIES

NEW BULGARIAN UNIVERSITY

February 03, 2023

# Plan



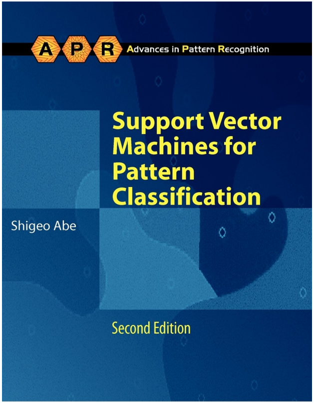
Theorem Hilbert

Support vector machine

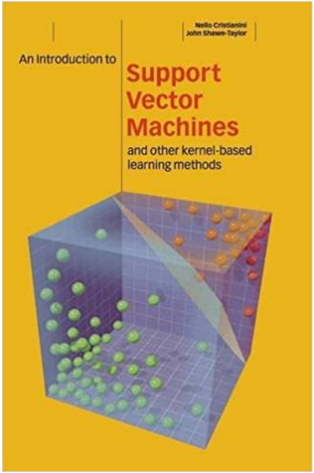
Quantum Circuits and Kernels

# Books






MyCopy powered by SpringerLink



# Theorem Hilbert

- C. Berg, J. P. R. Christensen, and P. Ressel. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions. Springer-Verlag, New York, 1984.
- Shigeo Abe, Support Vector Machines for Pattern Classification

If a symmetric function $H(x, x')$

[1] Shigeo Abe, Support Vector Machines for Pattern Classification

# Hilbert-Schmidt theory.[1]



If a symmetric function $H(x, x')$ satisfies

$$\sum_{i,j}^{M} h_i h_j H(x_i, x_j) \geq 0 \tag{1}$$

for each $M \geq 0$, $x_i$ and $h_i \in \mathbb{R}$ (,i.e. $H$ is nonnegative definite),

[1] Shigeo Abe, Support Vector Machines for Pattern Classification

If a symmetric function $H(x, x')$ satisfies

$$\sum_{i,j}^{M} h_i h_j H(x_i, x_j) \geq 0 \quad (1)$$

for each $M \geq 0$, $x_i$ and $h_i \in \mathbb{R}$ (,i.e. $H$ is nonnegative definite), then there exists a mapping $\mathbf{g(x)}$, such that $H$ can be factorized

$$H(x, x') = g^T(x)g(x').$$

[1]Shigeo Abe, Support Vector Machines for Pattern Classification

# Hilbert-Schmidt theory.[1]

If a symmetric function $H(x, x')$ satisfies

$$\sum_{i,j}^{M} h_i h_j H(x_i, x_j) \geq 0 \qquad (1)$$

for each $M \geq 0$, $x_i$ and $h_i \in \mathbb{R}$ (,i.e. $H$ is nonnegative definite), then there exists a mapping $\mathbf{g(x)}$, such that $H$ can be factorized

$$H(x, x') = g^T(x) g(x').$$

Therefore

$$\sum_{i,j} h_i h_j H(x_i, x_j) = (\sum_{i=1}^{M} h_i g^T(x_i))(\sum_{i=1}^{M} h_i g(x_i)) \geq 0$$

Eq. (1) - Mercer's condition, $H(x, x')$ - positive semidefinite (Mercer) kernel

[1]Shigeo Abe, Support Vector Machines for Pattern Classification

# D

# Positive Semidefinite Kernels and Reproducing Kernel Hilbert Space

Support vector machines are based on the theory of reproducing kernel Hilbert space. Here, we summarize some of the properties of positive semidefinite kernels and reproducing kernel Hilbert space based on [30].

## D.1 Positive Semidefinite Kernels

**Definition D.1.** Let $H(\mathbf{x}, \mathbf{x}')$ be a real-valued symmetric function with $\mathbf{x}$ and $\mathbf{x}'$ being $m$-dimensional vectors. For any set of data $\{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ and $\mathbf{h}_M = (h_1, \ldots, h_M)^T$ with $M$ being any natural number, if

$$\mathbf{h}_M^T H_M \mathbf{h}_M \geq 0 \tag{D.1}$$

is satisfied (i.e., $H_M$ is a positive semidefinite matrix), we call $H(\mathbf{x}, \mathbf{x}')$ a *positive semidefinite kernel*, where

$$H_M = \begin{pmatrix} H(\mathbf{x}_1, \mathbf{x}_1) & \cdots & H(\mathbf{x}_1, \mathbf{x}_M) \\ \vdots & \ddots & \vdots \\ H(\mathbf{x}_M, \mathbf{x}_1) & \cdots & H(\mathbf{x}_M, \mathbf{x}_M) \end{pmatrix}. \tag{D.2}$$

If (D.1) is satisfied under the constraint

$$\sum_{i=1}^{M} h_i = 0, \tag{D.3}$$

$H(\mathbf{x}, \mathbf{x}')$ is called a *conditionally positive semidefinite kernel*.

From the definition it is obvious that if $H(\mathbf{x}, \mathbf{x}')$ is positive semidefinite, it is also conditionally positive semidefinite. In the following we discuss several properties of (conditionally) positive semidefinite kernels that are useful for constructing positive semidefinite kernels.

**Theorem D.2.** *If*

$$H(\mathbf{x}, \mathbf{x}') = a, \tag{D.4}$$

*where* $a > 0$, $H(\mathbf{x}, \mathbf{x}')$ *is positive semidefinite.*

*Proof.* Because for any natural number $M$,

$$H_M = (\sqrt{a}, \ldots, \sqrt{a})^T (\sqrt{a}, \ldots, \sqrt{a}), \tag{D.5}$$

$H(\mathbf{x}, \mathbf{x}')$ is positive semidefinite. ■

**Theorem D.3.** *If* $H_1(\mathbf{x}, \mathbf{x}')$ *and* $H_2(\mathbf{x}, \mathbf{x}')$ *are positive semidefinite kernels,*

$$H(\mathbf{x}, \mathbf{x}') = a_1 H_1(\mathbf{x}, \mathbf{x}') + a_2 H_2(\mathbf{x}, \mathbf{x}') \tag{D.6}$$

*is also positive semidefinite, where* $a_1$ *and* $a_2$ *are positive.*

*Proof.* Because for any $M$, $h_i$, and $\mathbf{x}_i$

$$\mathbf{h}_M^T (a_1 H_{1M} + a_2 H_{2M}) \, \mathbf{h}_M = a_1 \mathbf{h}_M^T H_{1M} \mathbf{h}_M + a_2 \mathbf{h}_M^T H_{2M} \mathbf{h}_M \geq 0, \tag{D.7}$$

$H(\mathbf{x}, \mathbf{x}')$ is positive semidefinite. ■

**Theorem D.4.** *If* $H(\mathbf{x}, \mathbf{x}') = f(\mathbf{x}) \, f(\mathbf{x}')$, *where* $f(\mathbf{x})$ *is an arbitrary scalar function,* $H(\mathbf{x}, \mathbf{x}')$ *is positive semidefinite.*

*Proof.* Because for any $M$, $h_i$, and $\mathbf{x}_i$

$$\sum_{i,j=1}^{M} h_i \, h_j \, f(\mathbf{x}_i) \, f(\mathbf{x}_j) = \left( \sum_{i=1}^{M} h_i \, f(\mathbf{x}_i) \right)^2 \geq 0, \tag{D.8}$$

$H(\mathbf{x}, \mathbf{x}')$ is positive semidefinite. ■

**Theorem D.5.** *If* $H_1(\mathbf{x}, \mathbf{x}')$ *and* $H_2(\mathbf{x}, \mathbf{x}')$ *are positive semidefinite,*

$$H(\mathbf{x}, \mathbf{x}') = H_1(\mathbf{x}, \mathbf{x}') \, H_2(\mathbf{x}, \mathbf{x}') \tag{D.9}$$

*is also positive semidefinite.*

*Proof.* It is sufficient to show that if $M \times M$ matrices $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ are positive semidefinite, $\{a_{ij} \, b_{ij}\}$ is also positive semidefinite.

Because $A$ is positive semidefinite, $A$ is expressed by $A = F^T F$, where $F$ is an $M \times M$ matrix. Then $a_{ij} = \mathbf{f}_i^T \mathbf{f}_j$, where $\mathbf{f}_j$ is the $j$th column vector of $F$. Thus for arbitrary $h_1, \ldots, h_M$,

$$\sum_{i,j=1}^{M} h_i \, h_j \, \mathbf{f}_i^T \, \mathbf{f}_j \, b_{ij} = \sum_{i,j=1}^{M} (h_i \, \mathbf{f}_i)^T (h_j \, \mathbf{f}_j) \, b_{ij} \geq 0. \; ■ \tag{D.10}$$

*Example D.6.* The linear kernel $H(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}$ is positive semidefinite because $H_M = (\mathbf{x}_1, \dots, \mathbf{x}_M)^T (\mathbf{x}_1, \dots, \mathbf{x}_M)$. Thus, from Theorems D.2 to D.5 the polynomial kernel given by $H(\mathbf{x}, \mathbf{x}') = (1+\mathbf{x}^T \mathbf{x}')^d$ is positive semidefinite.

**Corollary D.7.** *If $H(\mathbf{x}, \mathbf{x}')$ and $H'(\mathbf{y}, \mathbf{y}')$ are positive semidefinite kernels, where $\mathbf{x}$ and $\mathbf{y}$ may be of different dimensions, $H(\mathbf{x}, \mathbf{x}')\, H'(\mathbf{y}, \mathbf{y}')$ is also a positive semidefinite kernel.*

**Corollary D.8.** *Let $H(\mathbf{x}, \mathbf{x}')$ be positive semidefinite and satisfy*

$$|H(\mathbf{x}, \mathbf{x}')| \leq \rho, \tag{D.11}$$

*where $\rho > 0$. Then if*

$$f(y) = \sum_{i=1}^{\infty} a_i\, y^i \tag{D.12}$$

*converges for $|y| \leq \rho$, where $a_i \geq 0$ for all integers $i$, the composed kernel $f(H(\mathbf{x}, \mathbf{x}'))$ is also positive semidefinite.* ■

*Proof.* From Theorem D.5, $H^i(\mathbf{x}, \mathbf{x}')$ is positive semidefinite. Then from Theorem D.5,

$$\sum_{i=0}^{N} a_i\, H^i(\mathbf{x}, \mathbf{x}') \tag{D.13}$$

is positive semidefinite for all integers $N$. Therefore, so is $f(H(\mathbf{x}, \mathbf{x}'))$. ■

From Corollary D.8, especially for positive semidefinite kernel $H(\mathbf{x}, \mathbf{x}')$, $\exp(H(\mathbf{x}, \mathbf{x}'))$ is also positive semidefinite.

In the following we clarify the relations between positive and conditionally positive semidefinite kernels.

**Lemma D.9.** *Let*

$$H(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}, \mathbf{x}') + K(\mathbf{x}_0, \mathbf{x}_0) - K(\mathbf{x}, \mathbf{x}_0) - K(\mathbf{x}', \mathbf{x}_0). \tag{D.14}$$

*Then $H(\mathbf{x}, \mathbf{x}')$ is positive semidefinite, if and only if $K(\mathbf{x}, \mathbf{x}')$ is conditionally positive semidefinite.*

*Proof.* For $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ and $\mathbf{h}_M = (h_1, \dots, h_M)^T$ with

$$\sum_{i=1}^{M} h_i = 0, \tag{D.15}$$

we have

$$\mathbf{h}_M^T H_M \mathbf{h}_M = \mathbf{h}_M^T K_M \mathbf{h}_M. \tag{D.16}$$

Thus, if $H(\mathbf{x}, \mathbf{x}')$ is positive semidefinite, $K(\mathbf{x}, \mathbf{x}')$ is conditionally positive semidefinite.

On the other hand, suppose that $K(\mathbf{x}, \mathbf{x}')$ is conditionally positive semidefinite. Then for $\{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ and $\mathbf{h}_M = (h_1, \ldots, h_M)^T$ with

$$h_0 = -\sum_{i=1}^{M} h_i, \tag{D.17}$$

we have

$$\begin{aligned} 0 &\leq \sum_{i,j=0}^{M} h_i\, h_j\, K(\mathbf{x}_i, \mathbf{x}_j) \\ &= \sum_{i,j=1}^{M} h_i\, h_j\, K(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{M} h_i\, h_0\, K(\mathbf{x}_i, \mathbf{x}_0) + \sum_{j=1}^{M} h_0\, h_j\, K(\mathbf{x}_0, \mathbf{x}_j) \\ &\quad + h_0^2\, K(\mathbf{x}_0, \mathbf{x}_0) \\ &= \sum_{i,j=1}^{M} h_i\, h_j\, H(\mathbf{x}_i, \mathbf{x}_j). \end{aligned} \tag{D.18}$$

Therefore, $H(\mathbf{x}, \mathbf{x}')$ is positive semidefinite. ■

**Theorem D.10.** *Kernel $K(\mathbf{x}, \mathbf{x}')$ is conditionally positive semidefinite if and only if $\exp(\gamma\, K(\mathbf{x}, \mathbf{x}'))$ is positive semidefinite for any positive $\gamma$.*

*Proof.* If $\exp(\gamma\, K(\mathbf{x}, \mathbf{x}'))$ is positive semidefinite, $\exp(\gamma\, K(\mathbf{x}, \mathbf{x}')) - 1$ is conditionally positive semidefinite. So is the limit

$$K(\mathbf{x}, \mathbf{x}') = \lim_{\gamma \to +0} \frac{\exp(\gamma\, K(\mathbf{x}, \mathbf{x}')) - 1}{\gamma}. \tag{D.19}$$

Now let $K(\mathbf{x}, \mathbf{x}')$ be conditionally positive semidefinite and choose some $\mathbf{x}_0$ and $H(\mathbf{x}, \mathbf{x}')$ as in Lemma D.9. Then for positive $\gamma$

$$\gamma K(\mathbf{x}, \mathbf{x}') = \gamma H(\mathbf{x}, \mathbf{x}') - \gamma K(\mathbf{x}_0, \mathbf{x}_0) + \gamma K(\mathbf{x}, \mathbf{x}_0) + \gamma K(\mathbf{x}', \mathbf{x}_0). \tag{D.20}$$

Thus,

$$\begin{aligned} \exp(\gamma K(\mathbf{x}, \mathbf{x}')) &= \exp(\gamma H(\mathbf{x}, \mathbf{x}'))\, \exp(-\gamma K(\mathbf{x}_0, \mathbf{x}_0)) \\ &\quad \times \exp(\gamma K(\mathbf{x}, \mathbf{x}_0))\, \exp(\gamma K(\mathbf{x}', \mathbf{x}_0)). \end{aligned} \tag{D.21}$$

From Theorems D.4 and D.5 and Corollary D.8, $\exp(\gamma K(\mathbf{x}, \mathbf{x}'))$ is positive semidefinite. ■

*Example D.11.* Kernel $H(\mathbf{x}, \mathbf{x}') = -\|\mathbf{x} - \mathbf{x}'\|^2$ is conditionally positive semidefinite because for $\sum_i^M h_i = 0$,

$$\begin{aligned}
\mathbf{h}_M^T H_M \mathbf{h}_M &= -\sum_{i=1}^{M} h_i h_j \|\mathbf{x}_i - \mathbf{x}_j\|^2 \\
&= -\sum_{i,j=1}^{M} h_i h_j (\mathbf{x}_i^T \mathbf{x}_i - 2\mathbf{x}_i^T \mathbf{x}_j + \mathbf{x}_j^T \mathbf{x}_j) \\
&= 2\left(\sum_{i=1}^{M} h_i \mathbf{x}_i\right)^T \left(\sum_{i=1}^{M} h_i \mathbf{x}_i\right) \geq 0. \qquad \text{(D.22)}
\end{aligned}$$

Thus, $\exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$ is positive semidefinite.

## D.2 Reproducing Kernel Hilbert Space

Because a positive semidefinite kernel has the associated feature space called the *reproducing kernel Hilbert space (RKHS)*, support vector machines can determine the optimal hyperplane in that space using the kernel trick. In this section, we discuss reproducing kernel Hilbert spaces for positive and conditionally positive semidefinite kernels.

For the positive semidefinite kernels, the following theorem holds.

**Theorem D.12.** *Let $X$ be the input space and $H(\mathbf{x}, \mathbf{x}')\,(\mathbf{x}, \mathbf{x}' \in X)$ be a positive semidefinite kernel. Let $H_0$ be the space spanned by the functions $\{H_\mathbf{x} \,|\, \mathbf{x} \in X\}$ where*

$$H_\mathbf{x}(\mathbf{x}') = H(\mathbf{x}, \mathbf{x}'). \qquad \text{(D.23)}$$

*Then there exist a Hilbert space $H$, which is a complete space of $H_0$, and the mapping from $X$ to $H$ such that*

$$H(\mathbf{x}, \mathbf{x}') = \langle H_\mathbf{x}, H_{\mathbf{x}'} \rangle. \qquad \text{(D.24)}$$

*Here, instead of $\mathbf{x}^T \mathbf{x}'$, we use $\langle \mathbf{x}, \mathbf{x}' \rangle$ to denote the dot-product.*

*Proof.* Let $H_\mathbf{x}(\mathbf{x}') = H(\mathbf{x}, \mathbf{x}')$ and $H_0$ be a linear subspace generated by the functions $\{H_\mathbf{x} \,|\, \mathbf{x} \in X\}$. Then for $f, g \in H_0$ expressed by

$$f = \sum_{\mathbf{x}_i \in X} c_i H_{\mathbf{x}_i}, \qquad \text{(D.25)}$$

$$g = \sum_{\mathbf{x}'_j \in X} d_j H_{\mathbf{x}'_j}, \qquad \text{(D.26)}$$

we define the dot-product as follows:

$$\begin{aligned}
\langle f, g \rangle &= \sum_{\mathbf{x}'_j \in X} d_j f(\mathbf{x}'_j) \\
&= \sum_{\mathbf{x}_i, \mathbf{x}'_j \in X} c_i d_j H(\mathbf{x}_i, \mathbf{x}'_j) \\
&= \sum_{\mathbf{x}_i \in X} c_i g(\mathbf{x}_i). \qquad \text{(D.27)}
\end{aligned}$$

Now we show that (D.27) satisfies the properties of the dot-product. Clearly, (D.27) is symmetric and linear. Also, according to the assumption of $H(\mathbf{x}, \mathbf{x}')$ being positive semidefinite,

$$\langle f, f \rangle = \sum_{\mathbf{x}_i, \mathbf{x}_j \in X} c_i \, c_j \, H(\mathbf{x}_i, \mathbf{x}_j) \geq 0 \tag{D.28}$$

is satisfied. Here, the strict equality holds if and only if $f$ is identically zero. Thus, (D.27) is the dot-product. Hence, $H_0$ is a pre-Hilbert space and its completion $H$ is a Hilbert space, which is called *RKHS associated with* $H_{\mathbf{x}}$.

From (D.27) the following reproducing property is readily obtained:

$$\langle f, H_{\mathbf{x}} \rangle = f(\mathbf{x}). \tag{D.29}$$

In particular,

$$\langle H_{\mathbf{x}}, H_{\mathbf{x}'} \rangle = H(\mathbf{x}, \mathbf{x}'). \blacksquare \tag{D.30}$$

For a conditionally positive semidefinite kernel, for $f \in H_0$ the following theorem holds.

**Theorem D.13.** *Let $H(\mathbf{x}, \mathbf{x}')\,(\mathbf{x}, \mathbf{x}' \in X)$ be a conditionally positive semidefinite kernel. Then there exist a Hilbert space $H$ and a mapping $K_{\mathbf{x}}$ from $X$ to $H$ such that*

$$H(\mathbf{x}, \mathbf{x}') - \frac{1}{2} H(\mathbf{x}, \mathbf{x}) - \frac{1}{2} H(\mathbf{x}', \mathbf{x}') = -\|K_{\mathbf{x}} - K_{\mathbf{x}'}\|^2. \tag{D.31}$$

*Proof.* For $\mathbf{x}_0$ we define

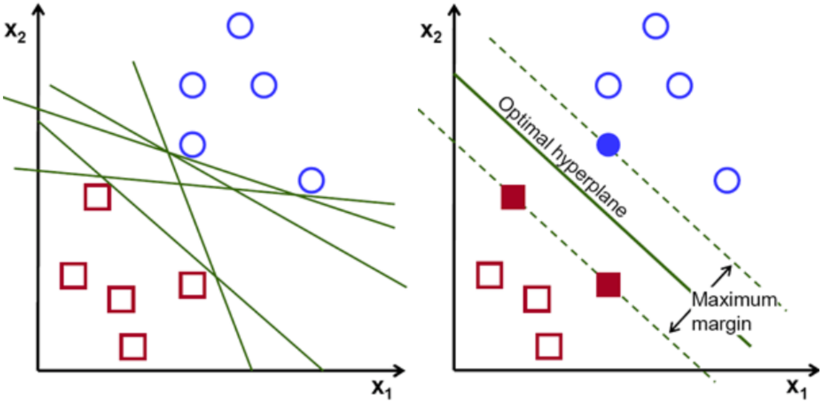
$$K(\mathbf{x}, \mathbf{x}') = \frac{1}{2} \left( H(\mathbf{x}, \mathbf{x}') + H(\mathbf{x}_0, \mathbf{x}_0) - H(\mathbf{x}, \mathbf{x}_0) - H(\mathbf{x}', \mathbf{x}_0) \right), \tag{D.32}$$

which is a positive semidefinite kernel from Lemma D.9. Let $H$ be the associated RKHS for $K(\mathbf{x}, \mathbf{x}')$ and $K_{\mathbf{x}}(\mathbf{x}') = K(\mathbf{x}, \mathbf{x}')$. Then

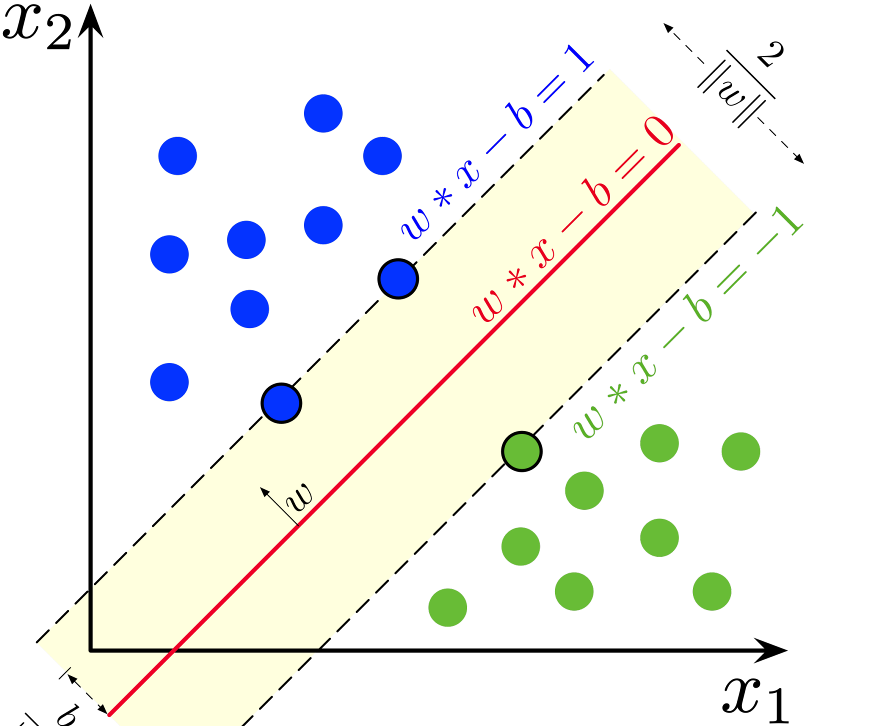
$$\begin{aligned} \|K_{\mathbf{x}} - K_{\mathbf{x}'}\|^2 &= K(\mathbf{x}, \mathbf{x}) + K(\mathbf{x}', \mathbf{x}') - 2K(\mathbf{x}, \mathbf{x}') \\ &= -H(\mathbf{x}, \mathbf{x}') + \frac{1}{2} H(\mathbf{x}, \mathbf{x}) + \frac{1}{2} H(\mathbf{x}', \mathbf{x}'). \end{aligned} \tag{D.33}$$

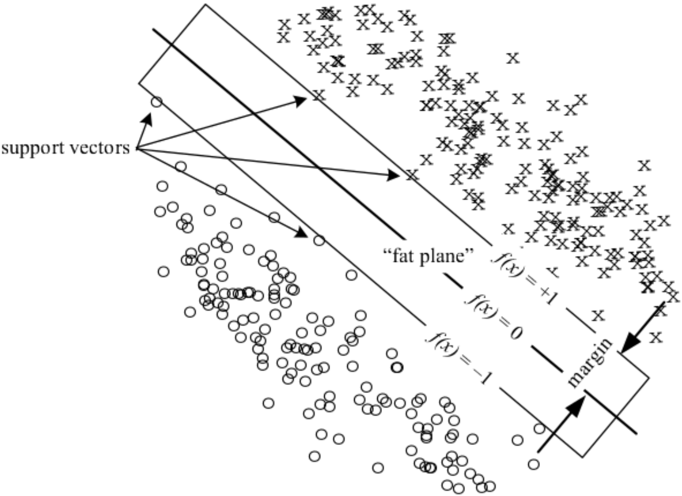
Thus the theorem holds. $\blacksquare$

# Support vector machine

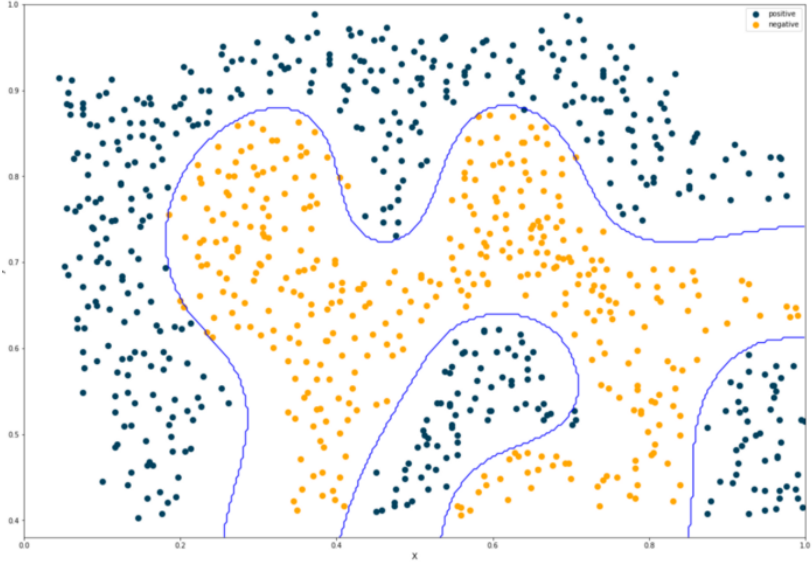


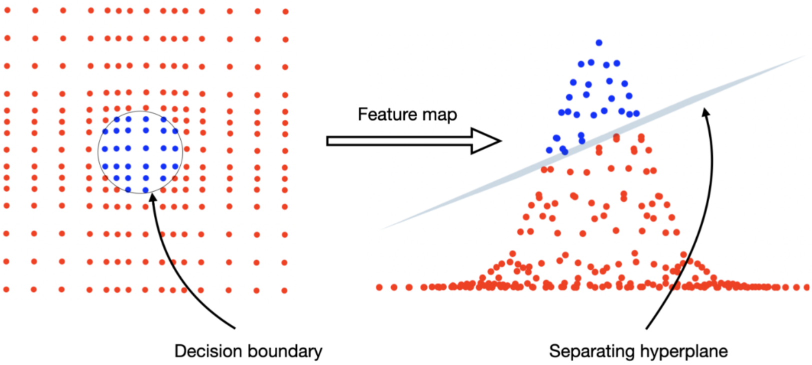
$x_2$
$x_1$
$x_2$
Optimal hyperplane
Maximum
margin
$x_1$


$x_2$
$x_1$
$w * x - b = 1$
$w * x - b = 0$
$w * x - b = -1$
$\frac{2}{\|w\|}$
$w$



support vectors
"fat plane"
f(x) = +1
f(x) = 0
f(x) = −1
margin


positive
negative
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.2
0.4
0.6
0.8
1.0
X




$$\Phi(x, y) \equiv (x, y, e^{-(x^2+y^2)}),\ R^2 \rightarrow R^3$$

Polynomial kernel

$$k_p(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^d$$

Polynomial kernel

$$k_p(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^d$$

$$\Phi(\mathbf{x}) = \begin{bmatrix} 1 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \sqrt{2}x_1 x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix}$$

$$R^2 \rightarrow R^6$$

Polynomial kernel

$$k_p(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^d$$

$$\Phi(\mathbf{x}) = \begin{bmatrix} 1 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \sqrt{2}x_1x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix}$$

$$R^2 \rightarrow R^6$$

$$k(\mathbf{x}, \mathbf{y}) = \left[1 \sqrt{2}x_1 \sqrt{2}x_2 \sqrt{2}x_1x_2 x_1^2 x_2^2\right] \cdot \begin{bmatrix} 1 \\ \sqrt{2}y_1 \\ \sqrt{2}y_2 \\ \sqrt{2}y_1y_2 \\ y_1^2 \\ y_2^2 \end{bmatrix} =$$

$$= 1 + 2\ \mathrm{x}_1 y_1 + 2x_2y_2 + 2x_1x_2y_1y_2 + x_1^2y_1^2 + x_2^2y_2^2 = (\mathbf{x} \cdot \mathbf{y} + 1)^2$$

# Lagrange theory (refresher) [2]

$$\text{minimize} : f(\mathbf{w})$$
$$\text{under the condition} : g_j(\mathbf{w}) \leq 0$$
$$h_k(\mathbf{w}) = 0$$

[2] Joseph Louis Lagrange; Giuseppe Lodovico Lagrangia

$$
\begin{aligned}
&\text{minimize} : f(\mathbf{w}) \\
&\text{under the condition} : g_j(\mathbf{w}) \leq 0 \\
&h_k(\mathbf{w}) = 0
\end{aligned}
$$

Introduce the Lagrangian:

$$\mathcal{L} \equiv \frac{1}{2} f(\mathbf{w}) + \sum_j \alpha_j g_j(\mathbf{w}) + \sum_k \beta_k h_k(\mathbf{w})$$

and minimize

$$\frac{\partial \mathcal{L}}{\partial w_i} = 0, \ \frac{\partial \mathcal{L}}{\partial \beta_k} = 0$$

[2] Joseph Louis Lagrange; Giuseppe Lodovico Lagrangia

Eliminating $\omega$ we obtain:

$$\begin{aligned} &\text{maximize} : \mathcal{L}(\alpha, \beta) \\ &\text{under the condition} : \alpha_j \geq 0 \text{ for each } j \end{aligned}$$

Eliminating $\omega$ we obtain:

$$\text{maximize} : \mathcal{L}(\alpha, \beta)$$
$$\text{under the condition} : \alpha_j \geq 0 \text{ for each } j$$

If we manage to find the optimal $\hat{\alpha}, \hat{\beta}$ then we can find $\hat{w}$ and $b$ from

$$f(\hat{w}) = \mathcal{L}(\hat{\alpha}, \hat{\beta})$$
$$\alpha_j g_j(\hat{\omega}) = 0 \text{ за всяко } j$$

Eliminating $\omega$ we obtain:

$$\begin{aligned}&\text{maximize} : \mathcal{L}(\alpha, \beta)\\&\text{under the condition} : \alpha_j \geq 0 \text{ for each } j\end{aligned}$$

If we manage to find the optimal $\hat{\alpha}, \hat{\beta}$ then we can find $\hat{w}$ and $b$ from

$$\begin{aligned}&f(\hat{w}) = \mathcal{L}(\hat{\alpha}, \hat{\beta})\\&\alpha_j g_j(\hat{\omega}) = 0 \text{ за всяко } j\end{aligned}$$

For a given $j$ from the second condition it follows that $\alpha_j \equiv 0$ or $g_j(\omega) \equiv 0$.

# SVM theory (in two slides)



$$\mathcal{L} = \frac{1}{2}\mathbf{w}.\mathbf{w} + \sum_i \alpha_i [1 - y_i(\mathbf{w}.\mathbf{x}_i + b)]$$

# SVM theory (in two slides)



$$\mathcal{L} = \frac{1}{2}\mathbf{w}.\mathbf{w} + \sum_i \alpha_i [1 - y_i(\mathbf{w}.\mathbf{x}_i + b)]$$

$$0 = \frac{\partial \mathcal{L}}{\partial w_i} = w_i - \sum_i \alpha_i y_i x_i \tag{2}$$

$$0 = \frac{\partial \mathcal{L}}{\partial b} = \sum_i \alpha_i y_i$$

$$\mathcal{L} = \frac{1}{2}\mathbf{w}.\mathbf{w} + \sum_i \alpha_i[1 - y_i(\mathbf{w}.\mathbf{x}_i + b)]$$

$$0 = \frac{\partial \mathcal{L}}{\partial w_i} = w_i - \sum_i \alpha_i y_i x_i \tag{2}$$

$$0 = \frac{\partial \mathcal{L}}{\partial b} = \sum_i \alpha_i y_i$$

After expressing $w$ and $b$ with $\alpha$:

$$\mathcal{L}(\alpha) = \sum_j \alpha_j - \frac{1}{2}\sum_{j,k} \alpha_j y_j (x_j.x_k) y_k \alpha_k \equiv$$

$$\equiv \mathbf{e}.\alpha - \frac{1}{2}\alpha.(\mathbf{y}).\mathbf{G}.(\mathbf{y}).\alpha$$

$$\mathcal{L} = \frac{1}{2}\mathbf{w}.\mathbf{w} + \sum_i \alpha_i[1 - y_i(\mathbf{w}.\mathbf{x}_i + b)]$$

$$0 = \frac{\partial \mathcal{L}}{\partial w_i} = w_i - \sum_i \alpha_i y_i x_i \tag{2}$$

$$0 = \frac{\partial \mathcal{L}}{\partial b} = \sum_i \alpha_i y_i$$

After expressing $w$ and $b$ with $\alpha$:

$$\mathcal{L}(\alpha) = \sum_j \alpha_j - \frac{1}{2}\sum_{j,k} \alpha_j y_j (x_j.x_k) y_k \alpha_k \equiv$$

$$\equiv \mathbf{e}.\alpha - \frac{1}{2}\alpha.(\mathbf{y}).\mathbf{G}.(\mathbf{y}).\alpha$$

$$G_{ij} \equiv x_i x_j$$

$$\text{minimize} : \frac{1}{2}\alpha.(\mathbf{y}).\mathbf{G}.(\mathbf{y}).\alpha - \mathbf{e}.\alpha$$

$$\text{under the condition} : \alpha_i > 0$$

$$\alpha.\mathbf{y} = 0$$

$$\text{minimize} : \frac{1}{2}\alpha.(\mathbf{y}).\mathbf{G}.(\mathbf{y}).\alpha - \mathbf{e}.\alpha$$

$$\text{under the condition} : \alpha_i > 0$$

$$\alpha.\mathbf{y} = 0$$

We find $w$ from (2) and $b$ from:

$$\alpha_i[y_i(wx_i + b) - 1] = 0$$

$$\text{minimize} : \frac{1}{2}\alpha.(\mathbf{y}).\mathbf{G}.(\mathbf{y}).\alpha - \mathbf{e}.\alpha$$
$$\text{under the condition} : \alpha_i > 0$$
$$\alpha.\mathbf{y} = 0$$

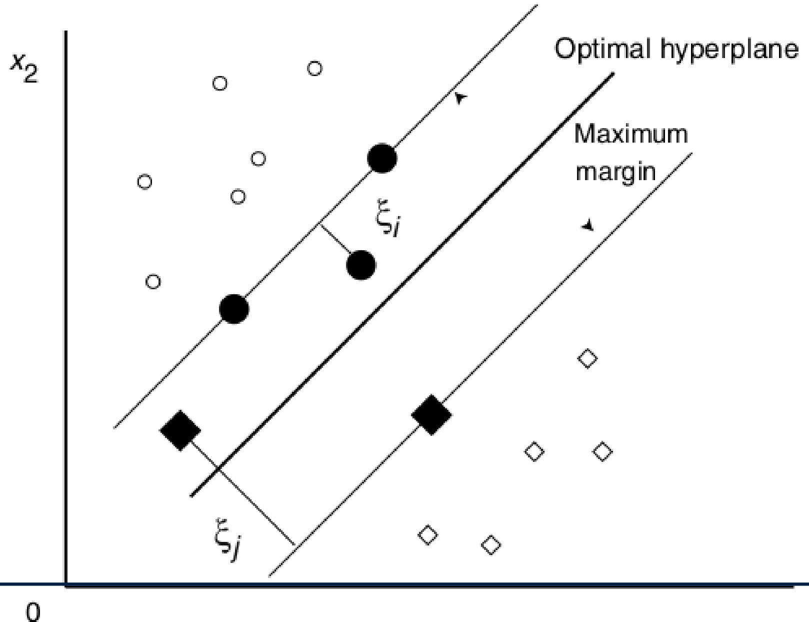
We find $w$ from (2) and $b$ from:

$$\alpha_i[y_i(wx_i + b) - 1] = 0$$

Notes:

- The points $i$ with $\alpha_i \neq 0$ are the support vectors
- $x_i$ enter the Lagrangian via the Gram matrix only.

# Soft margin SVM



# Soft margin SVM



Minimize

$$\mathcal{L}(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{m}\xi_i$$

$$-\sum_{i=1}^{m}\alpha_i(y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \xi_i) - \sum_{i=1}^{m}\beta_i\xi_i$$

# Soft margin SVM

Minimize

$$\mathcal{L}(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{m}\xi_i$$
$$-\sum_{i=1}^{m}\alpha_i(y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \xi_i) - \sum_{i=1}^{m}\beta_i\xi_i$$

In dual representation:

$$\mathcal{L}(\alpha) = \sum_{i=1}^{M}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{M}\alpha_i\alpha_j\mathbf{x}_i^T\mathbf{x}_j y_i y_j \quad (3)$$

under the conditions

$$\sum y_i\alpha_i = 0, C \geq \alpha_i \geq 0 \quad (4)$$

The Lagrangian

$$\mathcal{L}(\alpha) = \sum_{j} \alpha_j - \frac{1}{2} \sum_{j,k} \alpha_j y_j (x_j . x_k) y_k \alpha_k \tag{5}$$

The Lagrangian

$$\mathcal{L}(\alpha) = \sum_j \alpha_j - \frac{1}{2} \sum_{j,k} \alpha_j y_j (x_j . x_k) y_k \alpha_k \qquad (5)$$

is replaced with

$$\mathcal{L}(\alpha) = \sum_j \alpha_j - \frac{1}{2} \sum_{j,k} \alpha_j y_j H(x_j, x_k) y_k \alpha_k \qquad (6)$$

and the classes (i.e. $y_i$) can be separated not by the plane

$$D(\mathbf{x}) = \mathbf{w}\mathbf{x} + b \equiv \sum_i \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \qquad (7)$$

The Lagrangian

$$\mathcal{L}(\alpha) = \sum_j \alpha_j - \frac{1}{2} \sum_{j,k} \alpha_j y_j (x_j . x_k) y_k \alpha_k \tag{5}$$

is replaced with

$$\mathcal{L}(\alpha) = \sum_j \alpha_j - \frac{1}{2} \sum_{j,k} \alpha_j y_j H(x_j, x_k) y_k \alpha_k \tag{6}$$

and the classes (i.e. $y_i$) can be separated not by the plane

$$D(\mathbf{x}) = \mathbf{w}\mathbf{x} + b \equiv \sum_i \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \tag{7}$$

but by the surface

$$D(\mathbf{x}) = \mathbf{w} g(\mathbf{x}) + b \equiv \sum_i \alpha_i y_i H(\mathbf{x}_i, \mathbf{x}) + b \tag{8}$$

# Quantum Circuits and Kernels

`https://learn.qiskit.org/course/machine-learning/quantum-feature-maps-kernels`

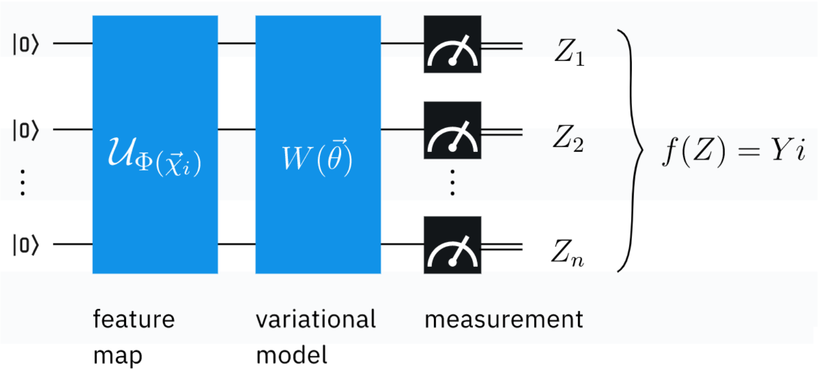
*Vojtech Havlicek, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow and Jay M. Gambetta, Supervised learning with quantum enhanced feature spaces, Nature 567, 209-212 (2019), doi.org:10.1038/s41586-019-0980-2, arXiv:1804.11326.*

# Quantum Circuits and Kernels

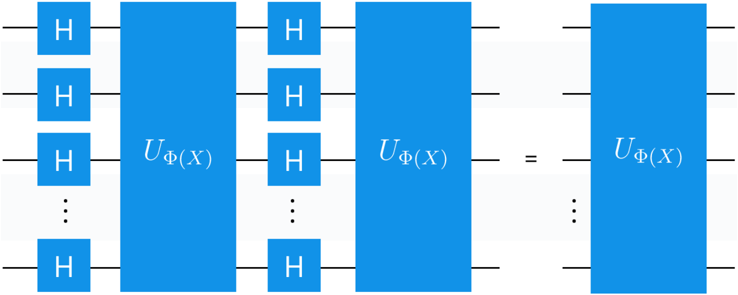
`https://learn.qiskit.org/course/machine-learning/quantum-feature-maps-kernels`

*Vojtech Havlicek, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow and Jay M. Gambetta, Supervised learning with quantum enhanced feature spaces, Nature 567, 209-212 (2019), doi.org:10.1038/s41586-019-0980-2, arXiv:1804.11326.*

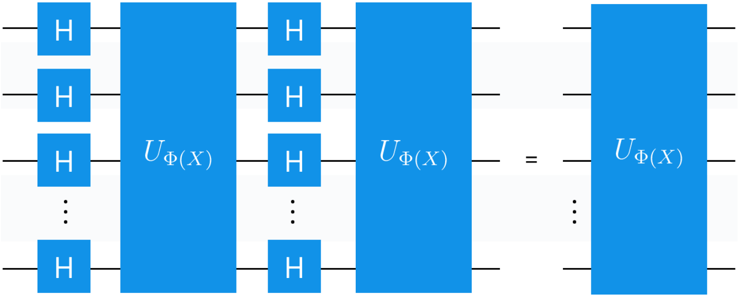
In quantum machine learning, a quantum feature map, $\phi(\vec{x})$, maps a classical feature vector, $\vec{x}$, to a quantum Hilbert space, $|\phi(\vec{x})\rangle\langle\phi(\vec{x})|$. The quantum feature map transforms $\vec{x} \rightarrow |\phi(\vec{x})\rangle$ using a unitary transformation $\vec{U}_{\phi}(\vec{x})$, which is typically a parameterized quantum circuit.



|0⟩
|0⟩
|0⟩
$\mathcal{U}_{\Phi(\vec{\chi}_i)}$
$W(\vec{\theta})$
$Z_1$
$Z_2$
$Z_n$
$f(Z) = Yi$
feature map
variational model
measurement



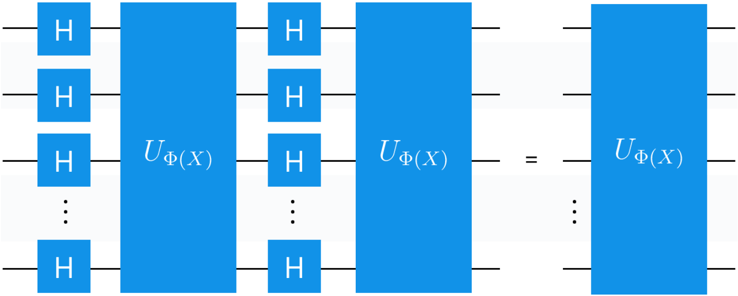

$$\mathcal{U}_{\Phi(\vec{x})} = \prod_{d} U_{\Phi(\vec{x})} H^{\otimes n}, \; U_{\Phi(\vec{x})} = \exp\left( i \sum_{S \subseteq [n]} \phi_S(\vec{x}) \prod_{k \in S} P_i \right)$$




$$\mathcal{U}_{\Phi(\vec{x})} = \prod_d U_{\Phi(\vec{x})} H^{\otimes n},\ U_{\Phi(\vec{x})} = \exp\left(i \sum_{S \subseteq [n]} \phi_S(\vec{x}) \prod_{k \in S} P_i\right)$$

Within the entangling blocks, $U_{\Phi(\vec{x})}$: $P_i \in \{I, X, Y, Z\}$ denotes the Pauli matrices, the index $S$ describes connectivity between different qubits or data points: $S \in \{\binom{n}{k} \text{ combinations},\ k = 1, \dots n\}$, and by default the data mapping function $\phi_S(\vec{x})$ is

$$\phi_S : \vec{x} \mapsto \begin{cases} x_i & \text{if } S = \{i\} \\ (\pi - x_i)(\pi - x_j) & \text{if } S = \{i, j\} \end{cases}$$




$$\mathcal{U}_{\Phi(\vec{x})} = \prod_d U_{\Phi(\vec{x})} H^{\otimes n},\ U_{\Phi(\vec{x})} = \exp\left( i \sum_{S \subseteq [n]} \phi_S(\vec{x}) \prod_{k \in S} P_i \right)$$

Within the entangling blocks, $U_{\Phi(\vec{x})}$: $P_i \in \{I, X, Y, Z\}$ denotes the Pauli matrices, the index $S$ describes connectivity between different qubits or data points: $S \in \{\binom{n}{k} \text{ combinations},\ k = 1, \dots n\}$, and by default the data mapping function $\phi_S(\vec{x})$ is

$$\phi_S : \vec{x} \mapsto \begin{cases} x_i & \text{if } S = \{i\} \\ (\pi - x_i)(\pi - x_j) & \text{if } S = \{i, j\} \end{cases}$$

when $k = 2, P_0 = Z, P_1 = ZZ$, this is the `ZZFeatureMap` in Qiskit:

$$\mathcal{U}_{\Phi(\vec{x})} = \left( \exp\left( i \sum_{jk} \phi_{\{j,k\}}(\vec{x})\, Z_j \otimes Z_k \right) \exp\left( i \sum_j \phi_{\{j\}}(\vec{x})\, Z_j \right) H^{\otimes n} \right)^d$$

THANK YOU FOR YOUR ATTENTION!

БЛАГОДАРЯ ЗА ВНИМАНИЕТО!