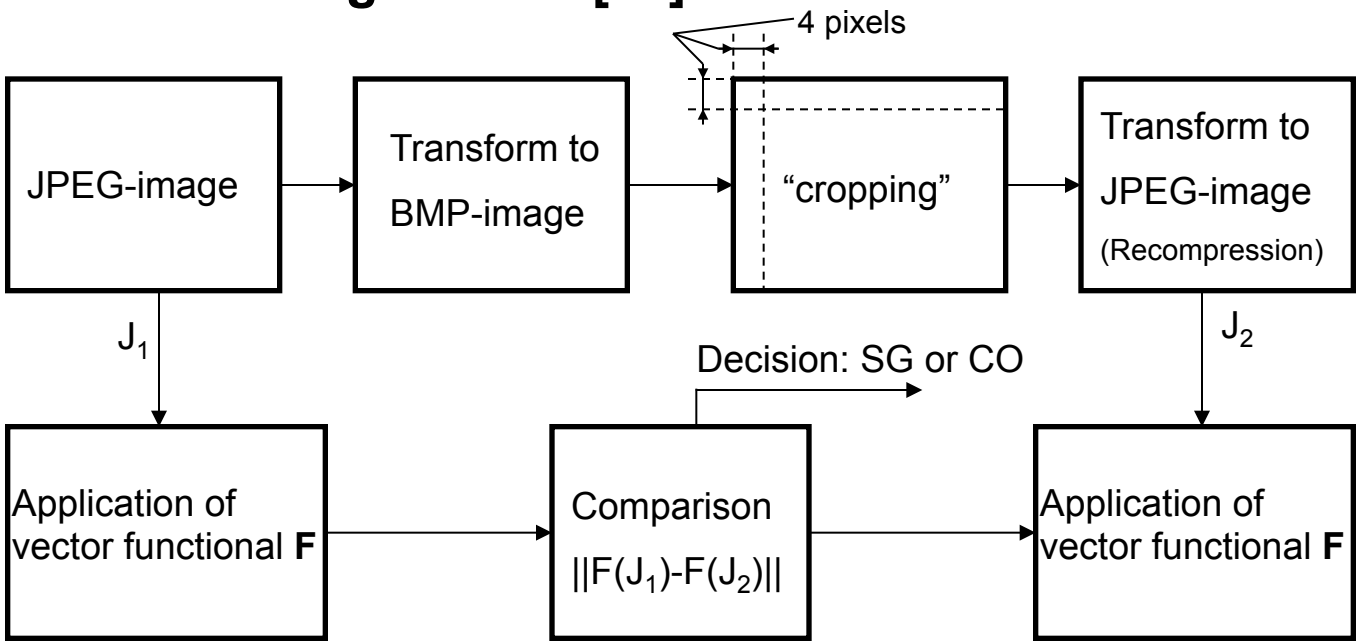
Lecture 7. Blind steganalysis

Three main classes of steganalytic methods:

1. *Targeted* steganalysis.
(It is developed after detail study of particular SG using differences between statistic of CO and SG. Examples of such methods are χ^2-attack on Jsteg and sample pair analysis of SG-LSB .
2. *Distinguishing* steganalysis.
(SG is processing in such a way to get approximation of CO, in which SG was embedded. Next an estimation of CO and SG are compared on their statistics.)
3. «Blind» steganalysis.
(Initially it is performed a "teaching" procedure on a large set of SG and corresponding to them CO that allows to specify the distinguishing algorithm optimal in some sense. Next, after presentation of sample, this algorithm takes a decision if it is SG or CO. *Fisher discriminant* or *Support Vector Machine* can be used as such algorithms).

Estimation of CO given SG [16].



The logic behind this choice: Fig. 1

The cropping and recompression should produce a "calibrated" image with most macroscopic features similar to the original cover image. This is because the cropped stego image is perceptually similar to the cover image and thus its DCT coefficients should have approximately the same statistical properties as the cover image. The cropping by 4 pixels is important because the 8x8 grid of recompression "does not see" the previous JPEG compression and thus the obtained DCT coefficients are not influenced by previous quantization (and embedding) in the DCT domain. (There are also other method of image calibration (see [57])).

Staganalytic classifier (Support Vector Machine-SVM) [17,18].

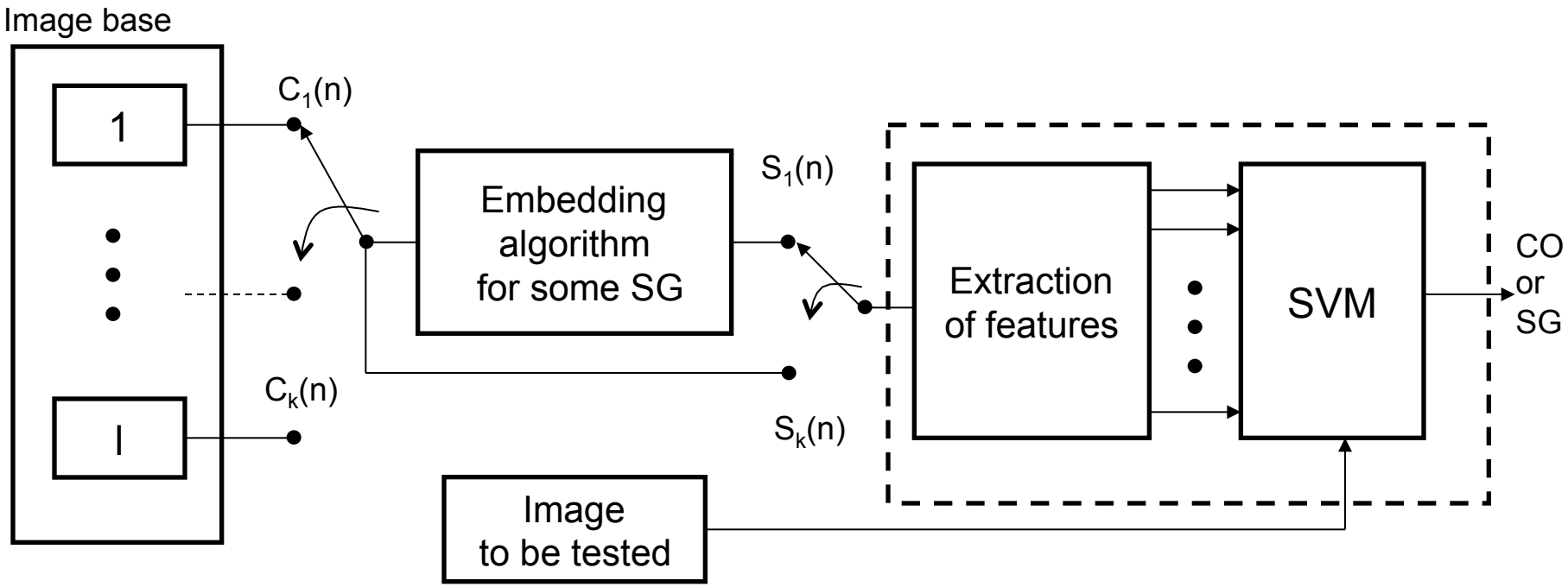


Рис. 2

Two phase of SG detecting:

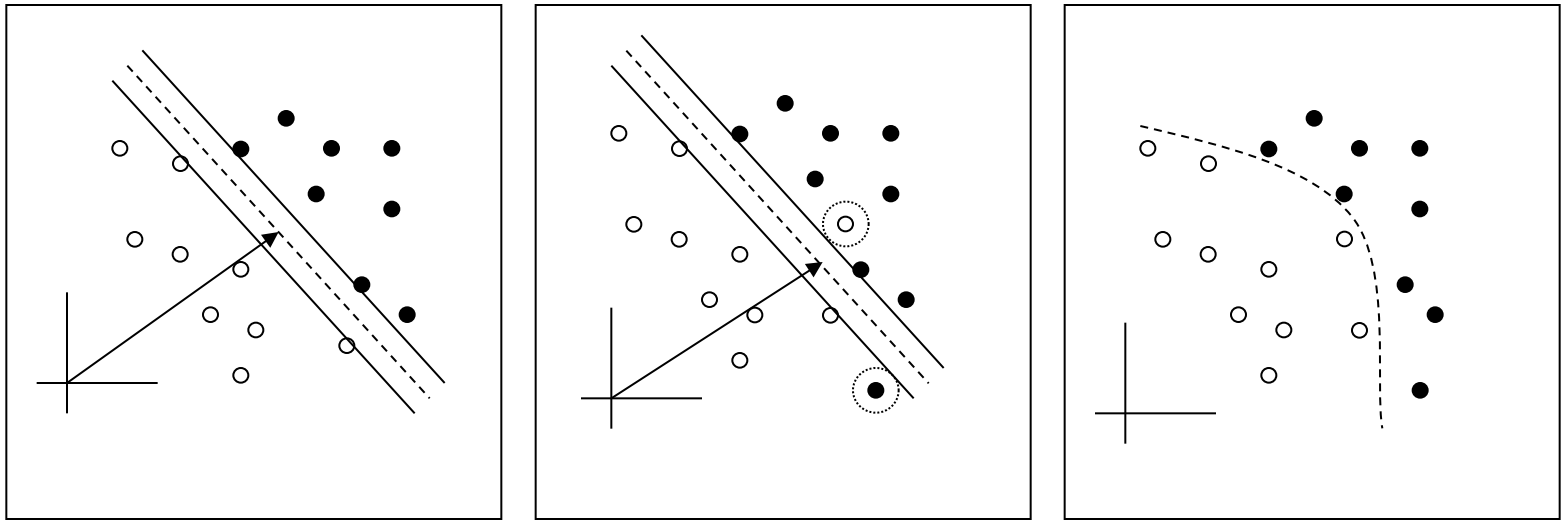
1. Training: $\{\mathbf{x}_i, y_i\}$, i=1,2,..k, $y_i \in \{-1,1\}$, $\mathbf{x}_i \in \mathbf{R}_d$ (feature vector). y_i=1, if $\mathbf{x}_i$=SG, y_i=-1, if $\mathbf{x}_i$=CO
2. Taking a decision: S(n)→$\mathbf{x} \in$CO or SG (?)

Classifier based on SVM.

The main types of SVM:

1. The linear separable SVM
2. The linear non-separable SVM
3. Nonlinear SVM

Illustration of the main types of SVM for two-dimensional feature space



a) Type 1 b) Type 2 c) Type 3

Fig. 3

SVM simply looks for the separating hyperplane (case1,2) or hypercurve (case 3) with largest margin between pairs ($\mathbf{x}_i$, +1) and ($\mathbf{x}_i$, -1).

SVM implementaion.

1. Linear separable SVM.

Equation for hyperplane in R^d:

$\mathbf{wx}+b=0$, (1)

where $\mathbf{w}=(w_1, w_2,.., w_d)$

$\mathbf{x}=(x_1, x_2,.., x_d)$

$b \in R^1$ (real value)

With geometric point of view **w** is a perpendicular to hyperplane , $|b|/||\mathbf{w}||$ - is the shortest distance from hyperplane to origin, $||.||$ - is Euclidean metric in R^d

Definition. The "*margin*" of a separating hyperplane between sets $X_{+1}=(\mathbf{x_i}, +1)$ and $X_{-1}=(\mathbf{x}_i, -1)$, i=1,2,..,k is $m=m_+ + m_-$, where m_+ is the shortest distance from the separating hyperplane to the closest point $\mathbf{x_i} \in X_{+1}$, while m_- is the shortest distance from the same hyperplane to the closest point $\mathbf{x_i} \in X_{-1}$.

SVM for the case 1 creates such hyperplane that maximizes *m* (margin).

If there exists a hyperplane separating samples of CO and SG (as it is the case 1), then the following inequalities hold:

$$\left.\begin{array}{l} \mathbf{w}^t\mathbf{x}_i + b \geq 1, \text{ if } y_i = 1 \\ \mathbf{w}^t\mathbf{x}_i + b \leq -1, \text{ if } y_i = -1 \end{array}\right\} (\mathbf{w}^t\mathbf{x}_i + b)y_i - 1 \geq 0, i = 1, 2, ..., k \qquad (2)$$

For any hyperplane satisfying to (2) the margin is simply 2/||**w**||.
A maximization of the margin is equivalently to a minimization of the following Lagrangian with respect to **w** and b :

$$L(\mathbf{w}, b, \alpha_1, ..\alpha_k) = \frac{1}{2} \| \mathbf{w} \|^2 - \sum_{i=1}^{k} \alpha_i (\mathbf{w}^t\mathbf{x}_i + b) y_i + \sum_{i=1}^{N} \alpha_i \qquad (3)$$

SVM (for the case 1) solves namely this problem although for some modification of, (3) .

Classification rule:
*y=sgn(**wx**+b)* (4)
We remember that if y=1, then x=SG, and if y = -1, then x = CO.

2. Linear non-separable SVM (Fig. 3b)

Then the restrictions (2) has to be modified with "slack"variables as follows:

$$\mathbf{w}^t\mathbf{x}_i + b \geq 1 - \xi_i, if\ \ y_i = 1$$

$$\mathbf{w}^t\mathbf{x}_i + b \leq -1 + \xi_i, if\ \ y_i = -1$$

Such training pairs that occur on "incorrect side" have $\xi_i \geq 1$ (see Fig. 3b).
The problem of hyperplane finding is to minimize general training error $\sum_i \xi_i$, while still maximizing the margin $\frac{\| \mathbf{w} \|^2}{2} + C\sum_i \xi_i$

where C –is a user selected scalar value whose chosen value controls the relative penalty for training errors.

Solution to such problem is equivalent to a minimization problem with respect to **w**, **b** and ξ_i of the following Lagrangian:

$$L(\mathbf{w}, \mathbf{b}, \xi, \mu, \alpha) = \frac{1}{2}\| \mathbf{w} \|^2 + C\sum_i \xi_i - \sum \alpha_i \{(\mathbf{w}\mathbf{x}_i + b)y_i - 1 + \xi_i\} - \sum_i \mu_i \xi_i \quad (5)$$

SVM also solves this problem.

Classification rule for non-separable SVM keeps the same as for separable SVM but the results of such classification can be less reliable .

2. Non-linear SVM (Fig. 3c)

The general idea is to map training exemplars $\mathbf{x}_j$ into a higher (possible infinite) dimensional Euclidean space H using nonlinear function $\Phi(\mathbf{x}_i)$ in which a linear SVM is the employed:

$$\Phi : R^d - H,$$

Then the main problem is to select the mapping function Φ(..).

Classification rule:

$y=\mathrm{sgn}(\mathbf{w}^t\Phi(\mathbf{z})+b)$ (6)

Remark 1. Transform of Lagrangian in this case requires an introducing so called "*kernel" function* k(x, x')=Φ(x)Φ(x').

Example.

-Polynomial kernels: $K(\mathbf{x},\mathbf{x}')=\left(\langle\gamma\,\mathbf{x},\mathbf{x}'\rangle+r\right)^d,\gamma>0$;

-Gaussian kernels: $K(\mathbf{x},\mathbf{x}')=\exp\left(-\gamma\|\mathbf{x}-\mathbf{x}'\|^2\right),\gamma>0$;

-Sigma-type kernels: $K(\mathbf{x},\mathbf{x}')=th\left(K_0+K_1\langle\mathbf{x},\mathbf{x}'\rangle\right)$;

where γ, *d*, K_0, K_1 are kernel's parameters.

Let us consider 2D -space $\mathbf{X}=\mathbf{R}^2$ and quadratic kernel

$$K(\mathbf{x},\mathbf{x}')=\langle\mathbf{x},\mathbf{x}'\rangle^2,\ \text{where}\ \mathbf{x}=(x_1,x_2),\mathbf{x}'=(x_1',x_2')$$

Transform squared scalar product:

$$K(\mathbf{x},\mathbf{x}') = \left(x_1 x_1^{'} + x_2 x_2^{'}\right)^2 = x_1^{\ 2} x_1^{'2} + x_2^{\ 2} x_2^{'2} + 2 \cdot x_1 x_1' x_2 x_2' = \quad \left\langle \left(x_1^{\ 2}, x_2^{\ 2}, \sqrt{2} x_1 x_2\right), \left(x_1^{'2}, x_2^{'2}, \sqrt{2} x_1^{'} x_2^{'}\right) \right\rangle$$

We can see that kernel $K(\mathbf{x},\mathbf{x}') = \langle \mathbf{x},\mathbf{x}' \rangle^2$ can be expressed as ordinary scalar product in $\mathbf{R}^3$, whereas space transform $\mathbf{Y}:\mathbf{R}^2 \rightarrow \mathbf{R}^3$, corresponding to this kernel is :

$$(x_1 x_2) \rightarrow \left(x_1^{\ 2}, x_2^{\ 2}, \sqrt{2} x_1 x_2\right)$$

It is worth to noting that planes in **H** be correspond to square surface in original space **X**. In particular such kernel is able to divide linearly inside and outside parts of oval that would be impossibly in original 2D space.

A library for SVM can be found over Internet:
Ching-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines, 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm
Remark 2. It is worth to note that SVM can be used not only for SG detecting but in solution to many other problems of recognition (stock-exchange games, person identification by their biometrics ctr).

Functionals F(.) used in SVM-based Classifier.

A choice of the set of functionals is based on analysis of embedding methods using in different SG . In [16] the final set of 23 functionals used (see Table below). But first of all this choice is, intended for steganalysis of such SG as F5, Outguess.

Functional/feature name	Functional F
Global histogram	$H / \| H \|_{L_1}$
Individual histograms for 5 DCT modes	$\frac{h^{21}}{\| h^{21} \|_{L_1}}, \frac{h^{31}}{\| h^{31} \|_{L_1}}, \frac{h^{12}}{\| h^{12} \|_{L_1}}, \frac{h^{22}}{\| h^{22} \|_{L_1}}, \frac{h^{13}}{\| h^{13} \|_{L_1}}$
Dual histograms for 11 DCT values (-5..5)	$\frac{g^{-5}}{\| g^{-5} \|_{L_1}}, \frac{g^{-4}}{\| g^{-4} \|_{L_1}}, \ldots, \frac{g^{4}}{\| g^{4} \|_{L_1}} \frac{g^{5}}{\| g^{5} \|_{L_1}}$
Variation	V
L1 and L2 blockiness	B_1, B_2
Co-occurence	N_{00}, N_{01}, N_{11}

Formulas for different functionals:

$$g_{ij}^{d} = \sum_{k=1}^{B} \delta(d, d_k(i,j))$$

$$V = \frac{\sum_{i,j=1}^{8} \sum_{k=1}^{|I_r|-1} |\, d_{I_r(k)}(i,j) - d_{I_r(k+1)}(i,j)\,| + \sum_{i,j=1}^{8} \sum_{k=1}^{|I_c|-1} |\, d_{I_c(k)}(i,j) - d_{I_c(k+1)}(i,j)\,|}{|\,I_r\,| + |\,I_c\,|}$$

$$B_{\alpha} = \frac{\sum_{i=1}^{\lfloor (M-1)/8 \rfloor} \sum_{j=1}^{N} \left| x_{8i,j} - x_{8i+1,j} \right|^{\alpha} + \sum_{j=1}^{\lfloor (N-1)/8 \rfloor} \sum_{i=1}^{M} \left| x_{i,8j} - x_{i,8j+1} \right|^{\alpha}}{N\lfloor (M-1)/8 \rfloor + M\lfloor (N-1)/8 \rfloor}$$

$$C_{st} = \frac{\sum_{k=1}^{|I_r|-1} \sum_{i,j=1}^{8} \delta\left(s, d_{I_r(k)}(i,j)\right)\delta\left(t, d_{I_r(k+1)}(i,j)\right) + \sum_{k=1}^{|I_c|-1} \sum_{i,j=1}^{8} \delta\left(s, d_{I_c(k)}(i,j)\right)\delta\left(t, d_{I_c(k+1)}(i,j)\right)}{|\,I_r\,| + |\,I_c\,|}$$

$$N_{00} = C_{0,0}(J_1) - C_{0,0}(J_2)$$
$$N_{01} = C_{0,1}(J_1) - C_{0,1}(J_2) + C_{1,0}(J_1) - C_{1,0}(J_2) + C_{-1,0}(J_1) - C_{-1,0}(J_2) + C_{0,-1}(J_1) - C_{0,-1}(J_2)$$
$$N_{11} = C_{1,1}(J_1) - C_{1,1}(J_2) + C_{1,-1}(J_1) - C_{1,-1}(J_2) + C_{-1,1}(J_1) - C_{-1,1}(J_2) + C_{-1,-1}(J_1) - C_{-1,-1}(J_2)$$

i, j=1,..,8, k=1,..,B

B – the total number of 8x8 blocks in the image,

$d_k(i,j)$ – quantized DCT coefficients,

Q(i,j) – quantization matrix,

H_r – global histogram,

$r=L,..,R$, $L=min_{k,i,j}(d_k(i,j))$, $R=max_{k,i,j}(d_k(i,j))$,

h_r^{ij} – individual histogram (for fixed *i* and *j*),

g_{ij}^{d} – the number of times the value *d* occurs as *(i,j)*-th DCT coefficient over all *B* blocks in the JPEG image,

$\delta(u,v)=1$, if $u=v$, else- *0,*

I_r, I_c – the vectors of block indices while scanning the image by rows and by columns , respectively,

B_{α} – blockiness measure,

M, N – image dimensions,

x_{ij} – grayscale values of the decompressed JPEG,

C – co-occurrence set.

SPAM features functionals [58]

SPAM features model transition probabilities between neighboring pixels along 8 directions: ↑, ↗, →, ↘, ↓, ↙, ←, ↖.

Let $I \in X$ be an image of size n_1 x n_2 pixels.

Then the formula to compute a functionals, say for one direction "→" is

$$C^{I\rightarrow}_{d_1,d_2} = \Pr\left\{D^{\rightarrow}_{ij} = d_1,\ D^{\rightarrow}_{i,j+1} = d_2\right\},$$

where $d_1, d_2 \in [-T, T]$ and $\forall (i,j) \in \{1,\ldots,n_1\} \times \{1,\ldots,n_2 - 1\}$, $D^{\rightarrow}_{ij} = I_{ij} - I_{i,j+1}$.

The total number of functionals is $8(2T+1)^2$ for the first order features.

So, for T = 3, every image can be described by 392 functionals.

Test performance of SG .

Remember the criteria of SG efficiency:

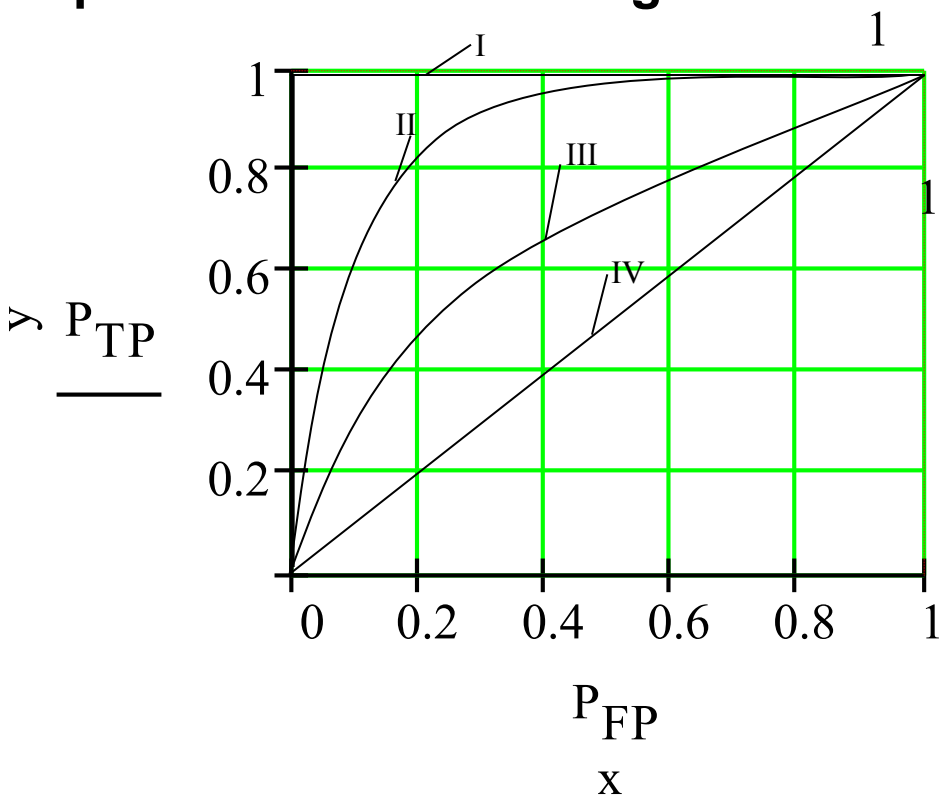
P_m – the probability of SG missing, otherwise- "*false negative"* probability P_{FN}
P_{fa} – the probability of false alarm, otherwise- "*false positive*" probability P_{FP}

Criteria which are equivalently to previous ones:
1- $P_m = P_{TP}$ (*true positive probability*)
1- $P_{fa} = P_{TN}$ (*true negative probability)*

The question arises : How can be expressed the connection between P_m and P_{fa}?

Definition. *ROC-curve*(*The Receiver Operating Characteristic curve*) is a plotting the fraction of true positive rate (the TP probability) versus the fraction of false positive rate (the FP probability), where Ptp is plotted on x-coordinate while Pfp is plotted on y-coordinate, .

Example of ROC-curve design



I – ideal classifier

II – good classifier

III – mediocre classifier

IV – classifier that is equal to random guessing

ROC-curve "quality" can be evaluated by parameter ρ (*detection reliability*):

ρ =2A-1, where A is the area under the ROC-curve.

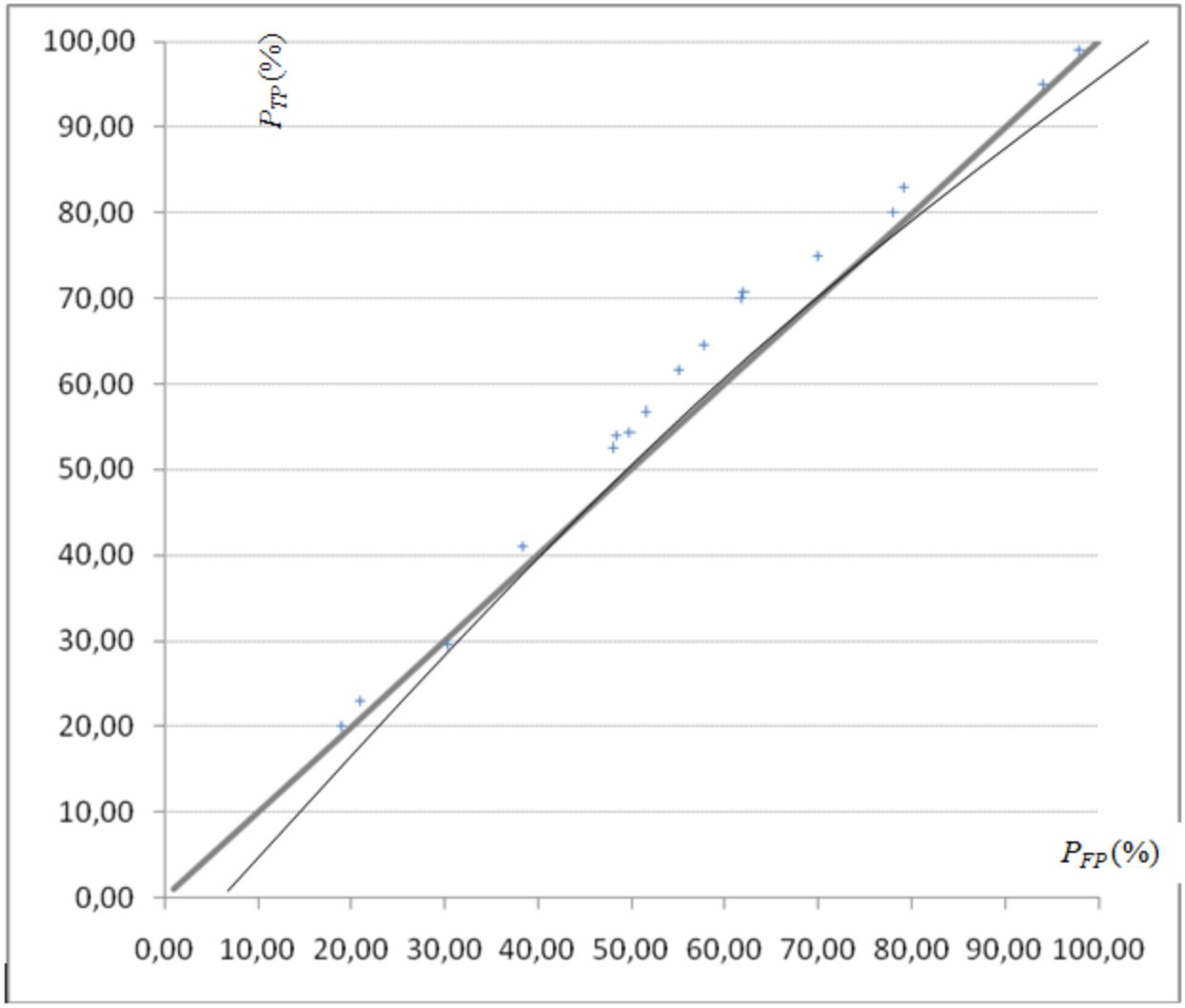
It is easy to prove that $\rho_I = 1$, $\rho_{IV} = 0$, $\rho_I > \rho_{II} > \rho_{III} > \rho_{IV}$.

Experiential testing of SVM efficiency. (Multiple detection)

Functionals	CO				F5			
	CO	F5	outguess	MB	CO	F5	outguess	MB
Global histogram	93.18	3.41	3.41	0.00	0.00	98.21	1.79	0.00
Ind.hist. for (1,2)	91.81	1.28	6.91	0.00	0.00	40.33	59.67	0.00
Ind.hist. for (2,1)	99.36	0.51	0.13	0.00	0.00	100.00	0.00	0.00
Ind.hist. for (1,3)	95.31	0.26	4.43	0.00	0.00	45.24	54.76	0.00
Ind.hist. for (3,1)	96.59	2.81	0.60	0.00	0.00	100.00	0.00	0.00
Ind.hist. for (2,2)	97.53	0.13	2.35	0.00	0.00	83.87	16.13	0.00
Dual hist. for -5	100.00	0.00	0.00	0.00	0.00	99.96	0.04	0.00
Dual hist. for -4	100.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for -3	100.00	0.00	0.00	0.00	0.00	47.76	52.24	0.00
Dual hist. for -2	100.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for -1	99.70	0.30	0.00	0.00	0.09	90.87	9.05	0.00
Dual hist. for 0	76.33	23.67	0.00	0.00	16.05	83.95	0.00	0.00
Dual hist. for 1	100.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for 2	100.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for 3	100.00	0.00	0.00	0.00	0.00	45.45	54.55	0.00
Dual hist. for 4	100.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for 5	100.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00

Functionals	outguess				F5 (1 byte)			
	CO	F5	outguess	MB	CO	F5	outguess	MB
Global histogram	0.00	0.00	100.00	0.00	0.00	99.16	0.84	0.00
Ind.hist. for (1,2)	0.00	5.38	94.62	0.00	0.00	43.85	56.15	0.00
Ind.hist. for (2,1)	0.00	0.00	100.00	0.00	0.00	100.00	0.00	0.00
Ind.hist. for (1,3)	0.00	22.67	77.33	0.00	0.00	48.70	51.30	0.00
Ind.hist. for (3,1)	0.00	0.00	100.00	0.00	0.00	100.00	0.00	0.00
Ind.hist. for (2,2)	0.18	0.00	99.82	0.00	0.00	86.21	13.79	0.00
Dual hist. for -5	0.00	0.00	100.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for -4	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for -3	0.00	22.34	77.66	0.00	0.00	47.11	52.89	0.00
Dual hist. for -2	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for -1	0.09	18.39	81.52	0.00	0.04	91.45	8.51	0.00
Dual hist. for 0	7.82	92.18	0.00	0.00	16.35	83.65	0.00	0.00
Dual hist. for 1	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for 2	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for 3	0.00	21.10	78.90	0.00	0.00	46.80	53.20	0.00
Dual hist. for 3	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
Dual hist. for 5	0.00	0.00	100.00	0.00	0.00	100.00	0.00	0.00

Functionals	Outguess (1 byte)				MB			
	CO	F5	outguess	MB	CO	F5	outguess	MB
Global histogram	0.00	0.00	100.00	0.00	100	0.00	0.00	0.00
Ind.hist. for (1,2)	0.00	6.47	93.53	0.00	0.00	100	0.00	0.00
Ind.hist. for (2,1)	0.00	0.09	99.91	0.00	100	0.00	0.00	0.00
Ind.hist. for (1,3)	0.00	21.83	78.17	0.00	0.00	100	0.00	0.00
Ind.hist. for (3,1)	0.00	0.13	99.87	0.00	100	0.00	0.00	0.00
Ind.hist. for (2,2)	0.00	0.31	99.69	0.00	0.00	100	0.00	0.00
Dual hist. for -5	0.00	0.00	100.00	0.00	0.00	0.00	100	0.00
Dual hist. for -4	0.00	100.00	0.00	0.00	0.00	100	0.00	0.00
Dual hist. for -3	0.00	23.69	76.31	0.00	0.00	0.00	100	0.00
Dual hist. for -2	0.00	100.00	0.00	0.00	0.00	100	0.00	0.00
Dual hist. for -1	0.31	20.55	79.14	0.00	100	0.00	0.00	0.00
Dual hist. for 0	8.81	91.19	0.00	0.00	0.00	100	0.00	0.00
Dual hist. for 1	0.00	100.00	0.00	0.00	0.00	0.00	100	0.00
Dual hist. for 2	0.00	100.00	0.00	0.00	0.00	100	0.00	0.00
Dual hist. for 3	0.00	23.16	76.84	0.00	0.00	0.00	100	0.00
Dual hist. for 3	0.00	100.00	0.00	0.00	0.00	100	0.00	0.00
Dual hist. for 5	0.00	0.00	100.00	0.00	0.00	0.00	100	0.00



ROC-curve based on detection of SG-PQS by SVM