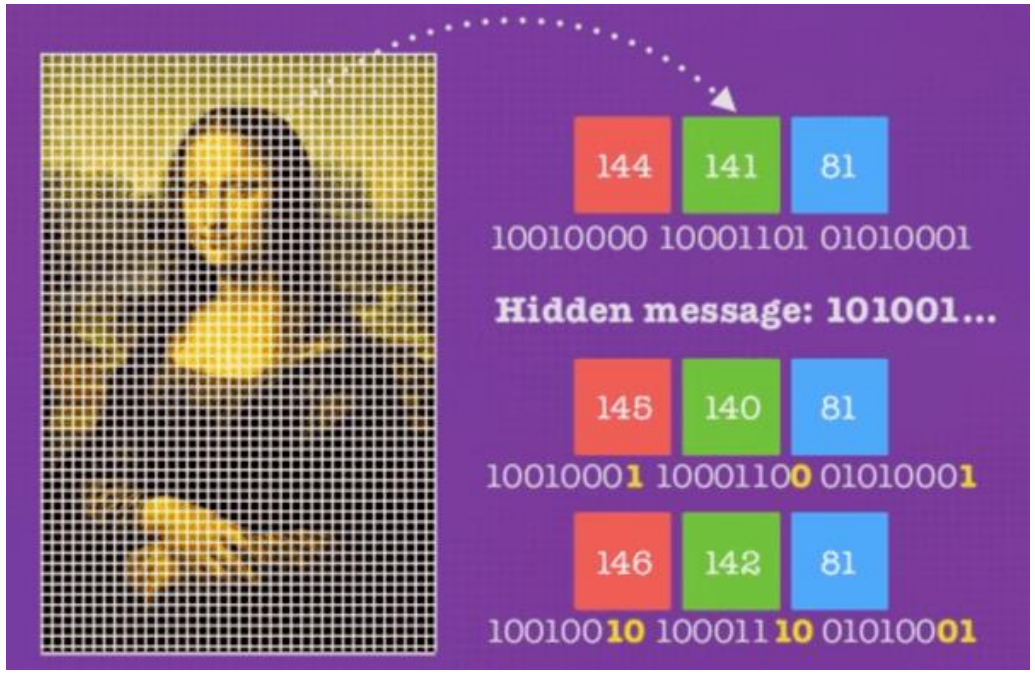
Jessica Fridrich

# Rich Models for Steganalysis of Digital Images

Woochul Shin

# Steganography vs Steganalysis




Cover Image ⟷ Stego Image

X Y

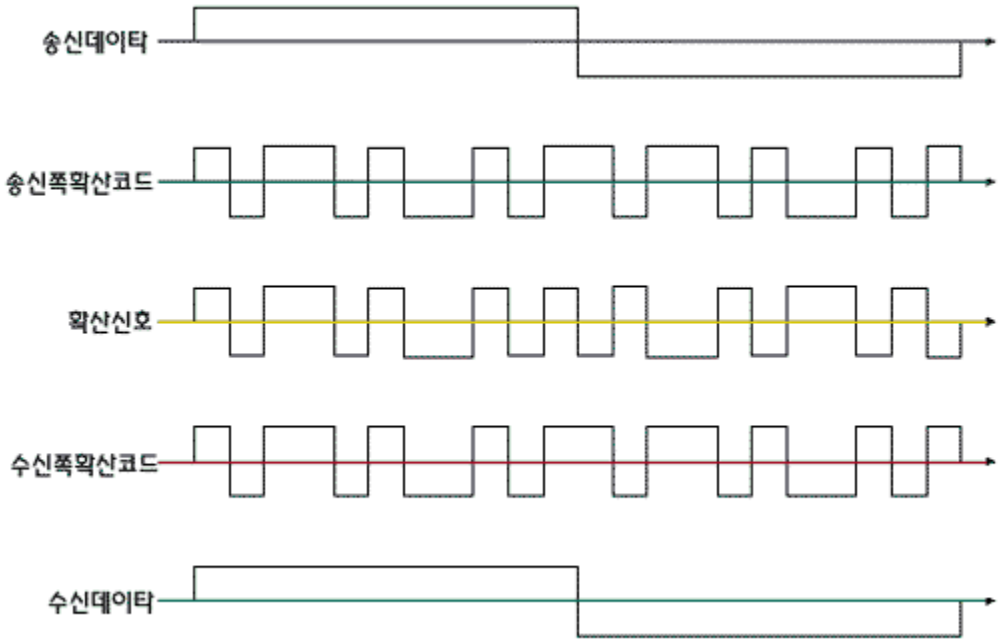
Map: $D : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty]$

Minimizing embedding impact:

$$D(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{n} \rho_i |x_i - y_i|.$$

($\rho_i$ : costs of pixel change)

# Steganography vs Steganalysis




<그림2-1>확산코드에 의한 통신 예

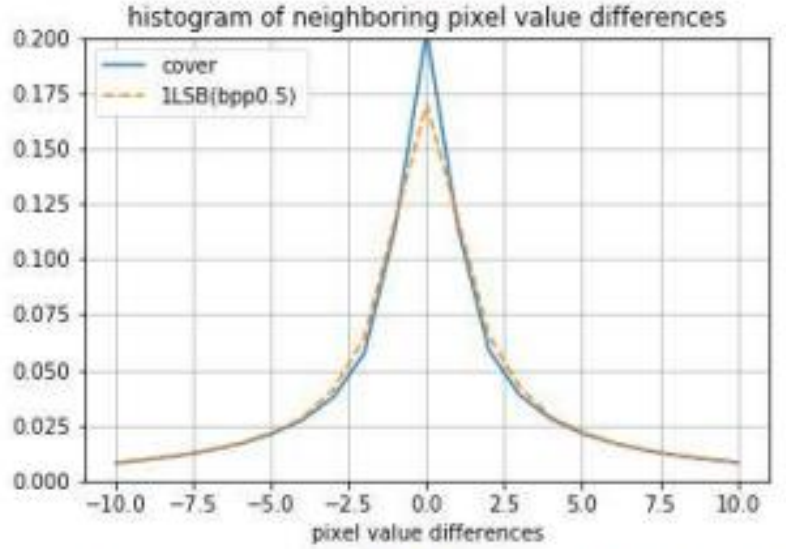



그림 1. BOSSbase 데이터 셋의 1000장의 커버 영상과 1-LSB bpp 0.5 삽입 스테고 영상의 이웃 픽셀 사이의 값 차이 분포[15]

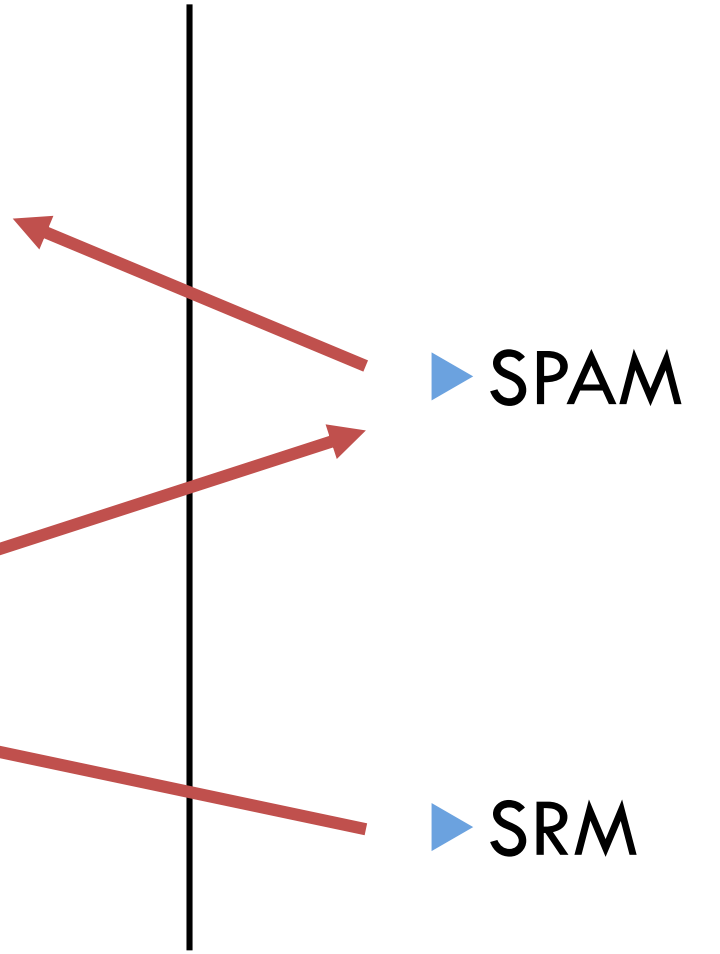
Fig. 1. Distribution of neighboring pixels differences between cover images and 1-LSB bpp 0.5 stego images in 1000 BOSSbase data-sets[15]

# Steganography vs Steganalysis

- LSB replacement
- LSB matching

  (±1 embedding)

- HUGO



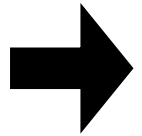
- SPAM
- SRM

# LSB Replacement

Hidden Message : “abc”  97 98 99

| | |
|---|---|
| 97 | 1100001 |
| 98 | 1100010 |
| 99 | 1100011 |


| | |
|---|---|
| 34 | .....0 |
| 157 | .....1 |
| 215 | .....1 |
| 136 | .....0 |



| | |
|---|---|
| 35 | .....1 |
| 157 | .....1 |
| 214 | .....0 |
| 136 | .....0 |

Odd: 0 or -1
Even: 0 or +1

∴ **asymmetry**

# LSB Matching(±1 embedding)

Hidden Message : “abc”  97 98 99

| | |
|---|---|
| 97 | 1100001 |
| 98 | 1100010 |
| 99 | 1100011 |


| | | | |
|---|---|---|---|
| 34 | .....0 | 1 | X |
| 157 | .....1 | 1 | O |
| 215 | .....1 | 0 | X |
| 136 | .....0 | 0 | O |

vs

Equal

Equal X: ±1 randomly
Equal O: do nothing

# SPAM(Subtractive Pixel Adjacency Matrix) (1)

$$\Pr(I_{i_1,j_1}, I_{i_2,j_2}) \Rightarrow \Pr(I_{i,j+1} - I_{i,j})$$

$$\begin{aligned} &I(I_{i,j+1} - I_{i,j}, I_{i,j}) \\ &\quad = H(I_{i,j+1} - I_{i,j}) - H(I_{i,j+1} - I_{i,j} | I_{i,j}) \\ &\quad = H(I_{i,j+1} - I_{i,j}) - H(I_{i,j+1} \; I_{i,j}) \end{aligned}$$

∴ Dependence between pixel difference and pixel value is small

$\mathbf{D}_{i,j}^{\rightarrow} = I_{i,j} - I_{i,j+1} \in [-T, T]$ Pixel difference array

$\mathbf{M}_{u,v}^{\rightarrow} = \Pr\left(\mathbf{D}_{i,j+1}^{\rightarrow} = u \, \mathbf{D}_{i,j}^{\rightarrow} = v\right)$ 1st order Markov process

$\mathbf{M}_{u,v,w}^{\rightarrow} = \Pr\left(\mathbf{D}_{i,j+2}^{\rightarrow} = u | \mathbf{D}_{i,j+1}^{\rightarrow} = v, \mathbf{D}_{i,j}^{\rightarrow} = w\right)$ 2nd order Markov process

$$\mathbf{F}_{1,\ldots,k} = \frac{1}{4}\left[\mathbf{M}^{\rightarrow} + \mathbf{M}^{\leftarrow} + \mathbf{M}^{\downarrow} + \mathbf{M}^{\uparrow}\right]$$
$$\mathbf{F}_{k+1,\ldots,2k} = \frac{1}{4}\left[\mathbf{M}^{\searrow} + \mathbf{M}^{\nwarrow} + \mathbf{M}^{\swarrow} + \mathbf{M}^{\nearrow}\right]$$

Features

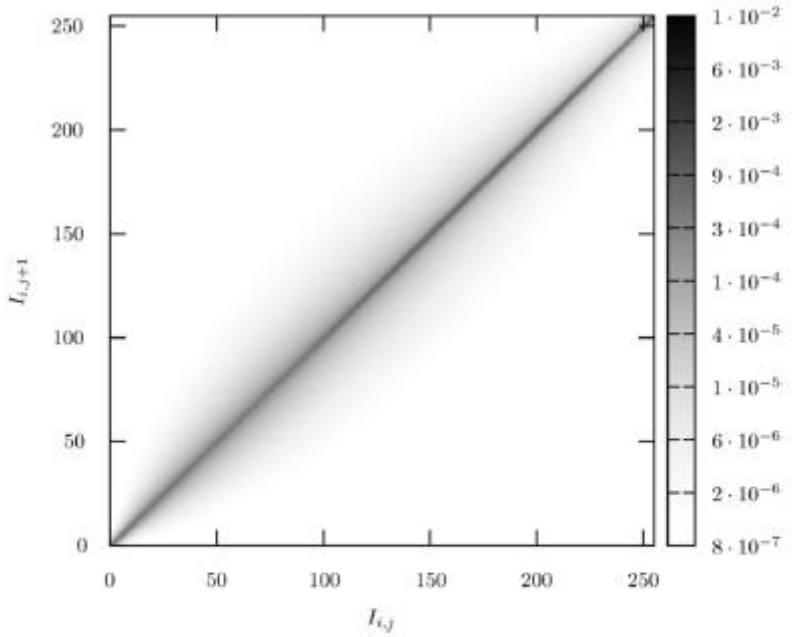
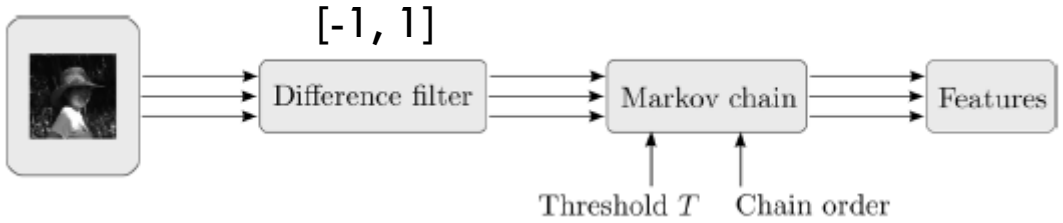



Fig. 1. Distribution of two horizontally adjacent pixels $(I_{i,j}, I_{i,j+1})$ in 8-bit grayscale images estimated from approximately 10 700 images from the BOWS2 database (see Section IV for more details about the database). The degree of gray at $(x, y)$ is the probability $\Pr(I_{i,j} = x \wedge I_{i,j+1} = y)$ at the logarithmic scale.

k = 2(2T+1)^2 features for 1st order Markov process, k = 2(2T+1)^3 features for 2nd order Markov process

# SPAM(Subtractive Pixel Adjacency Matrix) (2)

1. Extraction of features




2. SVM(Support Vector Machine) with rbf(radial basis function) kernel

$$k(x, y) = \exp\left(-\gamma \|x - y\|_2^2\right), \gamma > 0$$

Hand-crafted parameters :

| | |
|---|---|
| T(threshold), order of Markov process | : Extraction of features |
| C, gamma | : SVM |

# HUGO(Highly Undetectable steGO)

security through high-dimensions

# SRM(Spatial Rich Model) (1)

1. Extraction of features

Various types of relationships
among neighboring samples of noise residuals
obtained by linear and nonlinear filters
with compact supports

*Multiple submodels* where each submodel captures
different embedding artifacts

2. Random Forest + Ensemble Classifier(Bootstrap aggregation = bagging)

# SRM(Spatial Rich Model) (2)

## A. Submodels

1) Compute Residuals

$$R_{ij} = \hat{X}_{ij}(\mathcal{N}_{ij}) - cX_{ij}$$

Q. Why use residuals instead of pixel values?

A. The advantage of modeling the residuals is that the image content is largely suppressed in Residual matrix, which has a much narrower dynamic range allowing thus a more compact and robust statistical description.

2) Truncation and Quantization

$$R_{ij} \leftarrow \text{trunc}_T\left(\text{round}\left(\frac{R_{ij}}{q}\right)\right)$$

Truncation: To curb the residual's dynamic range to allow their description using co-occurrences with small T.

Quantization: To make residual more sensitive to embedding changes at spatial discontinuities in the image such as edges or textures.

# SRM(Spatial Rich Model) (3)

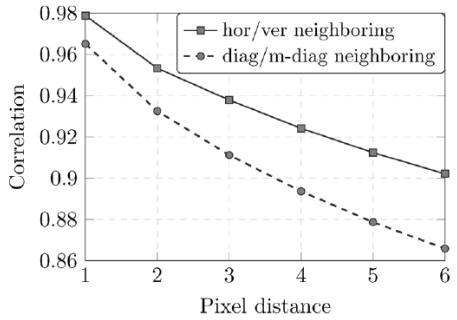



Fig. 1. Correlation between pixels based on their distance. Distance of diagonally neighboring pixels is in the multiples of the diagonal of two neighboring pixels. Results were averaged over 100 randomly selected images from BOSSbase ver. 0.92.

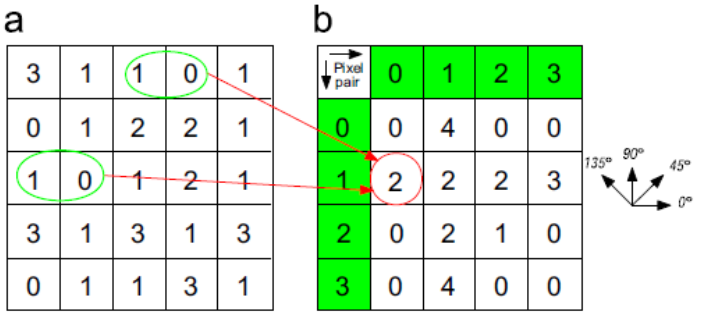



**Fig. 1.** Gray level co-occurrence matrix calculation example. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

3) Co-occurrences

Our submodels will be constructed from only horizontal and vertical co-occurrences of four consecutives residual samples processed using T=2

Thus, co-occurrence matrix has 4 dimensions with (2T+1)^4 = 625 elements

$$\mathbf{C}_{\mathbf{d}}^{(\mathrm{h})} = \frac{1}{Z}\Big|\{(R_{ij}, R_{i,j+1}, R_{i,j+2}, R_{i,j+3}) | R_{i,j+k-1} = d_k, k = 1, \ldots, 4\}\Big| \quad (3)$$

where $Z$ is the normalization factor ensuring that $\sum_{\mathbf{d}\in\mathcal{T}_4} \mathbf{C}_{\mathbf{d}}^{(\mathrm{h})} = 1$. The vertical co-occurrence $\mathbf{C}^{(\mathrm{v})}$ is defined analogically.

# SRM(Spatial Rich Model) (4)

## B. Residuals $R_{ij} = \hat{X}_{ij}(\mathcal{N}_{ij}) - cX_{ij}$

### 1) Residual Classes

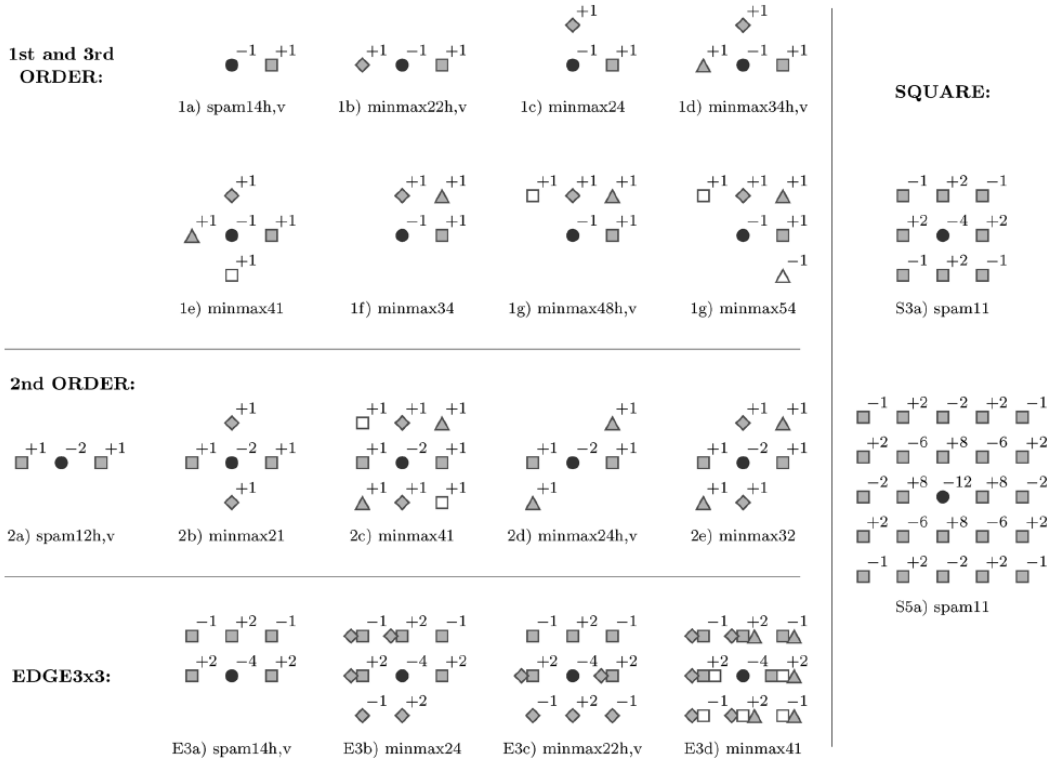



Fig. 2. Definitions of all residuals. Residuals 3a – 3h are defined similar to the first-order residuals, while E5a – E5d are similar to E3a – E3d, defined using the corresponding part of the $5 \times 5$ kernel displayed in S5a. See text for more details.

### 2) Residual Symmetries

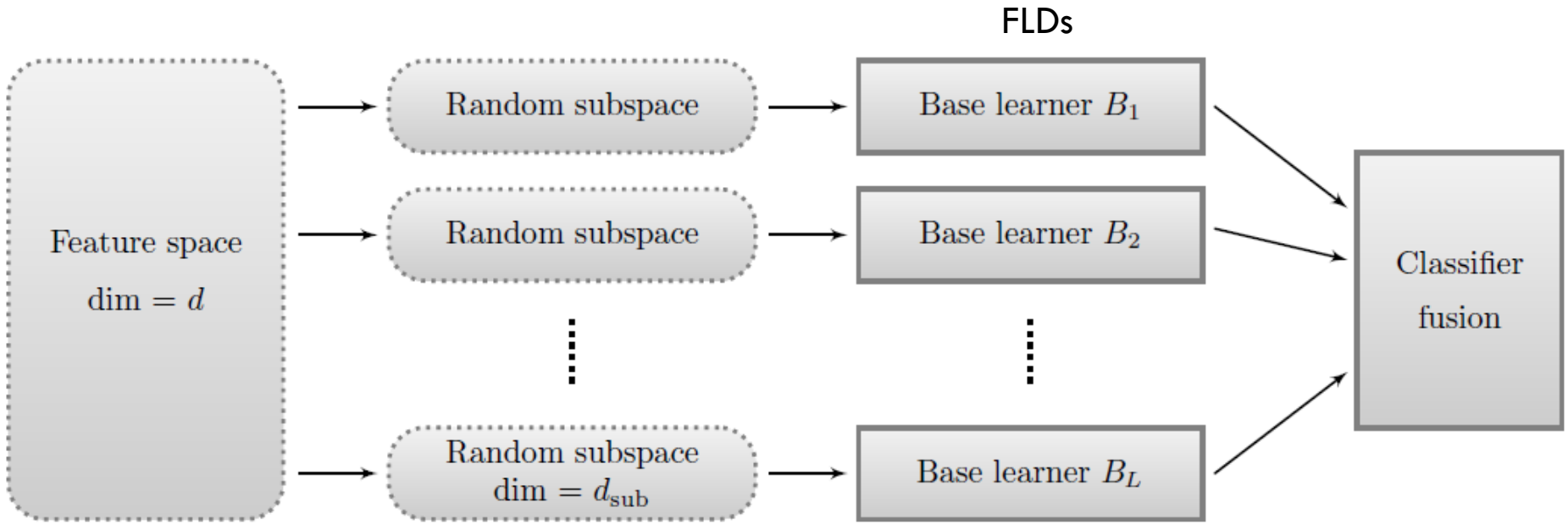
horizontal and vertical co-occurrence matrices can be added to a single matrix

### 3) Co-occurrence Symmetrization

(1) sign-symmetry (2) directional symmetry

⇨ Reduce number of submodels

# SRM(Spatial Rich Model) (5)

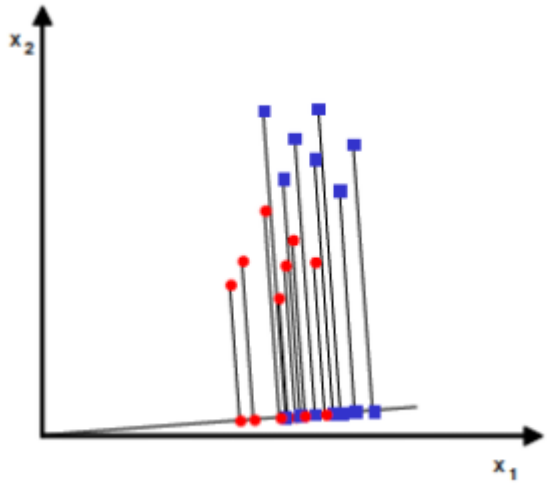


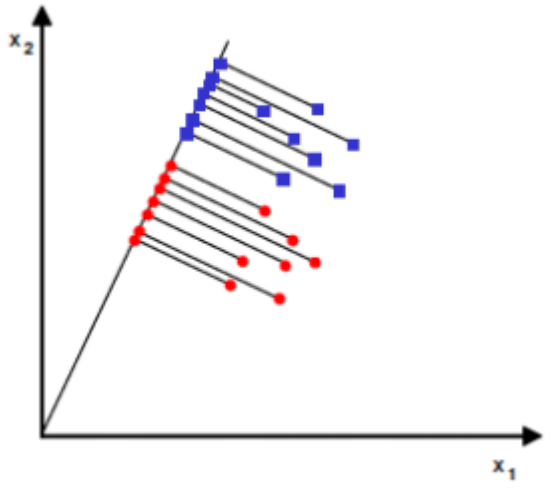

$$E_{\mathrm{OOB}}^{(L)} = \frac{1}{2N^{\mathrm{trn}}} \sum_{m=1}^{N^{\mathrm{trn}}} \left( B^{(L)}(\mathbf{x}^{(m)}) + 1 - B^{(L)}(\bar{\mathbf{x}}^{(m)}) \right)$$

➡ Out-Of-Bag(OOB) error: determines the number of base learners L

Submodel selection strategy: ALL, BEST-q, BEST-q-CLASS, Q1, CLASS-q, ITERATIVE-BEST-q

# FLD(Fisher Linear Discriminants analysis)




$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

# Ensemble + Bagging(Bootstrap Aggregation)

i.i.d

$$Var(X_i) = \sigma^2 \qquad Var(\bar{X}) = Var\left(\frac{1}{n}\sum_i X_i\right) = \frac{\sigma^2}{n}$$

Drop independence assumption i.e Xs are correlated by p

$$Var(\bar{X}) = p\sigma^2 + \frac{1-p}{n}\sigma^2$$

## Ways to ensemble

1) different algorithms
2) different training sets
3) Bagging (Random Forests)
4) Boosting (Adaboost, xgboost)

## Bagging - Bootstrap Aggregation

Have a true population P
Training set $S \sim P$
Assume $P = S$
Bootstrap samples $Z \sim S$

## Bootstrap Samples $Z_1, \ldots, Z_M$

Train model $G_m$ on $Z_m$

$$G(m) = \frac{\sum_{m=1}^{M} G_m(x)}{M}$$

## Random Forests

At each split, consider only a fraction of your total features

Decrease p

Decorrelate Models