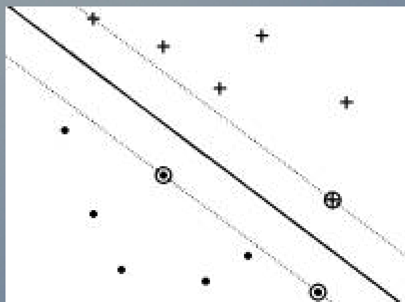
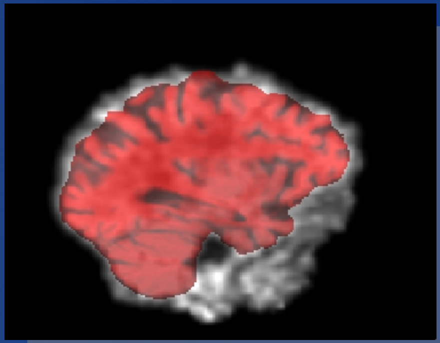
SVM classifiers and PET/MRI coregistration.





Magic-5 Workshop on Medical Imaging applications and tools.

2-4 December 2009.

CESARE OVANDO.

CINVESTAV Mexico city / INFN Genova.

SVM.

Support Vector Machines is a method for constructing a special kind of rule, called linear classifier, in a way that produces classifiers with theoretical guarantees of good predictive performance.

While for producing non-linear rules it's used the “kernel trick” to construct special kind of non-linear rules.

SVM.

The Support vector machines combines three ideas:

- The solution technique from optimal hyperplanes.
- The idea of convolution of the dot-product.
- The notion of soft margins.

Finding the optimal Hyperplane (separable case).

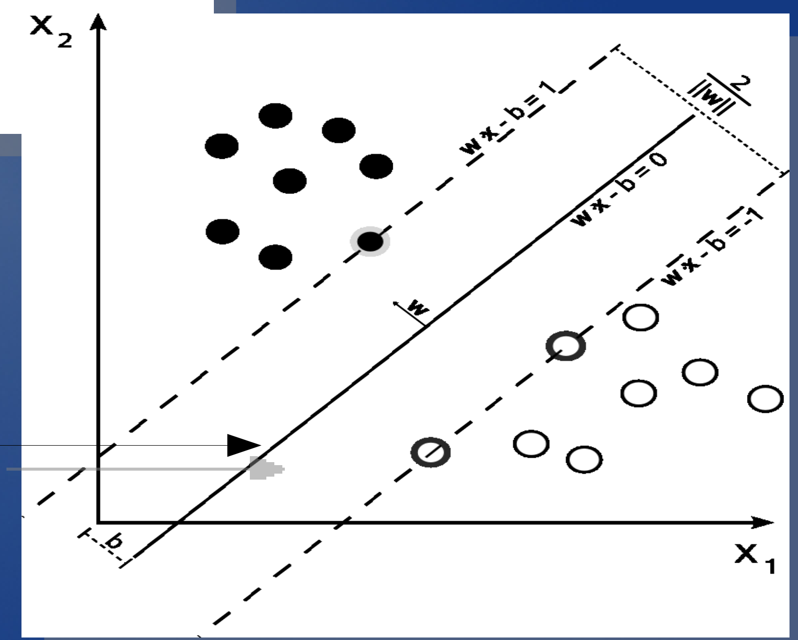
$$(y_1, \mathbf{x}_1), ..., (y_l, \mathbf{x}_l), \quad y_i \epsilon \{-1, +1\}, \quad , i = 1, ..., l.$$

$$i_F = sign[d(\mathbf{x})].$$

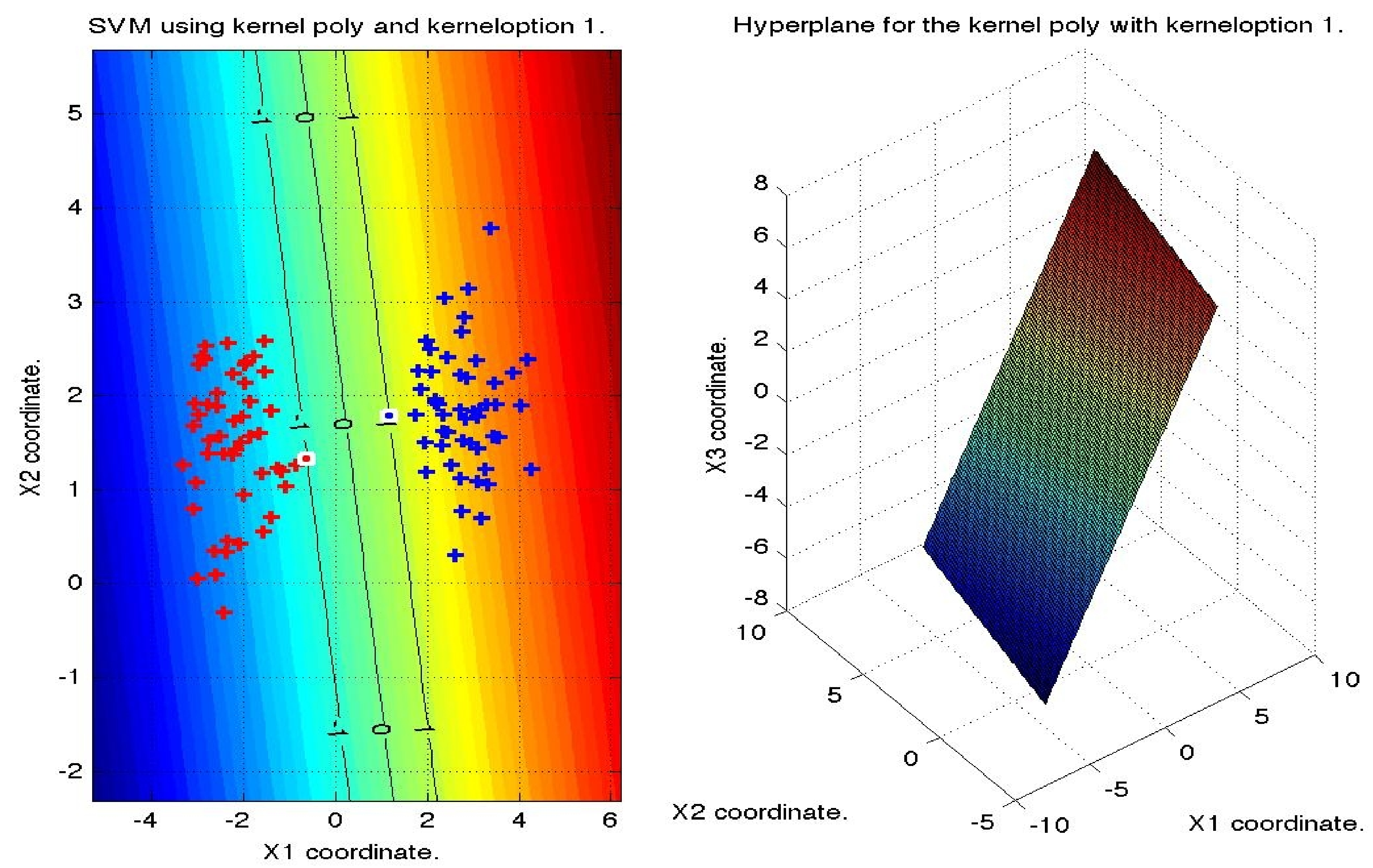
$$d(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\mathbf{x} + b.$$

$$y_i(\mathbf{w}^{\mathbf{T}}\mathbf{x} + b) \geq 1.$$

$$d(\mathbf{x}) = \mathbf{w}\mathbf{x}^{\mathbf{T}} + b = \sum_{i=1}^{l} y_i \alpha_i \mathbf{x}_i^{\mathbf{T}}\mathbf{x} + b.$$



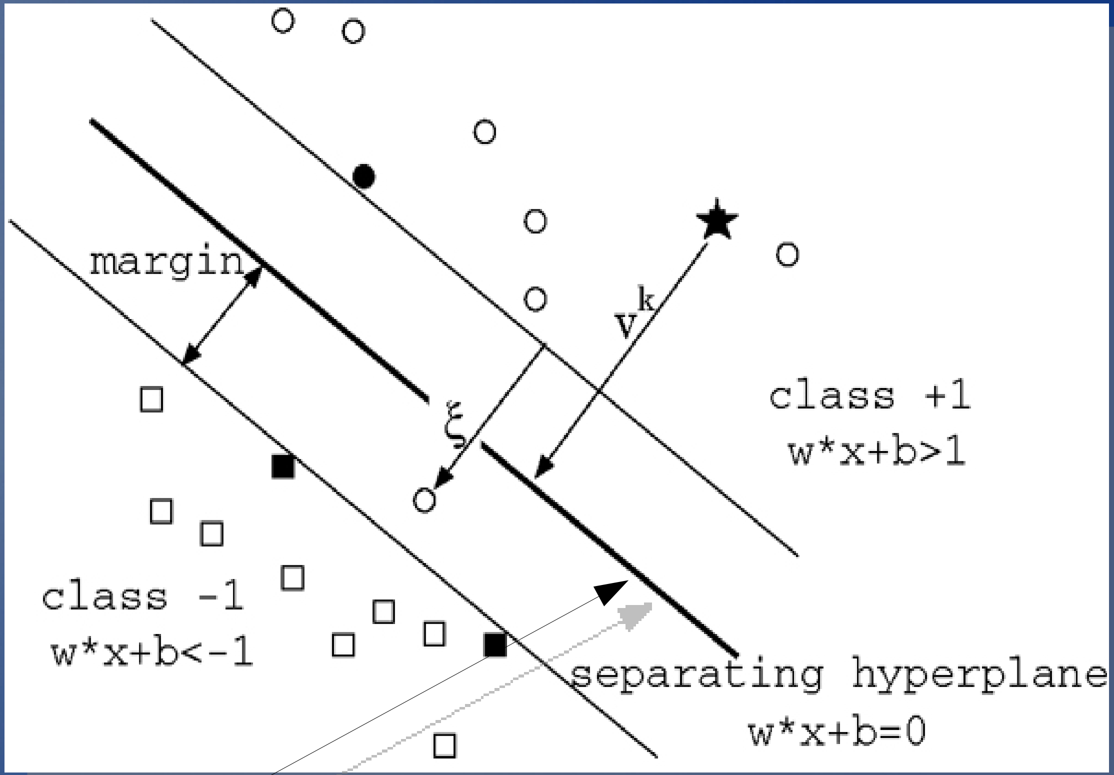
Finding the optimal Hyperplane (separable case).



Finding the optimal Hyperplane (non separable case).

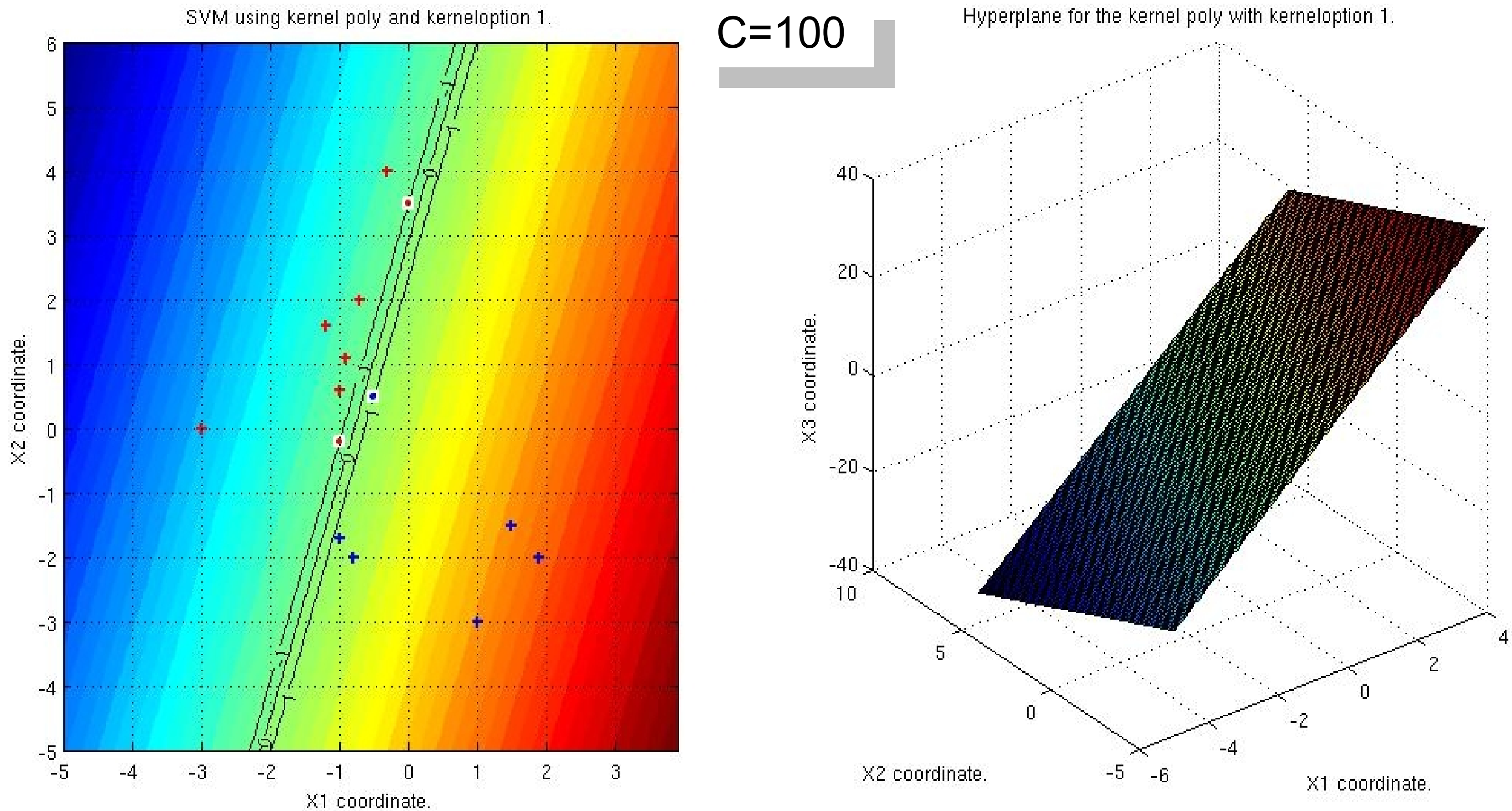
$$y_i(\mathbf{w}^T\mathbf{x} + b) \geq 1 - \xi_i.$$

$$\xi_i \geq 0.$$

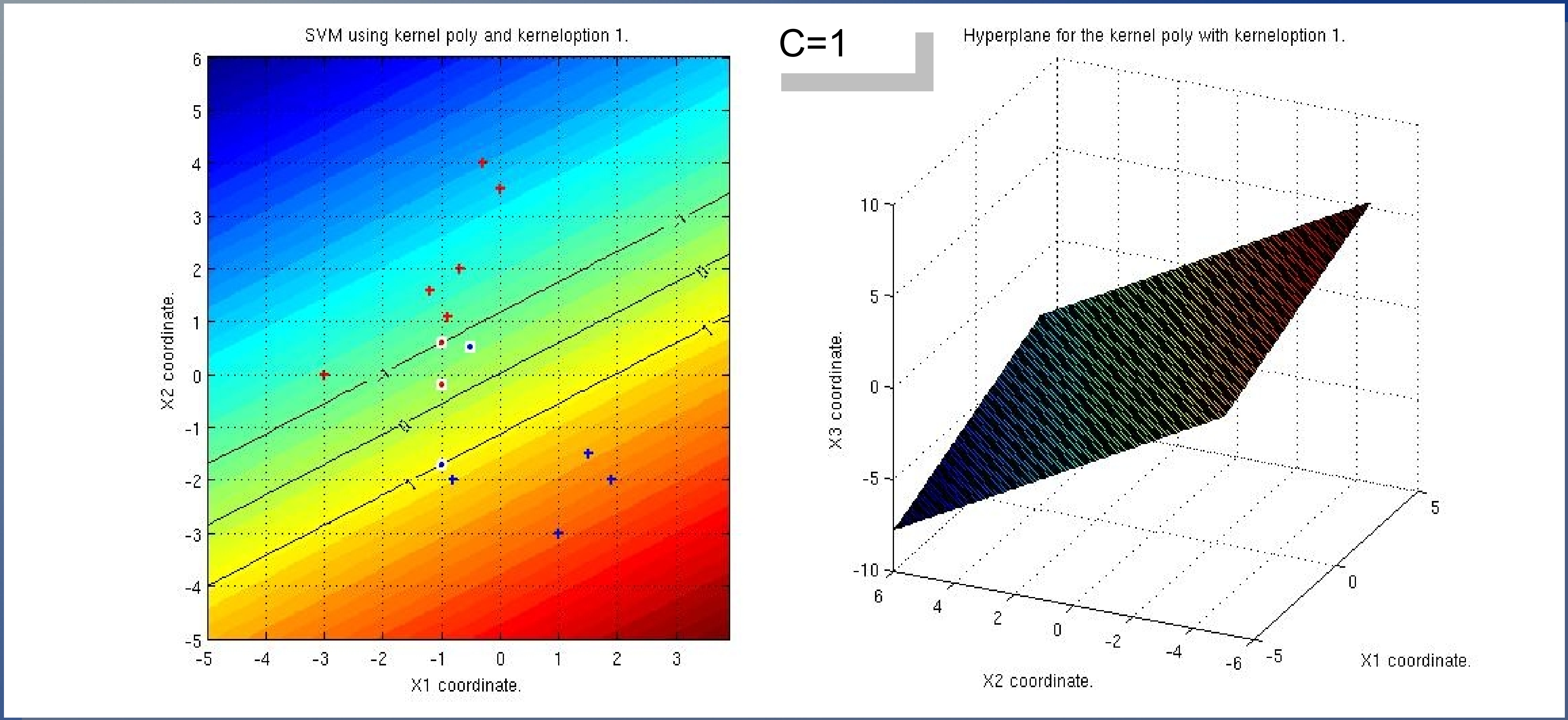


$$d(\mathbf{x}) = \mathbf{w}\mathbf{x}^T + b = \sum_{i=1}^{l} y_i \alpha_i \mathbf{x}_i^T \mathbf{x} + b.$$

Finding the optimal Hyperplane (non separable case).



Finding the optimal Hyperplane (non separable case).



Kernels: from linear to nonlinear classifiers.

Mercer's Theorem.

Any symmetric positive semi-definite function $K(\mathbf{x}, \mathbf{x}')$ is an inner product in some space (and viceversa).

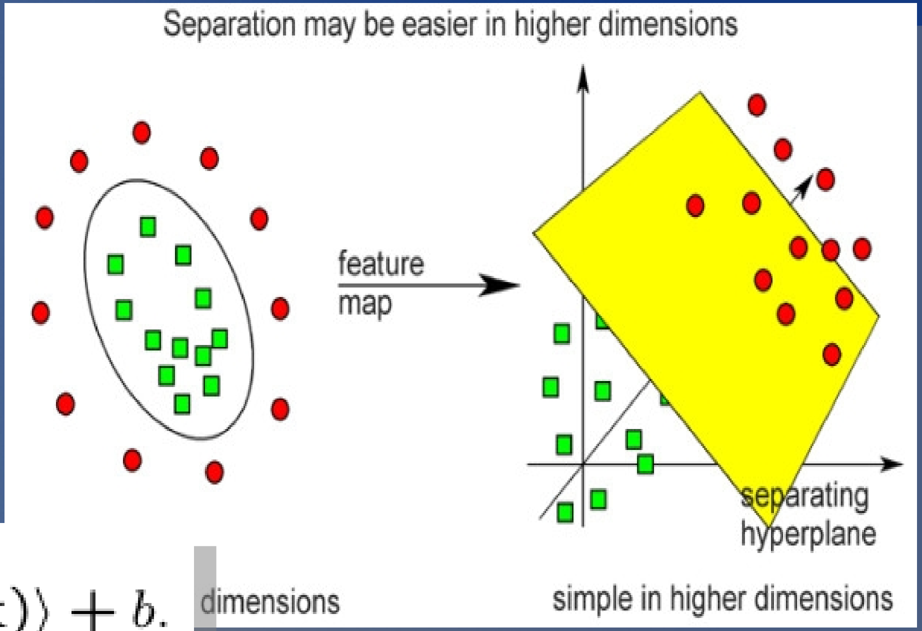
We don't need to know the actual mapping since qe can use kernel function to compute similarity in the feature space.

$$\mathbf{K}(\mathbf{x}, \mathbf{x_i}) = \mathbf{z}^\mathbf{T}\mathbf{z} = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x_i}) \rangle .$$

Kernels: from linear to nonlinear classifiers.

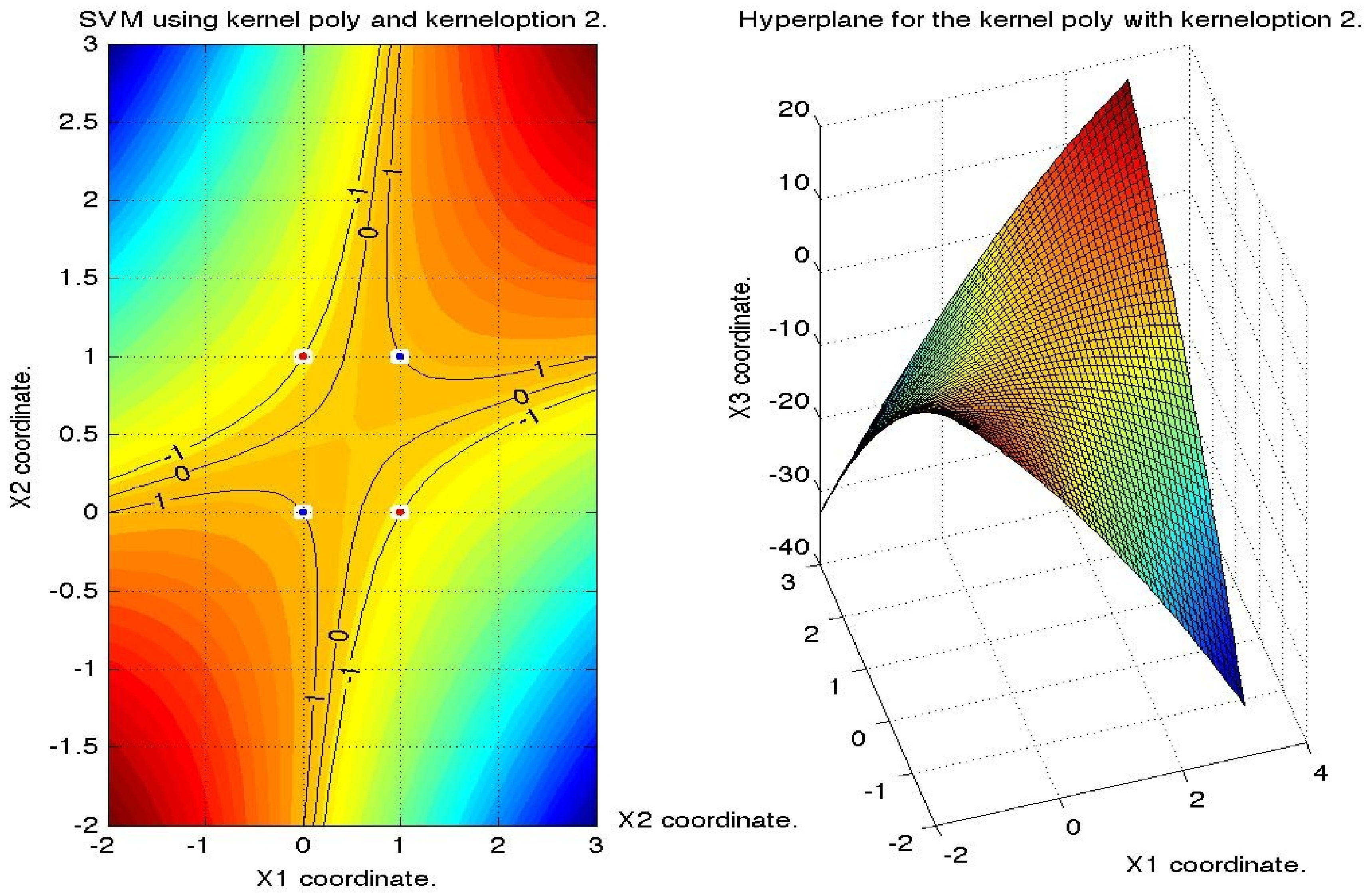
$$f(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\phi(\mathbf{x}) + b = \mathbf{w}^{\mathbf{T}}\mathbf{z} + b.$$

$$\mathbf{K}(\mathbf{x}, \mathbf{x_i}) = \mathbf{z}^{\mathbf{T}}\mathbf{z} = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x_i}) \rangle .$$



$$f(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\mathbf{z} + b = \sum_{i=1}^{l} \alpha_i y_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle + b.$$

$$f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i y_i \mathbf{K}(\mathbf{x_i}, \mathbf{x}) + b.$$


SVM using kernel poly and kerneloption 2.
X2 coordinate.
X1 coordinate.
Hyperplane for the kernel poly with kerneloption 2.
X3 coordinate.
X2 coordinate.
X1 coordinate.

SVM.

Once the support vectors have been found, the bound on the expected probability of commit-ing an error on a test example can be calcu-lated as follows:

$$E_l\left[p(error)\right] \leq \frac{E(number\, of\, support\, vectors)}{l}.$$

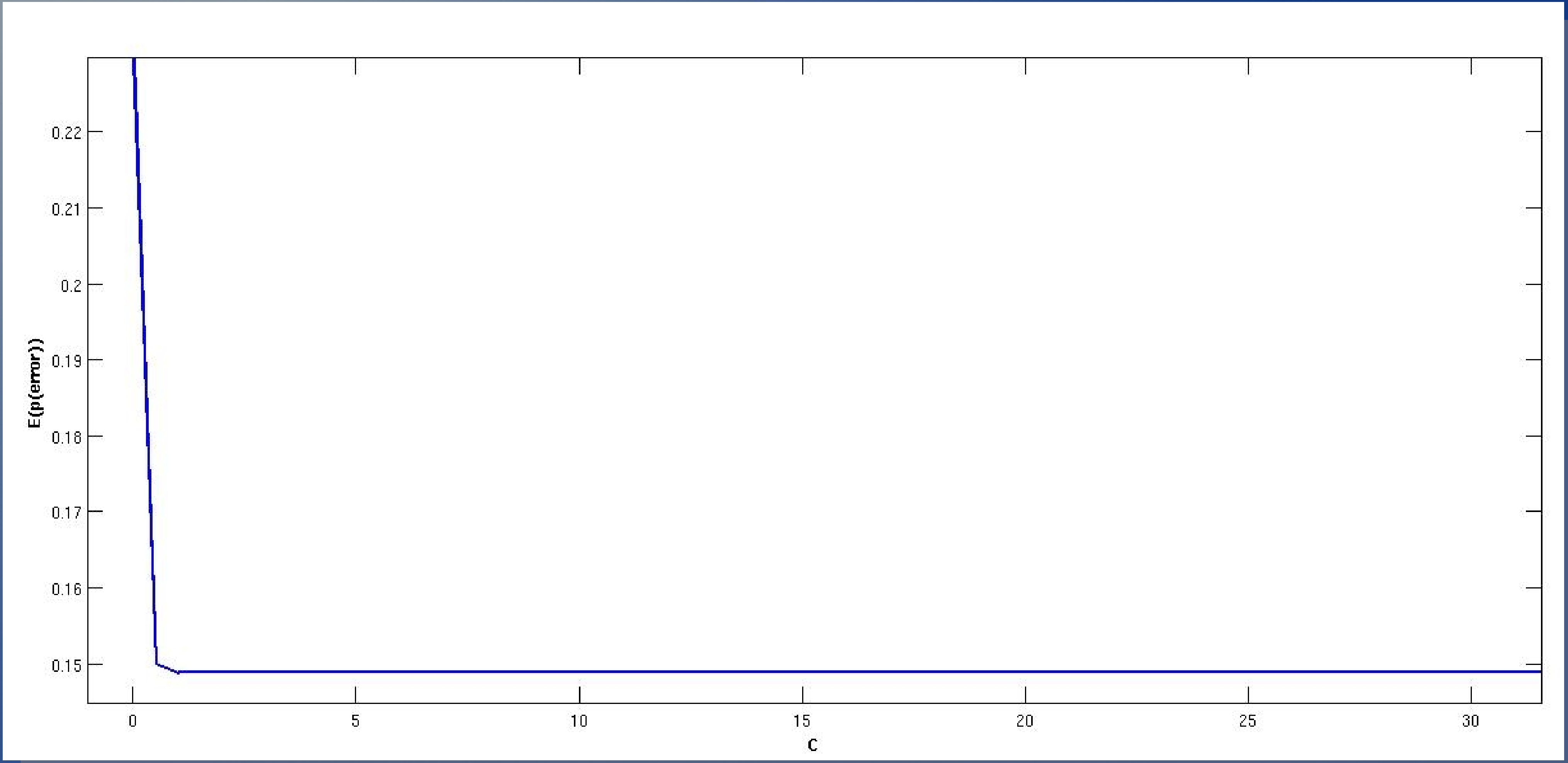
Results using SVMs on the data.

- Difference.
- Perimeter.
- Difference + Perimeter.

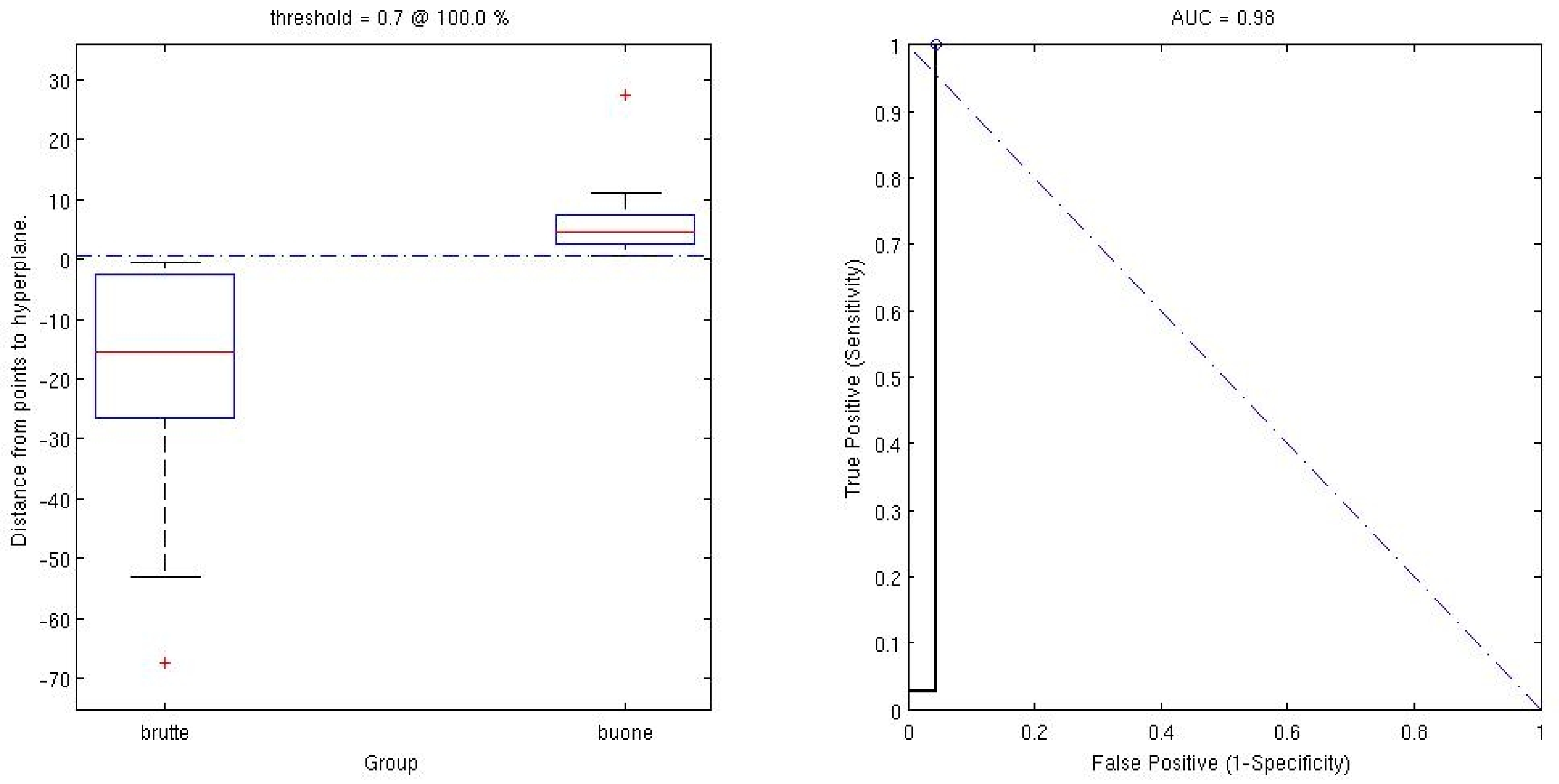
Difference.

Kernel.	C.	E(p(error))	ExpError.
Polynomial Degree 1.	50	0.2959	0.2373
Polynomial Degree 2.	1.0199	0.1488	0
Polynomial Degree 3.	0.01	0.2169	0
Gaussian σ = 1.	12.6337	0.3404	0.0702
Gaussian σ = 2.	34.3466	0.2779	0.1053

Difference.



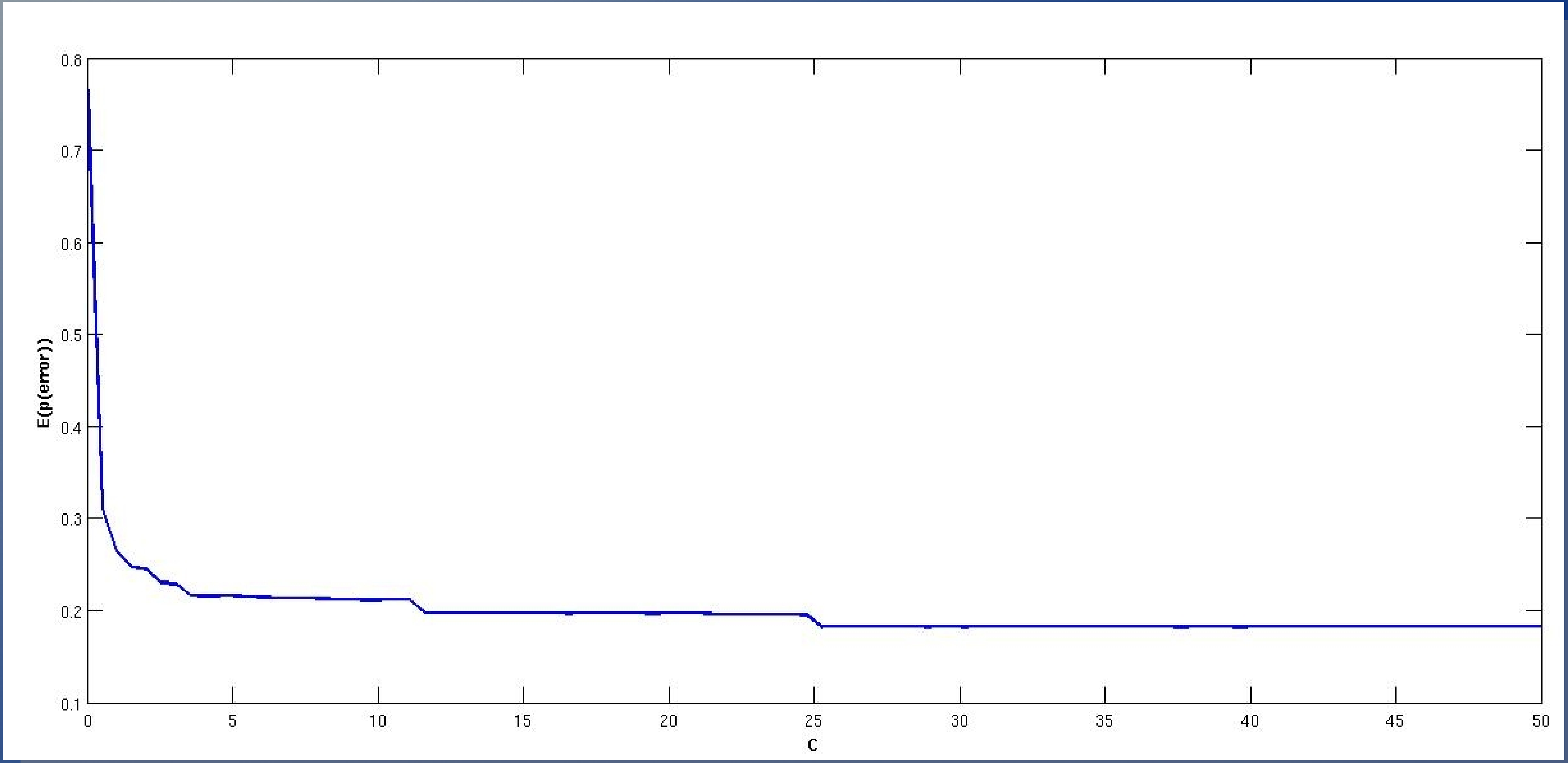
Difference.


threshold = 0.7 @ 100.0 %
Distance from points to hyperplane.
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
brutte
buone
Group
AUC = 0.98
True Positive (Sensitivity)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
False Positive (1-Specificity)

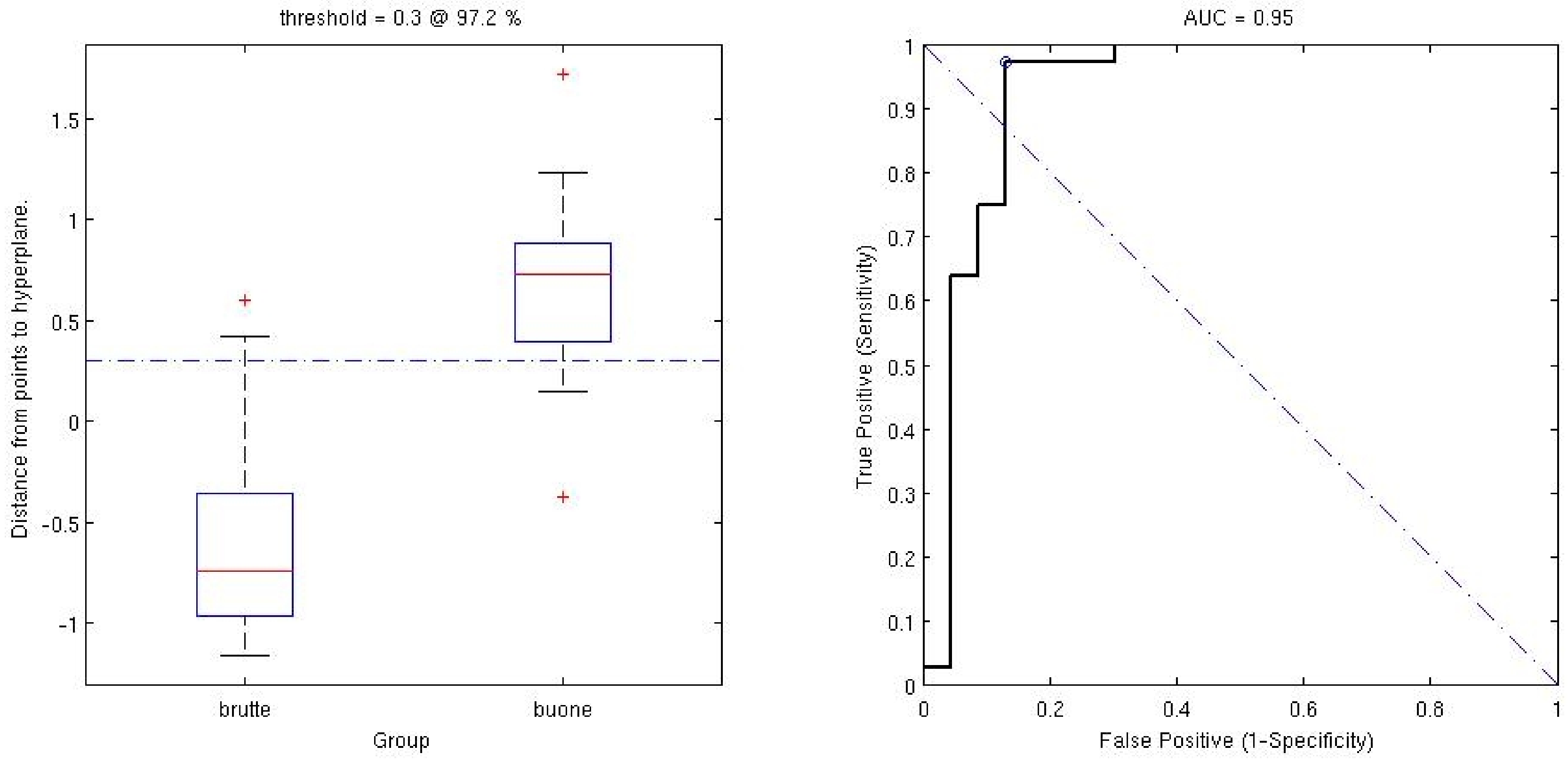
Perimeter.

Kernel.	C.	E(p(error))	ExpError.
Polynomial Degree 1.	39.3961	0.1818	0.1356
Polynomial Degree 2.	19.7030	0.1841	0.0678
Polynomial Degree 3.	4.5545	0.2005	0.0678
Gaussian σ = 5.	50	0.3657	0.0339
Gaussian σ = 10.	49.4951	0.3140	0.1356

Perimeter.



Perimeter.

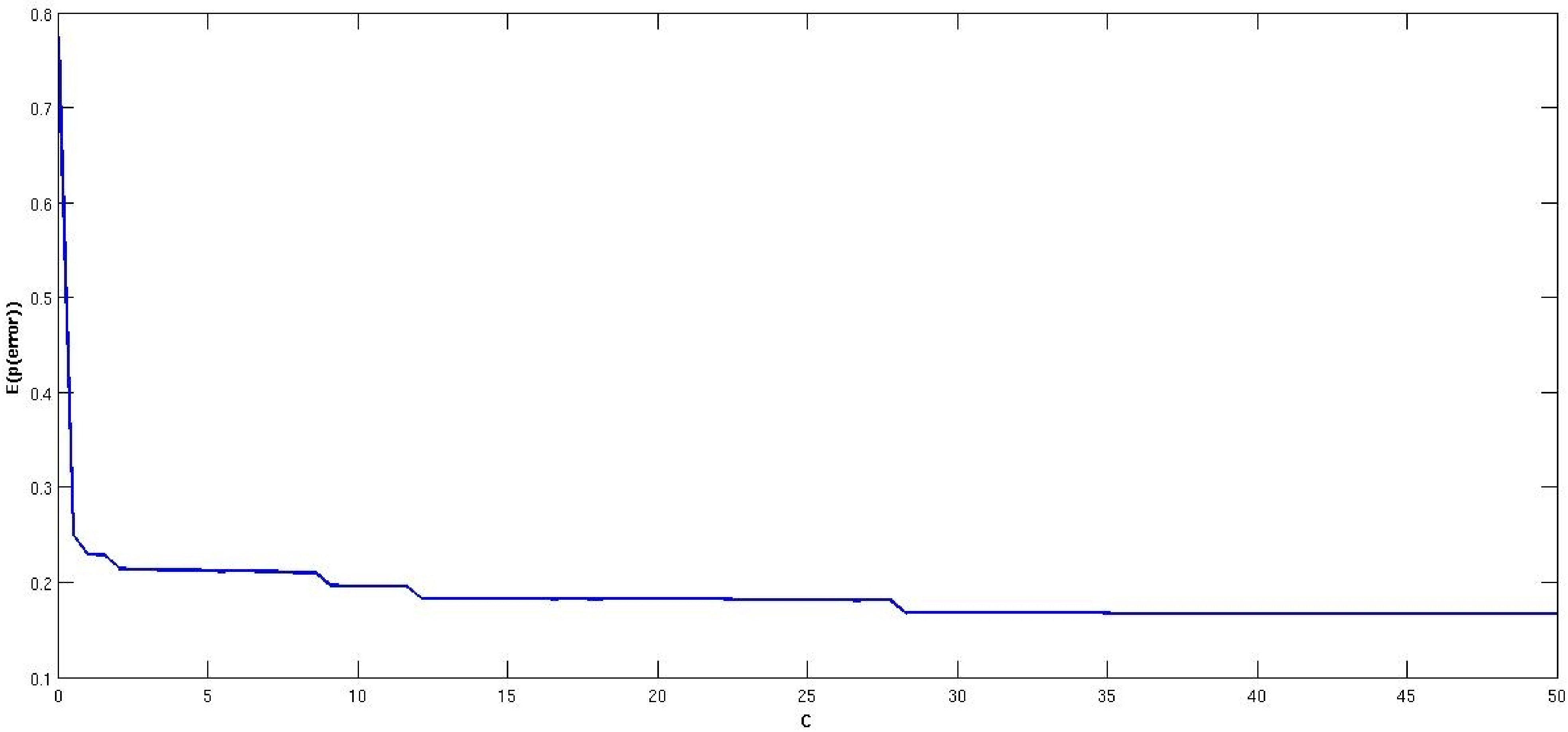


threshold = 0.3 @ 97.2 %
Distance from points to hyperplane.
1.5
1
0.5
0
-0.5
-1
brutte
buone
Group
AUC = 0.95
True Positive (Sensitivity)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
False Positive (1-Specificity)

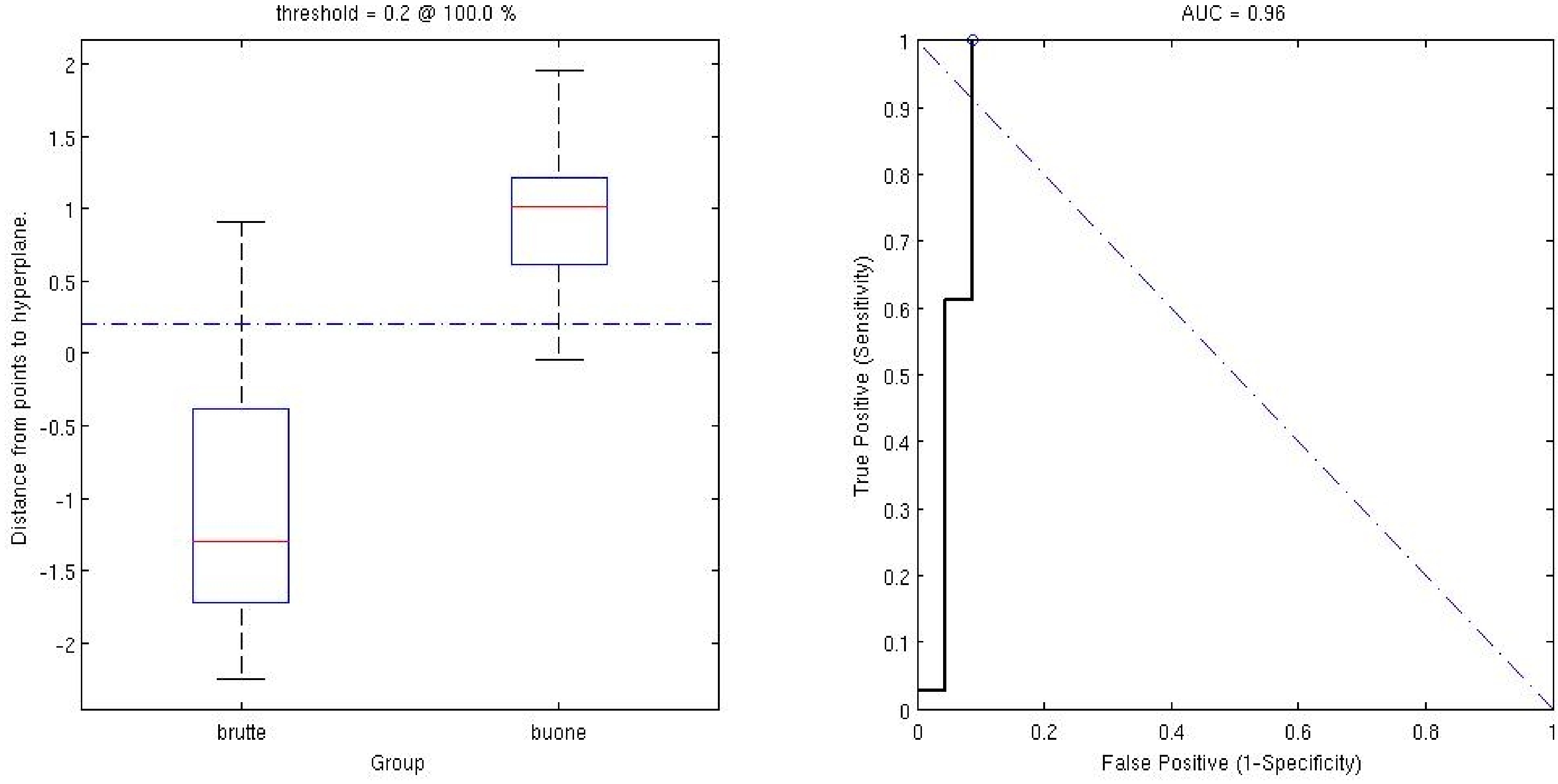
Difference + Perimeter.

Kernel.	C.	E(p(error))	ExpError.
Polynomial Degree 1.	41.4159	0.1666	0.0679
Polynomial Degree 2.	5.0595	0.2517	0
Polynomial Degree 3.	1.0199	0.3238	0
Gaussian σ=1.	13.7926	0.5944	0
Gaussian σ=2.	3.0397	0.3660	0.0339

Difference + Perimeter.



Difference + Perimeter.



Coregistration PET/MRI.

The coregistration PET/MRI is a necessary step for combining functional information from PET images with anatomical information in MR images.

Coregistration PET/MRI.

- Information-theoretic similarity measures, such as Mutual Information (MI), have been successful for coregistration since they don't assume linear intensity dependency between image modalities, such as intensity-difference or correlation-based metrics.
- Subsequent refinement is the Normalized Mutual Information (NMI).

Coregistration PET/MRI.

Information contributed by the image is the entropy

$$H(A) = -\sum_{a \epsilon A} p(a) log(p(a)).$$

$$H(B) = -\sum_{b \epsilon B} p(b) log(p(b)).$$

Coregistration PET/MRI.

Mutual Information.

Measure of the joint entropy with respect to the marginal entropies.

$$I(A,B) = H(A) + H(B) - H(A,B),$$

$$I(A,B) = \sum_{a \epsilon A} \sum_{b \epsilon B} p(a,b) log(\frac{p(a,b)}{p(a)p(b)}).$$

Coregistration PET/MRI.

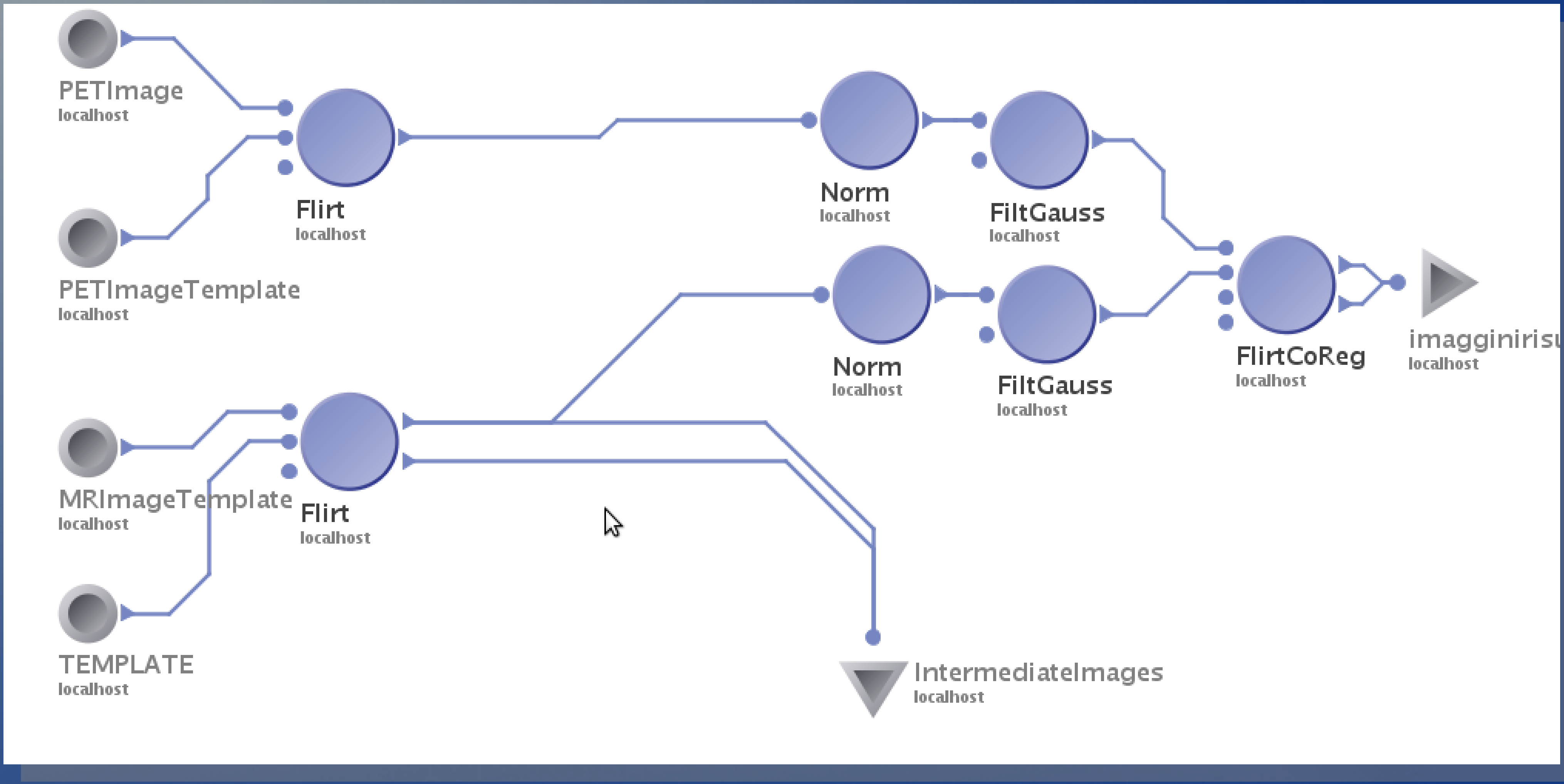
Normalized Mutual Information.

Maes et. al.

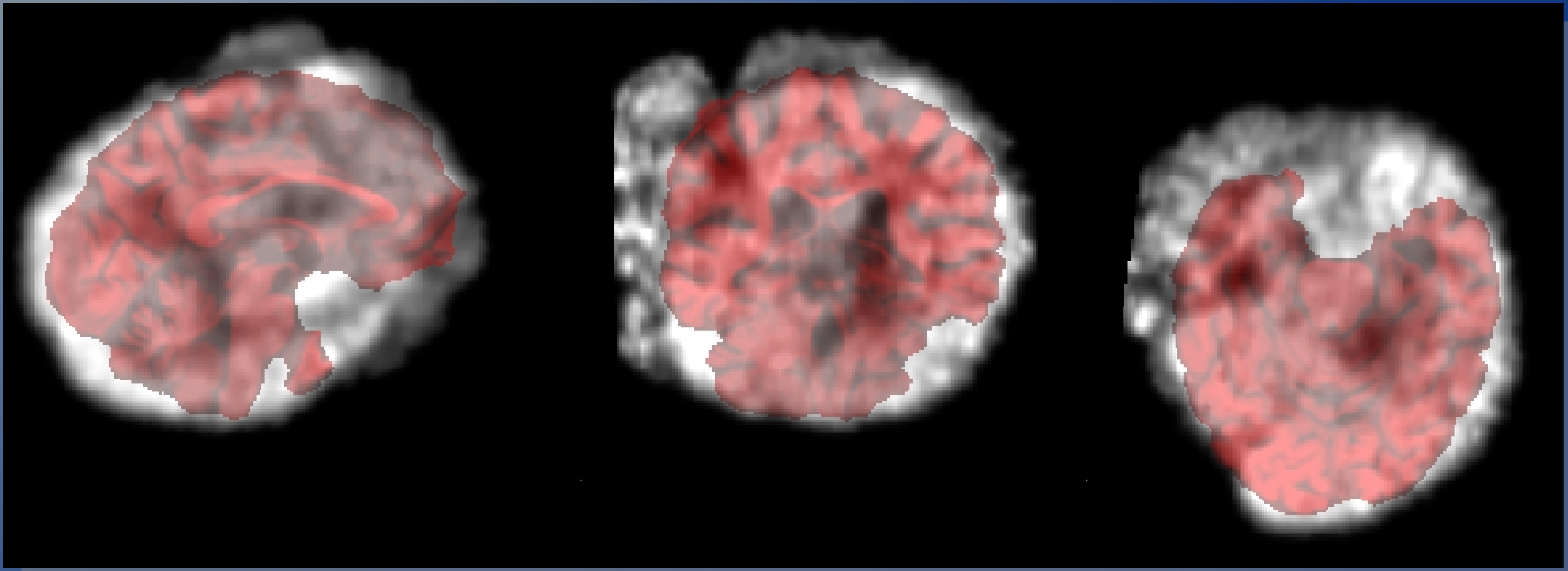
$$\hat{I}(A,B) = \frac{2I(A,B)}{H(A)+H(B)},$$

$$\widehat{I}(A,B) = H(A,B) - I(A,B).$$

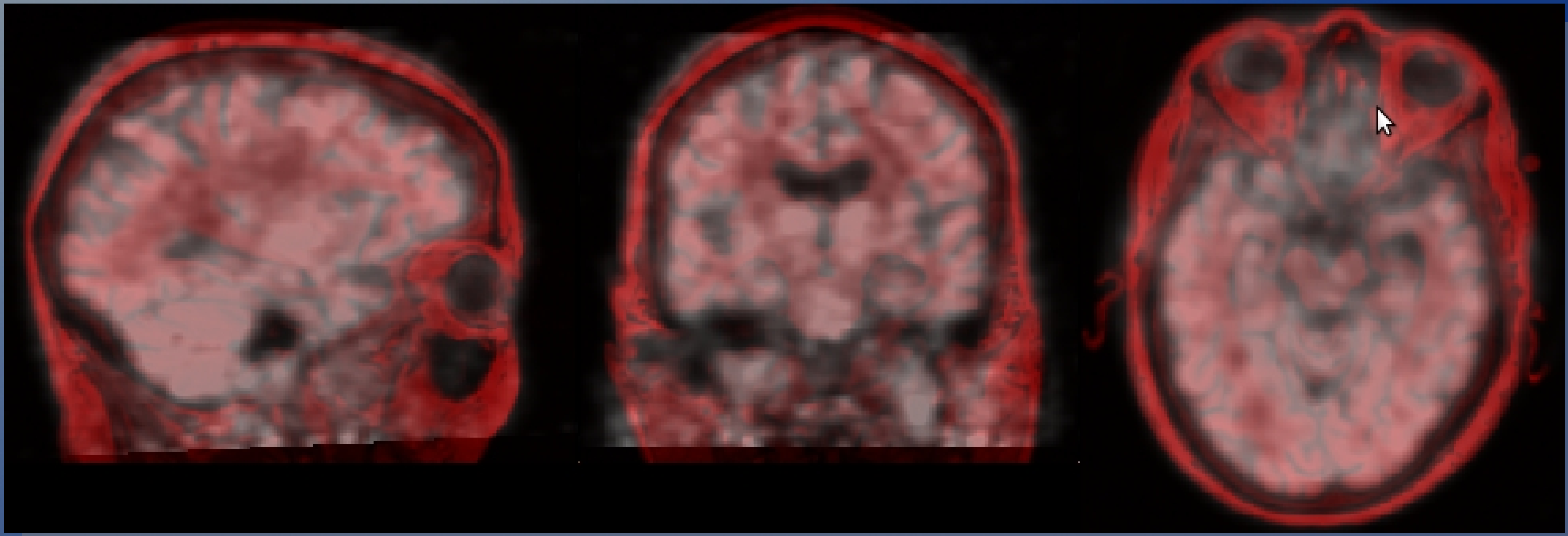
Coregistration PET/MRI.



Coregistration PET/MRI, using BET.



Coregistration PET/MRI, using Mtual Information.



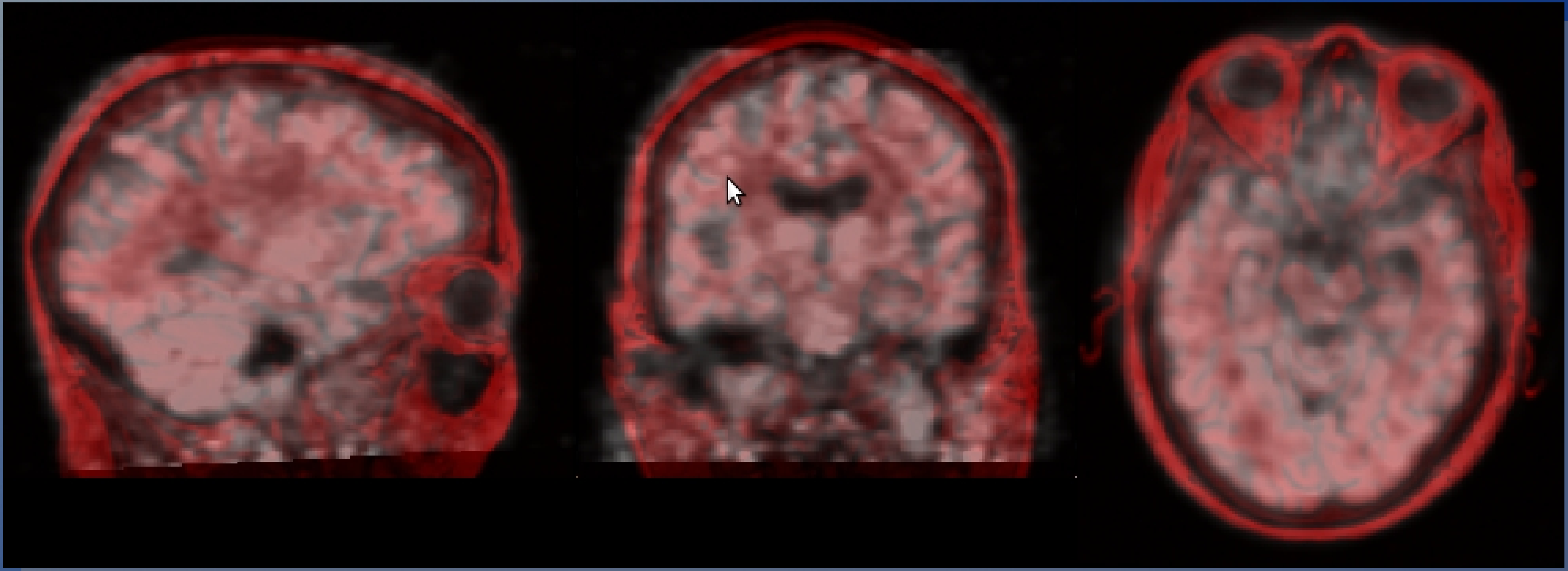
Coregistration PET/MRI.

Normalized Mutual Information.

Studholme

$$\hat{I}(A,B) = \frac{H(A) + H(B)}{H(A,B)}.$$

Coregistration PET/MRI, using Normalized Mutual Information.



Coregistration PET/MRI.

MI-based registration have three important limitations:

- Do not incorporate spatial information.
- Are sensitive to interpolation artifacts.
- Their computation are time-consuming.

Coregistration PET/MRI.

One way to improve the coregistration is the use of the Curvelet transform.

The Curvelet transform is a higher dimensional generalization of the Wavelet transform. It is design to represent piecewise smooth functions at different scales and angles. In this scheme, curved edges are approximated with very few coefficients.

Conclusions.

- Overview of Support Vectors Machines for classification, how we have used SVM on the data to find the optimal hyperplane and some results using SVMs.

- A Coregistration procedure using Mutual Information or Normalized Mutual Information, and a probably improvement (under study) to this coregistration procedure using the Curvelet Transform.

Thank you!