Pit Pattern Classification with Support Vector Machines and Neural Networks

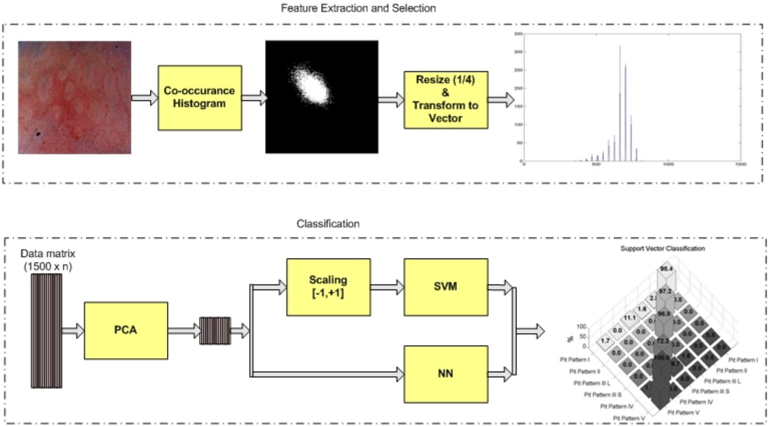
Christian Kastinger

February 1, 2007

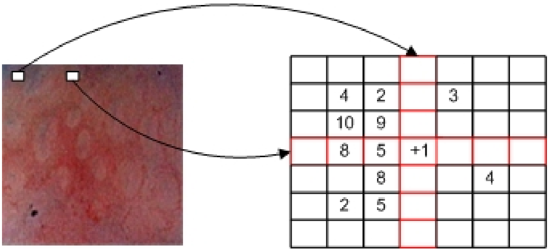
Objectives

Investigation of support vector machine and neural network for classification of pit pattern (256 x 256 RGB images) with previous feature extraction (co-occurrence histogram) and feature selection (principle component analysis) for a 2-class and a 6-class problem.

Classification Process



Co-occurrence histogram



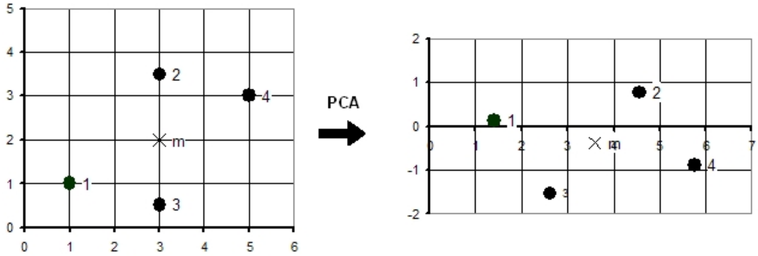
- Considers dependencies between adjacent pixels.
- Co-occurrence distance (between two samples) can be varied.
- Orientation (vertical,horizontal,...) is not fixed.

Principle Components Analysis (PCA)

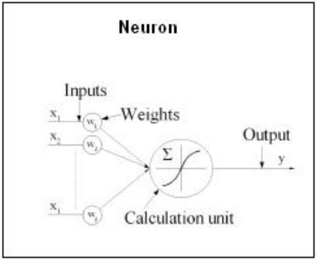
- PCA aims to provided a better representation with lower dimension.
- Compaction of information.
- Process
  - Create a new mean-adjusted data matrix $\tilde{\mathbf{X}}$.
  - Calculate a $m \times m$ covariance matrix $\mathbf{\Sigma}$ from the mean-adjusted data $\tilde{\mathbf{X}}$.
  - Compute n significant eigenvectors $\mathbf{W}$ from the covariance matrix $\mathbf{\Sigma}$.
  - Perform dimensionality reduction: $\mathbf{Y} = \mathbf{W}^T\tilde{\mathbf{X}}$.
- n is a tradeoff between "compression" and quality.

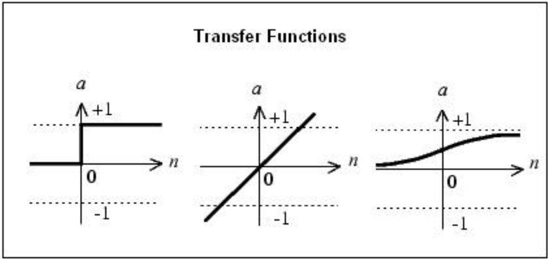
PCA Example

- Data points : $x = \begin{pmatrix} 1 & 3 & 3 & 5 \end{pmatrix}$; $y = \begin{pmatrix} 1 & 3.5 & 0.5 & 3 \end{pmatrix}$
- $\tilde{\mathbf{X}} = \begin{pmatrix} 2.667 & 1.333 \\ 1.333 & 2.1667 \end{pmatrix} \Longrightarrow \mathbf{W} = \begin{pmatrix} 0.639 & -0.77 \\ -0.77 & -0.639 \end{pmatrix}$
- $x' = \mathbf{W}^T * \tilde{x} = \begin{pmatrix} 1.409 & 4.546 & 2.63 & 5.767 \end{pmatrix}$
- $y' = \mathbf{W}^T * \tilde{y} = \begin{pmatrix} 0.131 & 0.778 & -1.63 & -1.532 & -0.885 \end{pmatrix}$



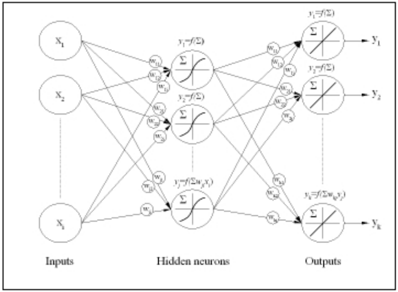
Basic Concept





- Inputs: $x_1, ..., x_j \in [0, 1]$
- Synapses: $w_1, ..., w_j \in R$
- Neuron: $net = \sum w_j * x_j$
- Output: $y = f(net - \theta)$
- Bias value: θ

Multi Layer Network



- Layer weights are adjusted during learning based on some input/output patterns.
- Learning typically starts at the output layer and moves toward the input layer (back-propagation).

Mathematical Model

- Activation of the hidden layer

$$net_j = \sum_i w_{ji} x_i$$

- Output of the hidden layer

$$y_j = f(net_j) = f\left(\sum_i w_{ji} x_i\right)$$

- Activation of the output layer

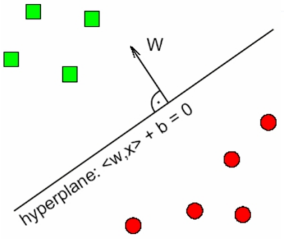
$$net_k = \sum_j w_{kj} f(net_j)$$

- Net output

$$y_k = f(net_k) = f\left(\sum_j w_{kj} f\left(\sum_i w_{ji} x_i\right)\right) = f\left(\sum_j w_{kj} y_j\right)$$

Basic Concept

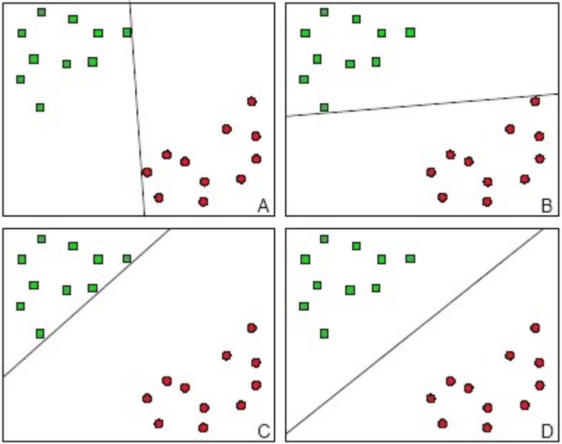
- A support vector machine is a learning algorithm which attempts to separate patterns by a hyperplane defined through:
  - normal vector w
  - offset parameter b.



Hyperplane definition:

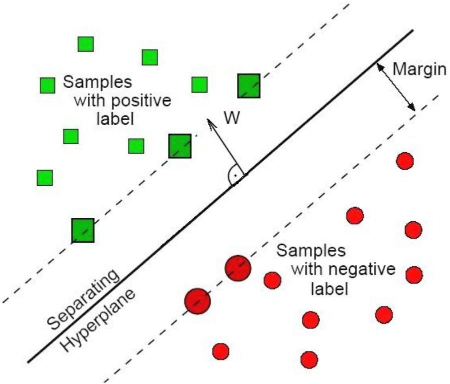
$$H = \{x \mid \langle w, x \rangle + b = 0\}$$

What is an Optimal Hyperplane?



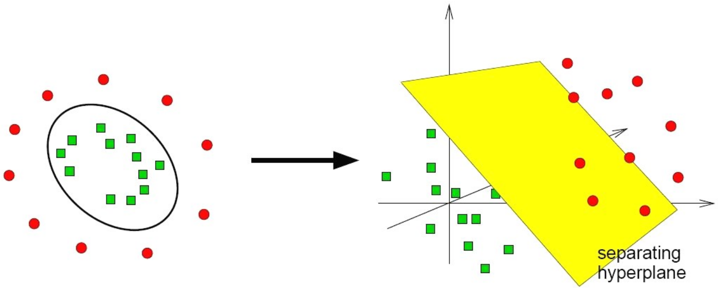
Separation with maximal Margin

- *Support vectors* are all points lying on the margin closest to the hyper plane.



Kernel Trick

- Nonlinear and complex separation in the 2-dimensional *input space*.
- Easier and often linear separation in higher dimensional *feature spaces*.



Kernel Examples

- Linear kernel

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}' = \langle \mathbf{x}, \mathbf{x}' \rangle$$

- Polynomial kernel of degree d

$$k(\mathbf{x}, \mathbf{x}') = (\gamma + \langle \mathbf{x}, \mathbf{x}' \rangle + coef0)^d$$

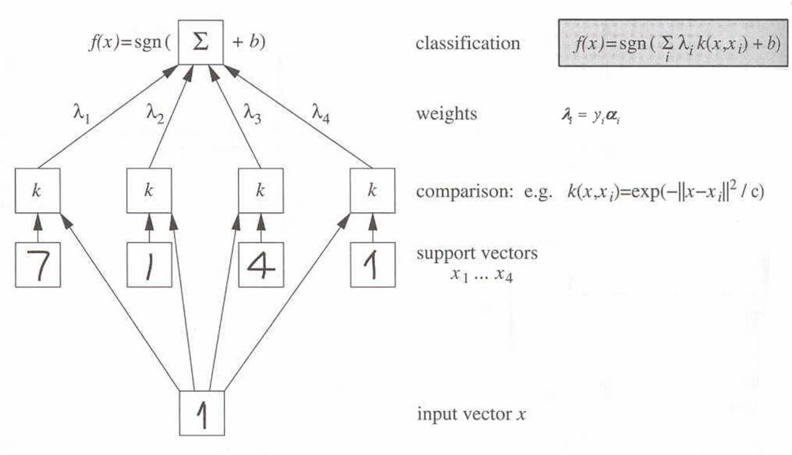
- Radial basis kernel (RBF)

$$k(\mathbf{x}, \mathbf{x}') = exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$$

- MLP or Sigmoid kernel

$$k(\mathbf{x}, \mathbf{x}') = tanh(\gamma \langle \mathbf{x}, \mathbf{x}' \rangle + coef0)$$

Classification Principle



Mathematical Model

- Dual optimization problem

$$\max_{\alpha \in \mathcal{R}^m} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i y_i \alpha_j y_j k(\mathbf{x}_i, \mathbf{x}_j)$$

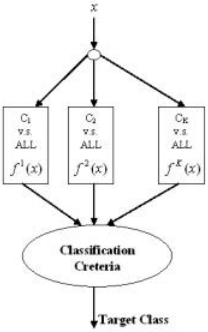
$$\text{subject to} \quad \alpha_i \geq 0, \quad \text{for all} \quad i = 1, ..., m, \quad \text{and} \quad \sum_{i,j=1}^{m} \alpha_i y_i = 0$$

- Decision function

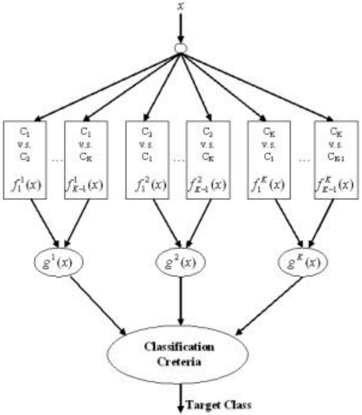
$$\begin{aligned} f(\mathbf{x}) &= sgn\left(\sum_{i=1}^{m} \alpha_i y_i \langle \phi(\mathbf{x}), \phi(\mathbf{x}_i) \rangle) + b\right) \\ &= sgn\left(\sum_{i=1}^{m} \alpha_i y_i k(\mathbf{x}, \mathbf{x}_i) + b\right) \end{aligned}$$

Multi-Class Approaches

- Decomposition into several binary classification tasks.

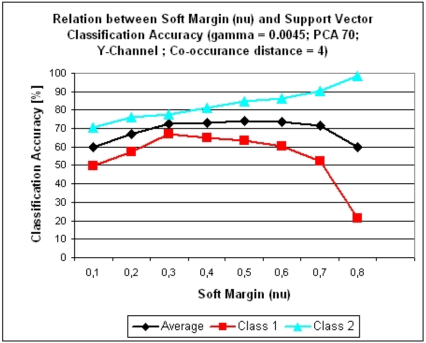


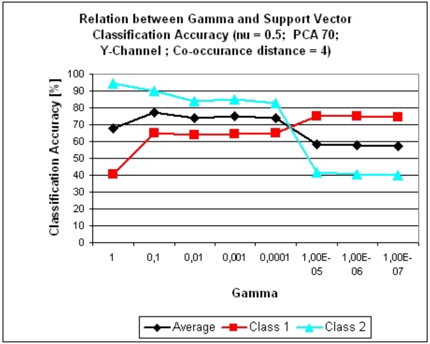
one-against-all



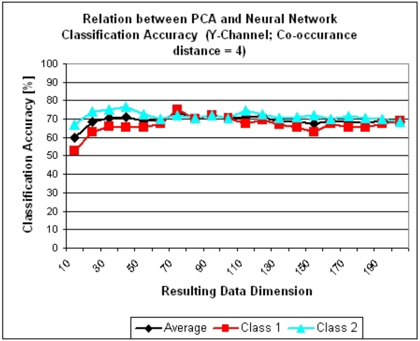
one-against-one

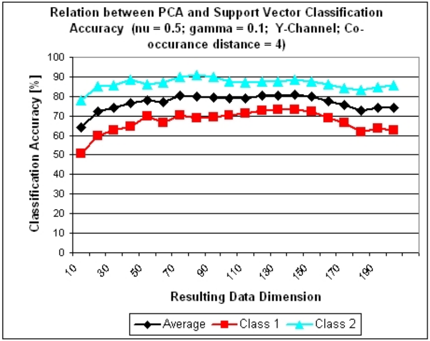
SVM Optimization



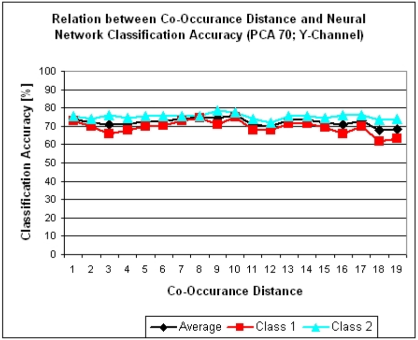


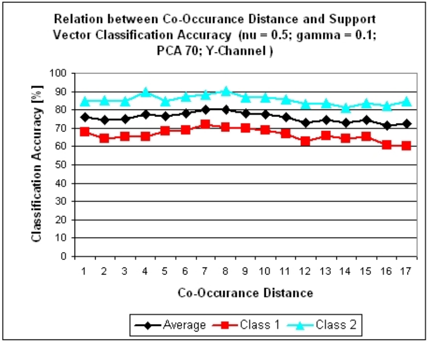
PCA Dependency



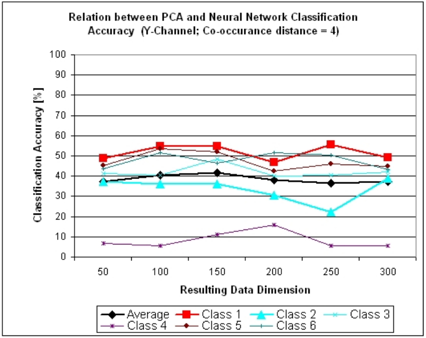


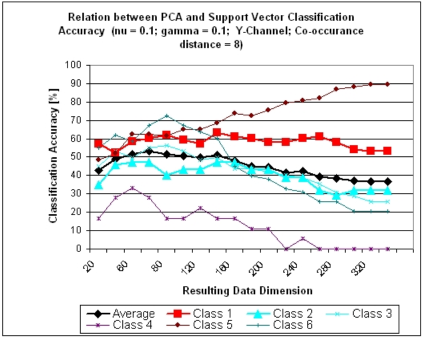
Co-occurrence Distance Dependency





PCA Dependency for 6 Classes





Additional Investigations

- Color-histogram (3-dimensional).
  - Too high data dimension.
  - Low classification results due to high data compression.
- Vertical co-occurrence histogram.
  - 3-5% lower classification results compared with horizontal histogram.
- Combination of horizontal and vertical co-occurrence histogram.
  - Lower classification results as with horizontal histogram.

Problems

- High data dimension.
- Data scaling.
- Time intensive parameter optimization.
- SVM accepts invalid input.
- Too less training samples

Summary

- SVM provides 10% better results than the NN.
- SVM parameter have to be optimized carefully.
- Better classification with PCA due to higher compaction of information.
- Low impact co-occurrence distance on classification accuracy.

Outlook

- Evaluation of SVM with a higher amount of pit patterns.
- Consideration of other feature extraction and selection strategies.
- Investigation of other neural network topologies.

Bibliography

- R.O. Duda, P.E. Hart and D.G. Stork. *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, 2001.
- C.C. Chang and C.J. Lin. *LIBSVM: a library for support vector machines*. URL: `http://www.csie.ntu.edu.tw/ cjlin/libsvm`, [Feb. 24, 2006].
- N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge: Cambridge University Press,2000.
- B. Schoelkopf and A.J. Smola. *Learning with kernels*. Cambridge,MA: MIT Press, 2002.
- J.E. Jackson *A User's Guide to Principal Components*. John Wiley & Sons Inc, 2003.
- P. Chang and J. Krumm. *Object recognition with color cooccurance histograms*. In Proceediungs of CVPR '99, 1999.

Questions?

Thank you for your attention.