Linear algebra in machine learning

Prof. Dr. Bart De Moor
ESAT-STADIUS, KU Leuven
Bart.DeMoor@kuleuven.be
www.bartdemoor.be

In this presentation, we will focus on least squares support vector machines.
These are machine learning algorithms, that basically, in their so-called primal formulation, start from a constrained optimization problem with a quadratic (least squares) objective function. The ingredients are a function that maps given data points to a higher dimensional (possibly infinite dimensional) feature space, a vector of unknown weights and a vector of error variables that, in the case of a classification problem, tolerates misclassifications.

First, we will show how standard (Gaussian) linear regression problems with linear constraints and a priori information on the distribution of the vector of unknowns and residuals, lead to constrained weighted least squares problems. These can be solved, either for the vector of unknowns, the vector of residuals or the Lagrange multipliers introduced for the constraints. This leads to three different, yet equivalent solutions.

We then use this result to solve a given least-squares optimization problem in its so-called dual form, which ultimately leads to (large) set of linear equations (in the literature called a Karush-Kuhn-Tucker system). The feature mapping is not to be known explicitly, but leads (via the so-called 'kernel trick') to a kernel function, that one can choose depending on the requirements and a priori information on the data and the problem. Examples are linear kernels, polynomial ones, Radial Basis Functions, etc.

So in the first part of the talk, we will elaborate how plain linear algebra allows to formulate non-linear classification and regression problems, by clever exploitation of the 'kernel trick'.

In the second part of the talk, we turn the problem around. We will show how the machinery of least squares support vector machines can be deployed to formulate kernel extensions of Principal Component Analysis (SVD), Canonical Correlation Analysis (Angles between subspaces) and Partial Least Squares. This allows us to generalize these well known 'linear' statistical techniques, to non-linear extensions, while we only use insights and algorithms from linear algebra.

We will comment on issues of sparsity and robustness, and illustrate our presentation with an abundance of applications from industrial and medical data sources.