

Learning with Structured Input and Output Data

Dr. Christoph Lampert
Institute of Science and Technology Austria (IST Austria)
Am Campus 1, A-3400 Klosterneuburg, Austria
Email: chl@ist.ac.at WWW: <http://www.ist.ac.at/~chl/>

In our everyday life, we are surrounded by structured data: text documents, digital images, and even chemical molecules are examples of data sources that consist not just vectors in some fixed vector space, but that consist of meaningful subparts (e.g. words, pixels, atoms). It is particularly characteristic for structured data that their information contents is often not just encoded in the values of the parts, also also in their relations to each other. While early Pattern Recognition techniques were only able to handle data in vector form, many modern techniques are capable of also working with structured input data. In the lecture we will introduce *kernel methods*, e.g. support vector machines, that allow us perform classification, regression or dimensionality reduction based only on a similarity measure between data samples. With a better understanding of the structured data and its part, we can develop better similarity measures, and this translated directly into better predictive performance. In the lecture we will take a geometric approach to kernel methods, that in particular allows us to also obtain an intuitive understanding of kernel methods as generalizations of linear methods while requiring a lot of formal mathematical background. A very more recent trend in Machine Learning is to use structured data not only as input, but also as the output of learning techniques. In the second part of the lecture we will show how many problems in image processing, computer vision and information retrieval can be cast as prediction problems with structured outputs, and we will introduce the *structured support vector machine* that is able to learn solutions to many such problems from structured training examples.