

# Generalisation Analysis as a Foundation for Classification

John Shawe-Taylor

Centre for Computational Statistics and Machine Learning  
University College London  
Gower Street, London WC1E 6BT  
England

**Abstract.** The presentation will review frequentist approaches to assessing the performance of classifiers on new data. This provides a statistical view that rests on the relatively agnostic assumption that data are drawn independently from an unknown but fixed distribution. The analysis is concerned with assessing the generalisation, that is the performance of a classifier on unseen data drawn from the same distribution. The implications of these results for well-known learning algorithms such as Support Vector Machines will be reviewed. Extensions of the approach to subspace methods such as (kernel) principal components analysis will be discussed as well as some results for clustering algorithms. In both cases we are again concerned with the usefulness of the deduced simplifications when handling new data. The emphasis will be on the framework that the analysis provides for designing principled algorithms that infer underlying structure from samples of data.

## Keywords

GENERALISATION, FREQUENTIST, SUPPORT VECTOR MACHINES,  
PRINCIPAL COMPONENTS ANALYSIS, CLUSTERING