## SUBMODULARITY AND ADAPTATION

## Jeff A. Bilmes

## University of Washington, Seattle, Washington 98195-2500, USA

## ABSTRACT

Convexity is a property of real-valued functions that enable their efficient optimization. Convex optimization moreover is a problem onto which an amazing variety of practical problems can be cast. Having strong analogs to convexity, submodularity is a property of functions on discrete sets that allows their optimization to be done in only polynomial time. Submodularity generalizes the common notion of diminishing returns. Like convexity, a large variety of discrete optimization problems can be cast in terms of submodular optimization. The first part of this talk will survey recent work taking place in our lab on the application of submodularity to machine learning, which includes discriminative structure learning and word clustering for language models. The second part of the talk will discuss recent work on a technique that for many years has been widely successful in speech recognition, namely adaptation. We will view adaptation in a setting where the training and testing time distributions are not assumed identical (unlike typical Bayes risk theory). We will derive generalization error and sample complexity bounds for adaptation which are specified in terms of a natural divergence between the train/test distributions. These bounds, moreover, lead to practical and effective adaptation strategies for both generative models (e.g., GMMs, HMMs) and discriminative models (e.g., MLPs, SVMs). Joint work with Mukund Narasimhan and Xiao Li.