**===Slide 1=============================**

Hi everyone!

My name is Tan Nguyen and this is the presentation of our paper “Greedy Convex Ensemble”, which is a joint work with my supervisors, Nan Ye and Peter Bartlett.

**===Slide 2=============================**

The problem that we studied in this paper is learning a convex ensemble of simple basis models.

It means we are given some set G of basis models, for example, the basis models could be artificial neurons or small neural networks.

The linear hull is the set of all possible linear combinations of elements in G. It is well known that the linear hull of even simplest basis models, such as artificial neurons, is a universal approximator.

On the other hand, we have convex hull, which is the set of all possible convex combinations of elements in G. Obviously, convex hull is a subset of the linear hull, where the combination weights are constrained to be non-negative and sum up to 1.

The convex hull is the subject of our study. From theoretical point of view, we would like to understand more about the capacity and the generalization property of convex hulls. From empirical point of view, we want to find the best approach to learn from a convex hull, and compare its performance with competing methods.

So, let’s get started!

**=== Slide 3=============================**

Our first theoretical contribution is given in Proposition 2 in the paper. It basically says that for linear threshold basis models, linear hulls have infinite Pseudodimension and Rademacher complexity, while convex hulls also have infinite Pseudodimension, but their Rademacher complexity is finite.

This result indicates that linear hulls have unbounded capacity, thus are prone to overfitting, while convex hulls have rich but bounded capacity, thus they can be seen as a regularized version of linear hulls with the convex constraint acts as a natural regularizer.

**=== Slide 4=============================**

Our second contribution is a generalization bound for convex hulls. It is given in Theorem 2 and 3 in the paper.

Theorem 2 basically says that for a general class of Lipschitz loss functions, with high probability, minimizing the empirical risk of f over the convex hull results in minimizing the bound of its expected risk.

Theorem 3 says that in the convex hull, the empirical risk minimizer converges to the expected risk minimizer at the rate of O(1/sqrt(n)), where n is the number of samples.

**=== Slide 5=============================**

Moving on to algorithms, in Section 4 of the paper, we discussed several approaches to learn a convex ensemble of basis models. The traditional way, which we call nonlinear greedy algorithm, simply chooses the new model in each iteration such that it minimizes the empirical risk. Another approach is to use Frank Wolfe algorithm, which in each iteration chooses the new model that aligns the most with the negative functional gradient of the empirical risk.

From empirical point of view, we found that the traditional nonlinear greedy algorithm doesn’t work as well as Frank Wolfe. And among variants of the Frank Wolfe algorithms, the Pairwise Frank-Wolfe works the best.

**=== Slide 6=============================**

We also performed an empirical comparison of the greedy convex ensemble using the Pairwise Frank Wolfe algorithm with other competing methods, such as neural network, boosting, random forest. The results show that greedy convex ensemble is competitive with other methods in the majority of cases while it requires little effort on hyper-parameters tuning.

**=== Slide 7=============================**

So, that was the brief overview of our paper. Thank you for watching and hope to see you at the conference.