Locality-Sensitive Hashing and Beyond

Abstract:

Locality-Sensitive Hashing (LSH) is a powerful technique for the approximate nearest neighbor search (ANN) in high dimensions. In this talk, I will present two recent results:

1) I will show a data structure for ANN for the Euclidean distance that provably outperforms the best possible LSH-based data structure. We proceed via designing a good *data-dependent* hash family.

2) I will show a practical and optimal LSH family for the cosine similarity (a.k.a. Euclidean distance on a sphere). It substantially outperforms the celebrated Hyperplane LSH family. Along the way, I will try to debunk two popular myths about LSH:

   * LSH-based data structures consume too much memory and are thus impractical;
   * Optimal LSH constructions are too complicated to be made practical.