Topics in High Dimensional Geometry

Hariharan Narayanan

February 13, 2015

This course will revolve around the following two topics.

1 Fitting low dimensional manifolds to high dimensional data

The field of manifold learning is based on the hypothesis that high dimensional data often lie in the vicinity of a low dimensional submanifold having controlled curvature and volume. We will develop the theory behind an algorithm for testing this hypothesis with statistical guarantees and bounds on the run-time, which do not depend on the dimension of the ambient space.

Topics covered will include some of Fefferman's work on *Whitney interpolation*, *random projections* and the *Johnson-Lindenstrauss Lemma*, material on *Dudley's entropy integral* and the work of Rudelson and Vershynin on *combinatorial entropy*, and some of Federer's work on the *reach* of subsets of Euclidean space.

2 High dimensional convex bodies

Let $K \subseteq \mathbb{R}^n$ be a convex body containing the unit ball, and contained in a ball of radius R. Suppose there is an oracle, which when presented with a point x outputs whether or not $x \in K$. Given such an oracle, we would like to do the following:

1. Produce a random sample from K from a distribution that is close to the uniform distribution in total variation distance.

2. Produce an approximation of the volume of K that holds with high probability.

3. Optimize a convex function f (given by an oracle) over K.

We will discuss a number of *Markov Chain algorithms* for 1. above, including Ball walk introduced by Lovasz and Simonovits and Hit-and-Run developed by Lovasz and Vempala. We will discuss *Markov Chain Monte Carlo algorithms* for 2. above. We will use sampling from distributions that are more and more peaked around the optimum as a method of addressing 3.

Prerequisites: Comfort with probability, linear algebra and analysis.