CHENNAI MATHEMATICAL INSTITUTE

Machine Learning

Sep 24, 2015.

(1) Write a program to generate n points uniformly at random in a ball of radius 1 in d dimensions starting with a unit variance Gaussian in d dimensions. Using $n = 200$ and $d = 50$, for each i and point $1 \leq j \leq 200$, plot the points (j, x_i^j). Find the α_i for which the points are within $|\alpha_i/\sqrt(50)|$. And output the values of the α_i obtained.

Now take 10 random lines through the origin and project each generated point onto the line. Plot the distribution of these projected points. For each of them fit a Gaussian model and compare with what is theoretically predicted

(2) Generate 500 points on the surface of a sphere of radius 1 in 50 dimensions, uniformly at random. Generate 10 more points randomly. For each of these 5 points calculate a narrow band around the equator assuming that the point is the North pole. How many of these 500 points are in the band corresponding to each of the equators. Plot a graph of band size versus number of points.

Also output how many points are in each of the 5 equators.

Run this for 100 dimensions as well.

(3) Repeat exercise 1 this time selecting not a line but a 10 random, 5 dimensional subspaces through the origin. Project the points on to these subspaces and fit a Gaussian model for each of them. Calculate the mean and the covariance matrix for each.

(4) Generate 50 points uniformly at random on the surface of a 200 dimensional sphere.Calculate the inter point distances. Now project the points on to k dimensional subspaces for k=1,2,3,5,10,50. Calculate the difference between the new distances and $\sqrt(\frac{k}{d})$ times the original distance.