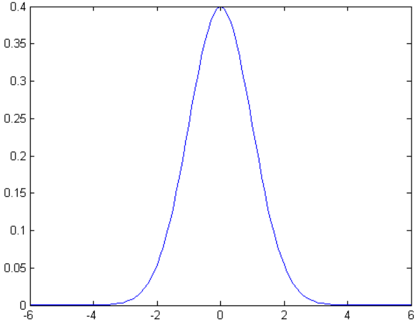
Projections of multinormal distribution using Euclidean distance

By
Yasin

December 2011

INTRODUCTION

In lower dimensional normal distributions, mass tends to concentrate at the center of of the distribution. For Example if we consider a gaussian distribution of dimension one, most of the mass concentrates at the center of the distribution.



According to (Lifshits,1995), 99.7% of the total mass lies within the segment $[\mu - 3\sigma, \mu + 3\sigma]$. When a random number is drawn from the distribution,the probability that it is situated very near the center point is high. i.e, its Euclidean distance from the center point is quite small. This is similar to what is observed in the 2-dimensional case of a gaussian distribution. However when the dimension grows, most of the points shift away from the center the center point. The distance of the randomly sampled points from the multinormal distribution from the the center point becomes larger on average. This what (Damien Francois and Vincent Wertz, 2007) refer to as the concentration of norm phenomenon where the Euclidean distance is substituted by the any norm.

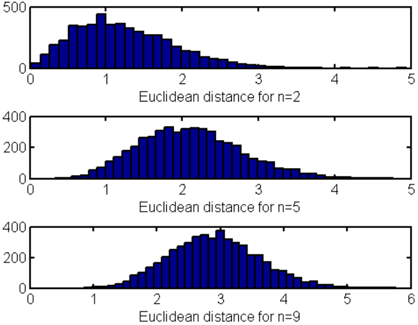


Figure: distribution of Euclidean distance from the center of the distribution of random vectors vectors drawn from n-dimensional gaussian distribution centered at the origin .

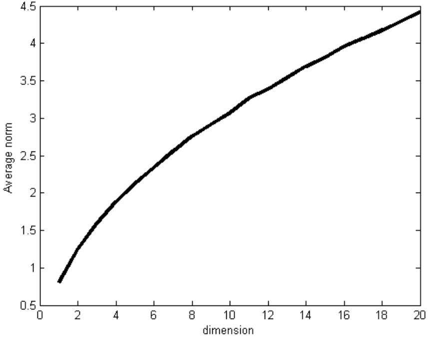


Figure: Variation of average distance of random vectors drawn from n-dimensional gaussian distribution centered at the origin .

distribution of distances

Lifshits, 1995; further gives the distribution of these norms in which the distances of the random vectors with reference to the center point of n-dimensional gaussian distribution follow the non normalized probability density of function

$$p(r) = (2\pi)^{-n/2} C_n r^{n-1} e^{-r^2/2} \tag{1}$$

where $C_n = \frac{(2\pi)^{n/2}}{\Gamma(n/2)}$.
Eqn1 gives the probability that a random vector drawn from an n-dimensional normal distribution lies at a distance r from the mean value of the distribution

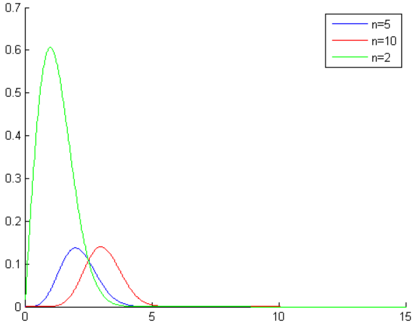


Figure: Nature of distributions of distances for various dimensions.

These distributions are in agreement with the simulation of the distributions presented earlier using histograms. For, the various dimensions, the corresponding distributions as given above and this further confirms that as the dimension grows, most of the points tend to concentrate at a distance further away from the center.

Conclusion

In conclusion therefore, for, the various dimensions, the corresponding distributions as given above and this further confirms that as the dimension grows, most of the points tend to concentrate at a distance further away from the center.

References

1. M. A. Lifshits(1995 pg 13-15) Gaussian random functions. Dordrecht, Kluwer academic plublishers
2. Damien Francois, vincencent Wertz and Michael Verleysen (2007) The concentration of Fractional distances. IEEE vol19 No.7