Muthiah M Muthaia Chettiar
David Lin

CS 485 Notes –April 19,2006

Consider a d-dimensional Gaussian, variance is σ^2. Centered at origin.

Pick a point at random: $x = [x_1, x_2, .. , x_d]$.

We ask what is the expected distance of point from the origin?

$E(dist^2) = E(x_1^2 + x_2^2 + .. + x_n^2)$
$= E(x_1^2) + E(x_2^2) + .. + E(x_d^2)$ (expectation of sum is sum of expectations)
$= d\, E(x_1^2)$ (all r.v are identically distributed)

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-x^2}{2\sigma^2}} dx$$

$$= d\sigma^2$$

For large d, the points are concentrated around expected value.

Alternative approach

$e^{\frac{-r^2}{2}} r^d$ (variance is 1).

Calculate the ratio of number of points at $\sqrt{d}$ and $\sqrt{d}$ + k

$$\frac{d}{dr} e^{\frac{-r^2}{2}} r^d = -re^{\frac{-r^2}{2}} r^d + d\, e^{\frac{-r^2}{2}} r^{d-1}$$

$$= e^{\frac{-r^2}{2}} r^{d-1}(-r^2 + d) = 0$$ (differentiate to find max)

$\Rightarrow \quad r = \sqrt{d}$

Radius of annulus at $\sqrt{d}$.

Calculate the ratio: $\dfrac{e^{\frac{(\sqrt{d}+k)^2}{2}}(\sqrt{d}+k)^d}{e^{\frac{-d}{2}}(\sqrt{d})^d}$

$$= \frac{e^{\frac{d}{2}-\sqrt{d}k-\frac{k^2}{2}}}{e^{\frac{d}{2}}}(1+\frac{k}{\sqrt{d}})^d$$

$$= e^{\frac{k^2}{2}}$$

$$(1+\frac{k}{\sqrt{d}})^d = (1+\frac{1}{\sqrt{d}/k})^{\sqrt{d}/k \cdot k\sqrt{d}} = e^{k\sqrt{d}}$$

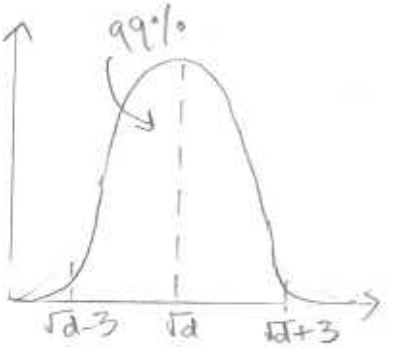


Figure:

Separating data of different processes:

We have multiple Gaussian processes various points and we need to identify which points belong to which processes.

To do this, calculate distance between each pair of points. (To calculate the mean, you need enough data points, when the space is of high dimension).

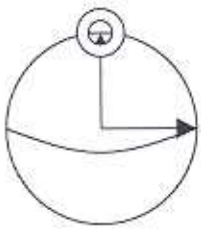
Points from Different Gaussians

Say we have a pair of points. We would like to know if they were generated by different distributions (Gaussians), and one way which we can do this is by calculating the distance between those points.

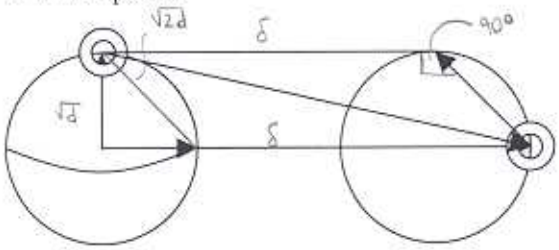
Note: In general case, we need to have enough points relative to the number of dimensions.

Assume the distribution is:
1) Gaussian
2) Spherically symmetric

We start by generating the first point and rotating the coordinates in order to put that point at the "north pole".



Since most of the area is at the equator, the second point is likely to be perpendicular to the first point:



Now for two separate spheres in high dimensions, almost all dimensions are perpendicular (as seen in the rightmost sphere above). The distance between the points in the two spheres is $\sqrt{\delta^2 + 4d}$.

In order to show that the points are in separate spheres, we want to show that $\sqrt{\delta^2 + 4d} \geq \sqrt{2d} + C$. $C \approx 6$ since the size of the annulus is 6.

$$\sqrt{2d}\left(1+\frac{\delta^2}{2d}\right)^{1/2} = \sqrt{2d}\left(1+\frac{\delta^2}{4d}+\ldots\right)^{1/2} \geq \sqrt{2d} + C$$

$$\sqrt{2d}\,\frac{\delta^2}{4d} \geq C \Rightarrow \delta^2 \geq C\frac{4d}{\sqrt{2d}} = C'\sqrt{d} \Rightarrow \delta \geq C''d^{1/4}$$

Next we examine using the SVD to find a rank l subspace that is close to containing the centers of these spheres. Using the SVD helps us by getting rid of random noise, which should help us show the axis for the sphere centers. After obtaining a rank l subspace, we then project the points onto that subspace.

In this subspace, the distances between the centers should be relatively the same due to them lying (hopefully) along the rank l approximation's axis. Distance in the space generated by the SVD is smaller by a factor of $\sqrt{d/l}$.

If 2 points are in the same Gaussian, then they are now $\sqrt{2l}$ apart. Likewise, if two points are in different Gaussians, they are $\sqrt{\delta^2 + 2l}$ apart. Thus our previous bound only depends on the number of Gaussians, which is l, not d.

Chernoff Bounds

Say we have independent random variables $x_1, x_2, \ldots, x_n$

How far can $\sum_{i=1}^{n} x_i$ deviate from $E\left(\sum_{i=1}^{n} x_i\right)$? If the variance of all x_i is small, not very far.

Let $S = \sum_{i=1}^{n} x_i$. We want to know $\Pr[S \geq (1+\delta)E(S)] \leq e^{\cdots}$ and $\Pr[S \leq (1-\delta)E(S)] \leq e^{\cdots}$.