Sheppard's Formula

BY GRAYSON YORK

(NOT FAIR GAME ON ASSESSMENTS EXCEPT AS EXTRA CREDIT, JUST USEFUL FOR OUR NEXT TOPIC)

What Is The Goal?

- We are given two vectors x,y
- We sample a random vector r, where each entry of r is drawn from a normal distribution
- What is the probability that the sign of x·r is the same as the sign of y·r?
- We will show that the following holds, where ϴ is the angle between the vectors:

$$\Pr[Sign(x \cdot r) = Sign(y \cdot r)] = 1 - \frac{\theta_{xy}}{\pi}$$

Why do we care?

- Useful for locality sensitive hashing
- Also relevant in political science (Majority is Stablest Theorem)
  - Take 2 voting populations p,q, where votes are random but have correlation ρ
  - The probability that p and q elect the same candidate under majority rule is $1 - \frac{\arccos \rho}{\pi}$
  - It turns out this is the highest probability possible for the results to align under reasonable assumptions about the voting rule

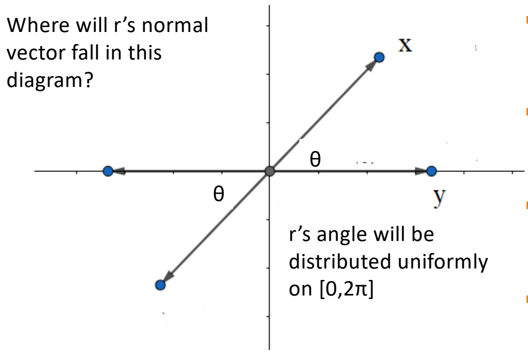
Quick Review of Dot Products

$$x \cdot y = \sum_{i=1}^{d} x_i y_i$$

$$\cos \theta_{xy} = \frac{x \cdot y}{\|x\|_2 \|y\|_2}$$

- The cosine above is positive when the angle between x and y is in the range -90 to 90 degrees (-π/2 to π/2 radians)
- Equivalently, the dot product is positive when the normal vector to one of the two vectors does not fall between the two.

When do two Random Vector Dot Products Share a sign?



- The random vector must fall within π/2 radians of both vectors, or more than π/2 radians away from both vectors
- Equivalently, the normal vector to the random vector cannot fall in the acute angle formed by the two vectors
- Because entries are gaussian, the angle of r and its normal vector will be uniformly distributed over [0, 2π]
- The probability of having the same sign is just the probability the normal vector does not fall in the acute part $1 - \frac{2\theta}{2\pi} = 1 - \frac{\theta}{\pi}$