CS 4850 prelim 1 review

High Dimensions

- Distance between random points concentrates about expected distance.
- Distance from origin to vertex of unit cube is $\frac{\sqrt{d}}{2}$.
- $$V(d) = \frac{\pi^{d/2}}{\frac{d}{2}\Gamma\left(\frac{d}{2}\right)}$$
- The gamma function (for our purposes) is defined as: $\Gamma(x) = (x-1)\Gamma(x-1)$ with base cases $\Gamma(1) = 1$ and $\Gamma(1/2) = \sqrt{\pi}$. For $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$.
- For any choice of the "North Pole", almost all of the volume and surface area of a d-dimensional sphere is near the equator. The intuitive reason for this is the following: WLOG, let the x_1 coordinate be in direction of the "North Pole" (ie, rotate so that it is). Then the larger the x_1 coordinate is, the fewer possibilities for the other coordinates.
- Almost all the volume of a sphere is located near the boundary: since the r^d term starts having a big impact, the volume of the sphere of radius $(1-\varepsilon)$ starts to have significantly smaller volume than the sphere of radius 1 as d gets sufficiently large. In particular, we get $\text{Vol}(B(0, 1-\frac{c}{d})) \leq e^{-c}\,\text{Vol}(d)$ for all c.
- To generate points uniformly at random on the sphere: One idea would be to pick the points uniformly at random on the cube, and then project onto the sphere. But then more points will be chosen by the vertices of the cube than points by the edges of the cube. And when d is sufficiently large, almost all of the volume of the cube is by the vertices, so almost all the points you generate on the sphere will be in the direction of the vertices. So that doesn't work. Another idea is to do the same thing, but throw away points that aren't in the sphere. But since as d increases the volume of the sphere goes to 0, for sufficiently large d, you will almost never find a point in the sphere this way. So you want to pick each coordinate using the Gaussian distribution. By multiplying all of these, we get that the probability distribution for a point $x_1, \ldots, x_d$ is:

  $$P(x_1, \ldots, x_d) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x_1^2 + \cdots + x_d^2}{2}}$$

  It is clear that this is spherically symmetric since the probability only depends on $x_1^2 + \cdots + x_d^2$, which is the distance of the point from the origin. So every point at the same distance from the origin has the same probability of being chosen.
- For d sufficiently large, the distance between two random points on the sphere is almost surely very close to $\sqrt{2}$. This is just a different way to say that for any "North Pole", the volume and surface area is near the equator.
- Let x be a random point on one sphere, and y a random point on another sphere. Let δ be the distance between the centers of the two spheres. Then the expected distance between x and y is $\sqrt{\delta^2 + d}$.
- Gaussian distributions are different in high dimesions. Since the volume of the sphere is almost 0 for sufficently large d, although the value of the Gaussian is maximum at the origin, there is very little mass there. You need to increase the radius of the sphere to $\sigma\sqrt{d}$ for the mass to be in the sphere. And almost none of the mass is past radius $\sigma\sqrt{d}$. Turns out that almost all of the mass of a Gaussian in d dimensions is concentrated in a thin annulus by the boundary of the sphere of width $1/\sqrt{d}$.
- To fit a Gaussian to a given set of points, pick μ to be the centroid of the points (ie, let $\mu = \frac{x_1 + \cdots + x_n}{n}$), and $\sigma = \sqrt{a/md}$ where a is the sum of the squares of the distances between the points and μ, and m is the actual mean of the Gaussian.

- Random Projection Theorem: You can project n points in d dimensions to $k \in \Omega(\log n)$ dimensions and preserve all pairwise distances between the points, up to the constant multiple of $\sqrt{k/d}$. The crux of the proof is that since the points are random, projecting to k dimensions will contract the length of the points by $\sqrt{k/d}$ on average.

Review of Statistics:

- Markov Inequality: For x a non-negative random variable:

$$\Pr[x \geq a] \leq \frac{E[x]}{a}$$

Proof:

$$E[x] = \int_0^\infty xP(x)dx = \int_0^a xP(x)dx + \int_a^\infty xP(x)dx \geq \int_a^\infty xP(x)dx \geq a\int_a^\infty P(x)dx = a\Pr[x \geq a]$$

- Variance: $\sigma^2 = E\left[|x-m|^2\right]$.

- Chebyshev's Inequality: $\Pr[|x-m| \geq a\sigma] \leq 1/a^2$. Proof:

  $\Pr[|x-m| \geq a\sigma] = \Pr\left[(x-m)^2 \geq a^2\sigma^2\right]$. From Markov's Inequality with $(x-m)^2$ as the random variable, we get:

$$\Pr\left[(x-m)^2 \geq a^2\sigma^2\right] \leq \frac{E[(x-m)^2]}{a^2\sigma^2} = \frac{\sigma^2}{a^2\sigma^2} = \frac{1}{a^2}$$

Random Graphs:

- $G(n,p)$ model: n vertices, where each each is added with probability p

- The probability that a given vertex has degree d is:

$$\binom{n}{d}p^d(1-p)^{n-d}$$

- Thresholds: If $\exists\ p(n)$ st :

$$\lim_{n\to\infty}\frac{p_1(n)}{p(n)} = 0 \quad \Rightarrow \quad G(n,p_1(n)) \text{ does not have the property}$$
$$\lim_{n\to\infty}\frac{p_2(n)}{p(n)} = \infty \quad \Rightarrow \quad G(n,p_2(n)) \text{ has the property}$$

Then $p(n)$ is the threshold for that property. If for $cp(n)$, the graph almost surely does not have the property when $c < 1$ and the graph almost surely does have the property when $c > 1$, then $p(n)$ is a sharp threshold.

- Second Moment Argument: Let X_n be a sequence of non-negative random variables converging to X (meaning if F_n and F are the cumulative distribution functions of X_n and X respectively, then $\lim_{n\to\infty} F_n(x) = F(x)$ for all x where F is continuous). If $E[X_n^2] \leq c_1 E[X_n]^2$ and $E[X_n] \geq c_2$ for all n, then for every n and a, $\Pr[X_n \geq c_2 a] \geq (1-a)^2/c_1$ (and hence, this also holds for X).

  Using a different inequality (which only holds some of the time, but will hold in all situations we will consider in this class), we know:

$$\Pr[X_n > 0] \geq \frac{E[X_n]^2}{E[X_n^2]}$$

  This was the inequality he used in class (this is why he showed that $\frac{E[x^2]}{E[x]^2} \leq 1$).

- Chart of thresholds:

$p(n)$	$G(n,p(n))$
$\frac{1}{n}$	cycles appear, and a giant component appears
$\frac{\ln n}{4n}$	consists of one giant component and a bunch of isolated vertices
$\frac{\ln n}{n}$	is connected, disapperance of isolated vertices
$\sqrt{\frac{2\ln n}{n}}$	has diamater 2

- Monotone Property: A property is monotone if adding an edge is not going to make the graph no longer have the property.
- Two important properties of monotone properties:

  a: If $p \leq q$, then $\Pr[G(n,p) \text{ has the property}] \leq \Pr[G(n,q) \text{ has the property}]$.

  b: It has a phase transition.

- Let $p(n,\varepsilon)$ be the function of n such that the probability that $G(n,p)$ has some property is ε. To simplify notation, let $p(\varepsilon)$ denote the same thing. Thenit can be shown that there $\exists\ m$ st $p(1-\varepsilon) \leq mp(\varepsilon)$. That implies that $p(1/2)$ is a threshold for the property.
- Other structures also have phase transitions. For example, $N(n,p)$, the subset of $\{1,\ldots,n\}$ where each number is added with probablility p. Also, Sudoku puzzles if you take a filled in grid and randomly remove numbers, there is a threshold at which the puzzle will become "easy". Also, CNF: When the number of clauses reaches some threshold in terms of the number of literals per clause k, there are very few satisfiable statements. However, that threshold is not known. We know that $2^k \ln 2$ is an upper bound on the threshold, and $2^k/k$ is a lower bound.
- Other models of random graphs:
  - "Growth Model without preferential attachment": Start with zero vertices at time 0. At each unit of time, a new vertex is created, and with probability δ, two vertices are chosen at random and connected by an edge.
  - "Growth Model with preferential attachment": At each unit of time, a new vertex is created, and with probability δ the new vertex is attached to another vertex chosen at random with probability proportional to their degrees.

    Let $d_i(t)$ be the degree of the i^{th} vertex at time t. Then $E\left[\sum d_i(t)\right] = 2\delta t$. So the probability of picking vertex i in time step t is $\frac{d_i(t)}{2\delta t}$. Solving a differential equation governing $d_i(t)$, we get $E[d_i(t)] = \delta\sqrt{t/i}$. Then we get:

$$\Pr[\text{degree } < d] = 1 - \frac{\delta^2}{d^2}$$
$$\Pr[\text{degree } = d] = 2\frac{\delta^2}{d^3}$$