1.5 Gaussian

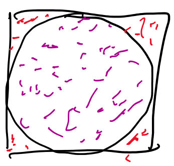
Wednesday, January 8, 2020 1:42 PM

Outline:

- Generating random points on a sphere
- Review of Gaussian r.v.
- Gaussian annulus thm
- Random projection + Johnson-Lindenstrauss Lemma
- Applications of Gaussians

Generating random points on sphere/ball

Last time, we ended with generating random points on a hypersphere.



In low dimensions, rejection sampling.

In high dimensions, this fails because the volume of the unit ball goes to 0.

Instead, recall the pdf of a 1D Gaussian $p(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$

Then, the pdf of a vector $\vec{x} = [x_1, \dots, x_d]$, where x_i's are ind. Gaussians, is

$$p(\vec{x}) = \frac{1}{(2\pi)^{d/2}} \exp\Bigg(\underbrace{-\frac{(x_1^2 + \cdots + x_d)^2}{2}}_{\text{spherically symmetric}}\Bigg).$$

Then for a hypersphere, normalize to unit length $\frac{\vec{x}}{|\vec{x}|}$.

For a unit ball, pick a radius ρ with density $\rho(r) = dr^{d-1}$ for r over $[0,1]$.

Because vol of radius r ball is $r^d V(d)$, where $V(d)$ is volume of d-dim ball

density at radius r is $\frac{d}{dr}\left(r^d V(d)\right) = d r^{d-1} V(d)$.

Alternately, know that density is proportional to r^{d-1}, then solve for

$$\int_{r=0}^{r=1} c r^{d-1} dr = 1$$

Then take $\vec{y} = \rho \frac{\vec{x}}{|x|}$ to get a

$$\int_{r=0}^{1} c\,r^{d-1}\,dr = 1$$

$$\frac{c}{d} r^d \Big|_0^1 = 1$$

$$\frac{c}{d} = 1 \Rightarrow c = d.$$

Then take $\vec{y} = \rho \frac{\vec{x}}{|x|}$ to get a random point in the ball.

Box-Muller transform: $\theta = U_1 \cdot 2\pi$

$r = \sqrt{-2\ln|U_2|}$, U_1, U_2 are uniform r.v. in $[0,1]$.

Gaussian Annulus Thm: For a d-dim. spherical Gaussian with unit variance in each direction, for any $\beta \le \sqrt{d}$, all but at most $3e^{-c\beta^2}$ of the probability mass lies within the annulus $\sqrt{d} - \beta \le |\vec{x}| \le \sqrt{d} + \beta$, where c is a fixed positive constant.

Note: we can get a much ~~weaker result~~ similar result using LLN.

$$\mathbb{E}(|\vec{x}|^2) = \sum_{i=1}^{d} \mathbb{E}(x_i^2) = d\,\mathbb{E}(x_1^2) = d.$$

Terminology: we call $\sqrt{d}$ the radius of the Gaussian.

Proof. Let $\vec{x} = (x_1, \dots, x_d)$ be a point selected from a unit variance Gaussian centered at the origin. Let $r = |\vec{x}|$.

WTS $\quad \text{Prob}(|r - \sqrt{d}| \ge \beta) \le 3e^{-c\beta^2}$.

Note that if $|r - \sqrt{d}| \ge \beta$, then $|r^2 - d| \ge \beta(r + \sqrt{d}) \ge \beta\sqrt{d}$.

So we only need to bound the probability that $|r^2 - d| \ge \beta\sqrt{d}$

(because $\text{Prob}(|r^2 - d| \ge \beta\sqrt{d}) \ge \text{Prob}(|r^2 - d| \ge \beta(r + \sqrt{d}))$)

$$r^2 - d = (x_1^2 + \dots + x_d^2) - d = (x_1^2 - 1) + \dots + (x_d^2 - 1).$$

Let $y_i = x_i^2 - 1$.

$$\mathbb{E}\,y_i = \mathbb{E}\,x_i^2 - 1 = 0.$$

Let's compute the moments of y_i.

Typo in book?

$$|\mathbb{E}(y_i^s)| \le \mathbb{E}(|y_i^s|) = \mathbb{E}|y_i|^s$$

If $|y_i| \le 1$, $|y_i|^s \le 1$

$$|\mathbb{E}(y_i^{\,s})| \le \mathbb{E}(|y_i^{\,s}|) = \mathbb{E}|y_i|^s$$

$$\le \mathbb{E}(1 + x_i^{2s})$$

If $|x_i| \le 1$, $|y_i|^s \le 1$
If $|x_i| \ge 1$, $|y_i|^s \le |x_i|^{2s}$

$$= 1 + \underbrace{\mathbb{E}(x_i^{2s})}_{\text{even moments of Gaussian}}$$

$$= 1 + (2s-1)!!$$

(double factorial, prod. of all integers from 1 to $2s-1$ that have the same parity)

$$\le 2^s s!$$

$$\Rightarrow |\mathbb{E}(y_i^{\,2})| \le 8 \qquad (\mathrm{var}(y_i))$$

(**Recall Master Tail Bounds Thm:** Let $x = x_1 + \cdots + x_n$, where x_i's have 0 mean and variance at most σ^2. Let $0 \le a \le \sqrt{2}\, n\sigma^2$. If $|\mathbb{E}(x_i^{\,s})| \le \sigma^2 s!$ for $s = 3, 4, \ldots, \lfloor (a^2/4n\sigma^2) \rfloor$, Then $\mathrm{Prob}(|x| \ge a) \le 3e^{-a^2/(12n\sigma^2)}$)

Let $w_i = \frac{y_i}{2}$. Then $\mathrm{var}(w_i) \le 2$ and $|\mathbb{E}(w_i^{\,s})| \le 2s!$

Thus, we can apply the Master Tail Bounds Thm with $a = \frac{\beta\sqrt{d}}{2}$, $\sigma^2 = 2$, $n = d$

$$\Rightarrow \mathrm{Prob}(|r - \sqrt{d}| \ge \beta) \le \mathrm{Prob}\left(|w_1 + \cdots + w_d| \ge \frac{\beta\sqrt{d}}{2}\right) \le \underbrace{3e^{-\beta^2/96}}_{\text{Can get a tighter constant bound, but no need.}}.$$

Random Projection & Johnson-Lindenstrauss Lemma

Let $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$.

For $\vec{v} \in \mathbb{R}^d$, want to find $\operatorname{argmin}_i |\vec{x}_i - \vec{v}|$.

We are allowed to preprocess a database of x_i's, but ideally want a query time that is a small function of $\log n$ and $\log d$.

Thm 2.10 (Random Projection)

Consider a matrix

$$A = \begin{bmatrix} u_{11} & \cdots & u_{1d} \\ \vdots & \ddots & \vdots \\ u_{k1} & \cdots & u_{kd} \end{bmatrix},$$

where each $u_{i,j}$ are randomly drawn from $N(0,1)$.

$[\ldots]_{k \times d}$ from $N(0,1)$.

Let $\vec{v} \in \mathbb{R}^d$. Then $\exists\, c > 0$ s.t. for $\varepsilon \in (0,1)$,

$$\text{Prob}\left(\Big|\, |A\vec{v}| - \sqrt{k}\,|\vec{v}| \,\Big| \geq \varepsilon \sqrt{k}\,|\vec{v}|\right) \leq 3e^{-ck\varepsilon^2}.$$

proof. WLOG, assume $|\vec{v}| = 1$, Say $\vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix}$

$$A\vec{v} = \begin{bmatrix} u_{11}v_1 + \cdots + u_{1d}v_d \\ \vdots \\ u_{k1}v_1 + \cdots + u_{kd}v_d \end{bmatrix}.$$

Let $w_i = u_{i1}v_1 + \cdots + u_{id}v_d$, a r.v.

Since u_{ij} are $N(0,1)$, $\mathbb{E}\, w_i = 0$

$$\text{Var}(w_i) = \sum_{j=1}^{d} v_j^2 \,\text{Var}(u_{ij}) = \sum_{j=1}^{d} v_j^2 = 1.$$

Thus, $w_i \sim N(0,1)$.

$\Rightarrow$ $A\vec{v}$ is a random vector from a k-dim spherical Gaussian with unit variance in each coordinate.

Then apply Gaussian Annulus Thm with d replaced by k. $\blacksquare$

Thm J-L For any $0 < \varepsilon < 1$, $n \in \mathbb{N}$, let $k \geq \frac{3}{c\varepsilon^2} \ln n$, with c as in Gaussian Annulus Thm.

For any set of n points in $\mathbb{R}^d$, the random projection $f(\vec{v}) : \mathbb{R}^d \to \mathbb{R}^k$ defined by $A\vec{v}$ has the prop. that for all pairs of points $\vec{v_i}$ and $\vec{v_j}$, with prob. at least $1 - \frac{3}{2n}$,

$$(1-\varepsilon)\sqrt{k}\,|\vec{v_i} - \vec{v_j}| \leq |f(\vec{v_i}) - f(\vec{v_j})| \leq (1+\varepsilon)\sqrt{k}\,|\vec{v_i} - \vec{v_j}|.$$

Proof. Apply a union bound for every pair of points, after using the random projection thm.

If $k \geq \frac{3 \ln n}{c\varepsilon^2}$, then $3e^{-ck\varepsilon^2} \leq \frac{3}{n^3}$.

There are $\binom{n}{2} < n^2/2$ pairs of points, so the prob. that any pair of points has large distortion is $< \frac{3}{2n}$ $\blacksquare$

What is the complexity?

What is the complexity?

$k = O\left(\frac{1}{\varepsilon^2}\log n\right)$, so we have a dimensionality reduction.

Pairwise comparisons in $\mathbb{R}^d$ take $O(d)$ time to compute distance.

In $\mathbb{R}^k$, take $O(k)$ time.

So to compare against all n points $O(dn) \to O\left(\frac{1}{\varepsilon^2} n \log n\right)$.

However, the projection $A\vec{v}$ takes $O(kd)$ time per point, requiring $O(kdn)$ preprocessing.

Luckily, [Kane, Nelson, 2010]

[Dasgupta-Kumar-Sarlós, 2010]

⋮

allow more computationally efficient choices of A, such as sparse Bernoulli $\left(-\frac{1}{2}, +\frac{1}{2}\right)$ matrices.

Next time:

- SVD