Johnson-Lindenstrauss Lemma

Lemma 1: Let $x_{1:n} \in \mathbb{R}^d$. Fix $\varepsilon, \delta \in (0,1)$,

$\ell \geq \frac{c_0}{\varepsilon^2} \log\left(\frac{c_1 n}{\delta}\right)$. Let $A \in \mathbb{R}^{\ell \times d}$ have i.i.d.

$\mathcal{N}(0, \frac{1}{\ell})$ entries. Then w.p. $\geq 1-\delta$

$$(1-\varepsilon)\|x_i - x_j\|^2 \leq \|Ax_i - Ax_j\|^2$$
$$\leq (1+\varepsilon)\|x_i - x_j\|^2 \quad \forall i,j \in [n]$$

A more general formulation of JL Lemma

Let $x_{1:n} \in \mathbb{R}^d$. Fix $\varepsilon \in (0,1)$ and

put $\ell \geq \frac{c_0}{\varepsilon^2} \log(c_1 n)$.

Then there exists a linear map $A \in \mathbb{R}^{\ell \times d}$ be s.t.

$$(1-\varepsilon)\|x_i - x_j\|^2 \leq \|Ax_i - Ax_j\|^2$$
$$\leq (1+\varepsilon)\|x_i - x_j\|^2$$

$\forall i,j \in [n]$

$\Rightarrow$ Gaussian Random Projection preserves distances with high probability.

$\Rightarrow$ But also very costly for large d.

Other Projections:

① Sparse JL Transform (Achlioptas, 2003)

Put $l \geq \dfrac{c_0 \log \frac{c_1 n}{\delta}}{\varepsilon^2}$

Let A_{ij} be i.i.d. r.v. from either one of the following two probability distributions:

$$A_{ij} = \begin{cases} \frac{1}{\sqrt{l}} & \text{w.p. } \frac{1}{2} \\ -\frac{1}{\sqrt{l}} & \text{w.p. } \frac{1}{2} \end{cases}$$

$$A_{ij} = \sqrt{3} \times \begin{cases} \frac{1}{\sqrt{l}} & \text{w.p. } \frac{1}{6} \\ 0 & \text{w.p. } \frac{2}{3} \rightarrow \text{sparsity} \\ -\frac{1}{\sqrt{l}} & \text{w.p. } \frac{1}{6} \end{cases}$$

Then w.p. at least $1-\delta$, $\forall i, j \in [n]$

$$(1-\varepsilon)\|x_i - x_j\|^2 \leq \|Ax_i - Ax_j\|^2 \leq (1+\varepsilon)\|x_i - x_j\|^2$$

② Fast JL Transform (FJLT)

Def: A Hadamard matrix is a $d \times d$ real matrix H that is orthogonal ($H^T H = I$) and has entries in $\{\pm \frac{1}{\sqrt{d}}\}$.

E.g.: $d=2$

$$H_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Nice property of Hadamard matrix:

matrix-vector multiplication can be performed in $O(d \log d)$ time

Linear map $A = \sqrt{\frac{d}{\ell}}\, PHD$

Compare with the Gaussian embedding $O(d\ell)$

D: random diagonal ± 1 matrix

H: the Hadamard matrix

P: samples ℓ coordinates

PCA

Proposition: $\Sigma = E[xx^T]$ → Second moment
not necessarily a covariance

$v_1, \cdots, v_k$ → top k eigenvalues of Σ

$\mathrm{Span}\{v_1, \cdots, v_k\} \in \underset{\dim(W)=k}{\arg\min} E\|P_W^{\perp} x\|^2$

$\mu = E[x]$

$\in \underset{P \in O(d,k)}{\arg\min} \underbrace{E\|(I-PP^T)x\|^2}$

$$\begin{aligned}\underset{P}{\arg\min} E\|(I-PP^T)x\|^2 &= E\|(I-PP^T)(x-\mu) + (I-PP^T)\mu\|^2 \\ &= E\|(I-PP^T)(x-\mu)\|^2 + E\|(I-PP^T)\mu\|^2\end{aligned}$$

= Reconstruction error of centered data

+ positive constant

$$\begin{aligned}\Sigma = E[xx^T] &= E[(x-\mu)(x-\mu)^T] + \mu\mu^T \\ &= \underbrace{Cov(x)}_{\Sigma_c} + \mu\mu^T\end{aligned}$$

Example: $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

with probability $\frac{1}{2}$

$x_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$$\mu = \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}$$

Centered data → $Cov(x) = \begin{bmatrix} 0.25 & 0 \\ 0 & 0 \end{bmatrix}$ → top eigenvector $= \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

Second Moment

$\Sigma = E[xx^T] = \begin{bmatrix} 2.5 & 1.5 \\ 1.5 & 1 \end{bmatrix}$ → top eigenvector

Uncented data

second moment

$= \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}$

$= \mu$

So uncentered PCA "bends" toward the mean (μ) direction

MLE Refresher

Let $x_{1:n}$ be i.i.d. samples from a normal distribution $x_i \sim N(\mu, \sigma^2)$ where both $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown.

Find MLE for μ & σ^2.

Likelihood Function:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

log-likelihood:

$$\ell(\mu, \sigma^2) = \log L(\mu, \sigma^2)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Maximize over μ:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\Rightarrow \hat{\mu}^{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Maximize over σ^2:

Substitute $\hat{\mu}^{MLE} = \bar{x}$ into ℓ

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0$$

$$\Rightarrow \hat{\sigma}^{2\,MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$\frac{1}{n-1}$

Not $n-1$, so $\hat{\sigma}^{2\,MLE}$ is not unbiased.

$$\Rightarrow \begin{cases} \hat{\mu}^{MLE} = \bar{x} \\ \hat{\sigma}^{2\,MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{cases}$$