Randomness

DS-GA 1013 / MATH-GA 2824 Optimization-based Data Analysis

http://www.cims.nyu.edu/~cfgranda/pages/OBDA_fall17/index.html

Carlos Fernandez-Granda

Gaussian random variables

Gaussian random vectors

Randomized projections

SVD of a random matrix

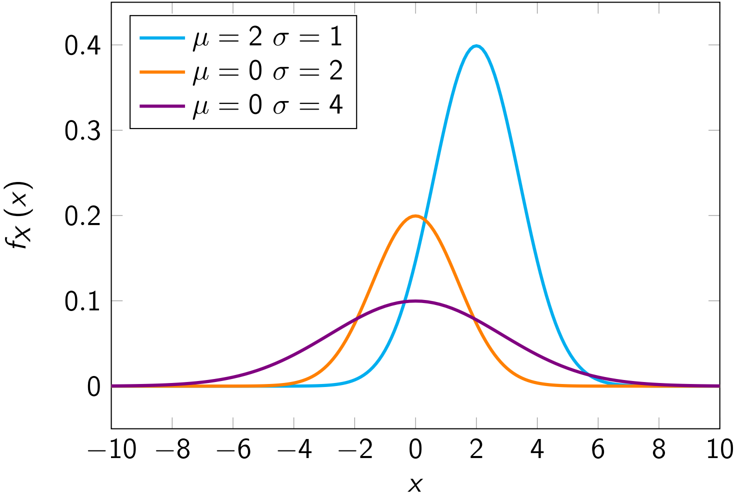
Randomized SVD

Gaussian random variables

The pdf of a Gaussian or normal random variable with mean μ and standard deviation σ is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Gaussian random variables



Linear transformation of Gaussian

If $\mathbf{x}$ is a Gaussian random variable with mean μ and standard deviation σ, then for any $a, b \in \mathbb{R}$

$$\mathbf{y} := a\mathbf{x} + b$$

is a Gaussian random variable with mean $a\mu + b$ and standard deviation $|a| \sigma$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$F_{\mathbf{y}}(y)$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$F_{\mathbf{y}}(y) = \mathrm{P}(\mathbf{y} \leq y)$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$
\begin{aligned}
F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\
&= \mathrm{P}(a\mathbf{x} + b \leq y)
\end{aligned}
$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$\begin{aligned} F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\ &= \mathrm{P}(a\mathbf{x} + b \leq y) \\ &= \mathrm{P}\left(\mathbf{x} \leq \frac{y - b}{a}\right) \end{aligned}$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$
\begin{aligned}
F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\
&= \mathrm{P}(a\mathbf{x} + b \leq y) \\
&= \mathrm{P}\left(\mathbf{x} \leq \frac{y-b}{a}\right) \\
&= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, \mathrm{d}x
\end{aligned}
$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$\begin{aligned} F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\ &= \mathrm{P}(a\mathbf{x} + b \leq y) \\ &= \mathrm{P}\left(\mathbf{x} \leq \frac{y-b}{a}\right) \\ &= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \,\mathrm{d}x \\ &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{(w-a\mu-b)^2}{2a^2\sigma^2}} \,\mathrm{d}w \qquad \text{change of variables } w = ax + b \end{aligned}$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$\begin{aligned} F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\ &= \mathrm{P}(a\mathbf{x} + b \leq y) \\ &= \mathrm{P}\left(\mathbf{x} \leq \frac{y-b}{a}\right) \\ &= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x \\ &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{(w-a\mu-b)^2}{2a^2\sigma^2}} \, \mathrm{d}w \qquad \text{change of variables } w = ax + b \end{aligned}$$

Differentiating with respect to y:

$$f_{\mathbf{y}}(y) = \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{(w-a\mu-b)^2}{2a^2\sigma^2}}$$

Central limit theorem

Let $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, ... be a sequence of iid random variables with mean μ and bounded variance σ^2

The sequence of averages $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, ... is defined as

$$\mathbf{a}_i := \frac{1}{i} \sum_{j=1}^{i} \mathbf{x}_j$$

Central limit theorem

The sequence $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$, ...

$$\mathbf{b}_i := \sqrt{i}(\mathbf{a}_i - \mu)$$

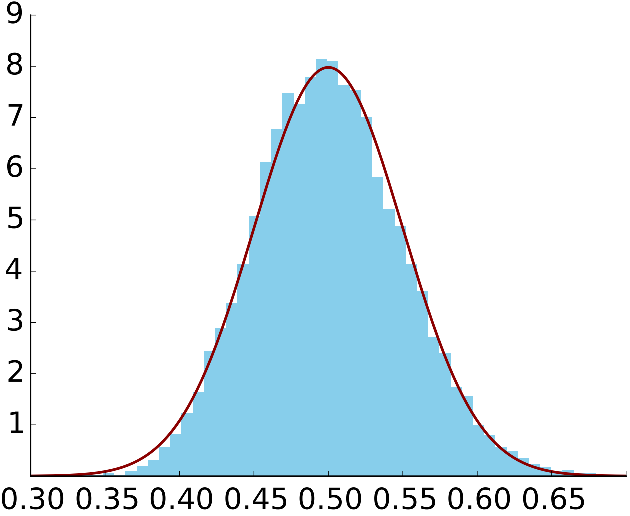
converges in distribution to a Gaussian random variable with mean 0 and variance σ^2

For any $x \in \mathbb{R}$

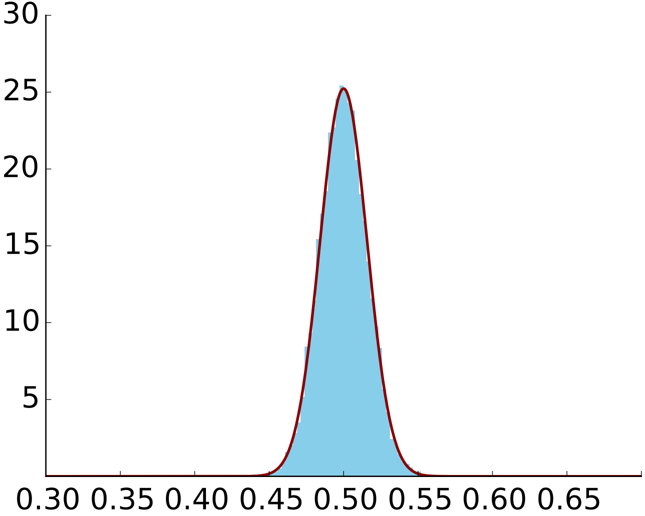
$$\lim_{i \to \infty} f_{\mathbf{b}_i}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

For large i the theorem suggests that the average $\mathbf{a}_i$ is approximately Gaussian with mean μ and variance $\sigma/\sqrt{n}$

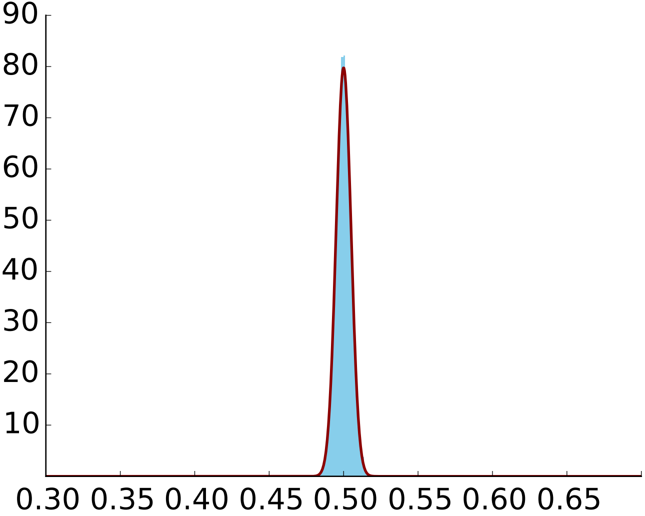
iid exponential $\lambda = 2$, $i = 10^2$



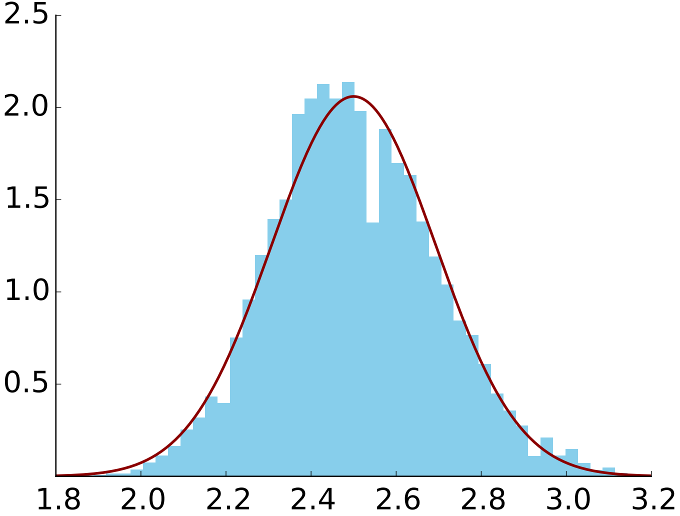
iid exponential $\lambda = 2$, $i = 10^3$



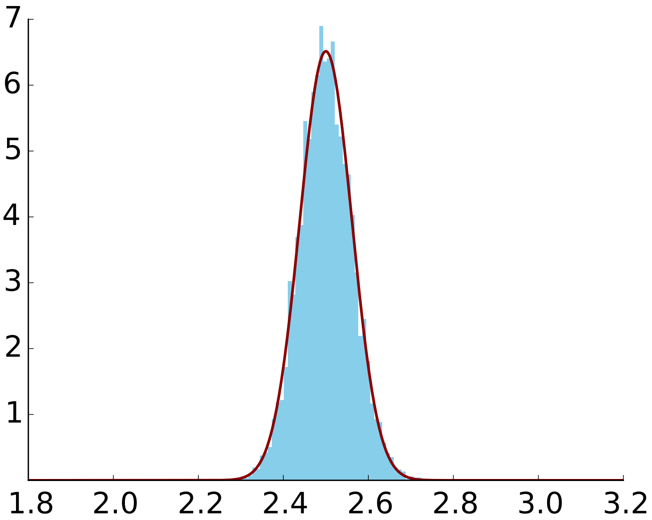
iid exponential $\lambda = 2$, $i = 10^4$



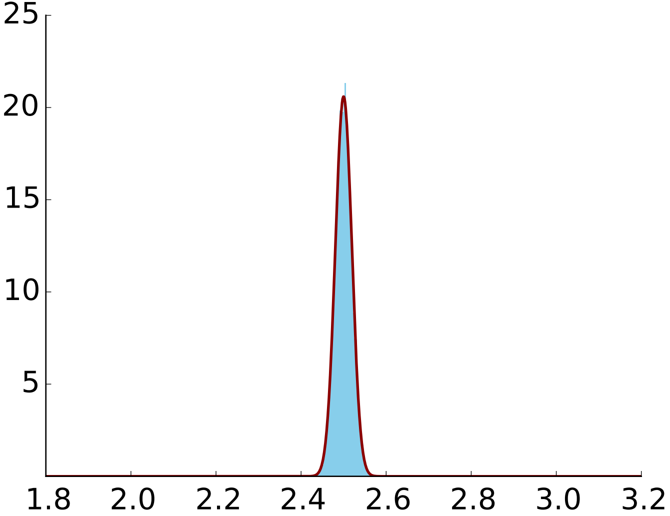
iid geometric $p = 0.4$, $i = 10^2$



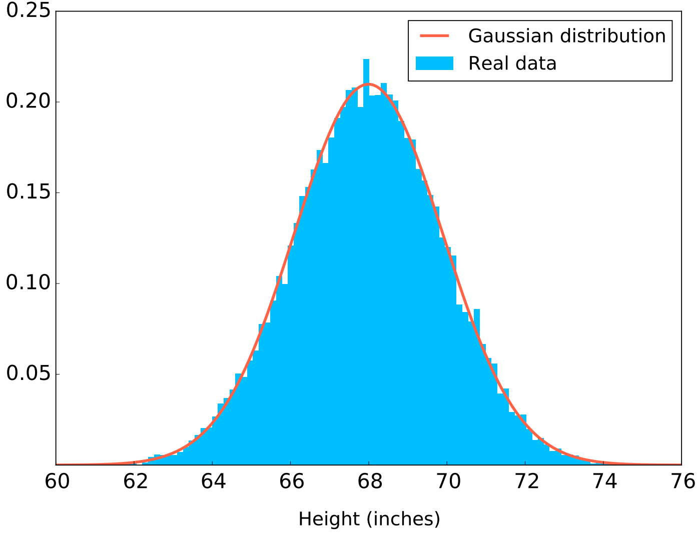
iid geometric $p = 0.4$, $i = 10^3$



iid geometric $p = 0.4$, $i = 10^4$



Histogram of heights



Gaussian random variables

Gaussian random vectors

Randomized projections

SVD of a random matrix

Randomized SVD

Gaussian random vector

A Gaussian random vector $\vec{\mathbf{x}}$ is a random vector with joint pdf

$$f_{\vec{\mathbf{x}}}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

where $\vec{\mu} \in \mathbb{R}^n$ is the mean and $\Sigma \in \mathbb{R}^{n \times n}$ the covariance matrix

Uncorrelation implies independence

If the covariance matrix is diagonal,

$$\Sigma_{\vec{x}} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix},$$

the entries are independent

Proof

$$\Sigma_{\vec{\mathbf{x}}}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix}$$

$$|\Sigma| = \prod_{i=1}^{n} \sigma_i^2$$

Proof

$f_{\vec{\mathbf{x}}}\left(\vec{x}\right)$

Proof

$$f_{\vec{\mathbf{x}}}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

Proof

$$
\begin{aligned}
f_{\vec{\mathbf{x}}}\left(\vec{x}\right) &= \frac{1}{\sqrt{\left(2\pi\right)^{n}\left|\Sigma\right|}} \exp\left(-\frac{1}{2}\left(\vec{x}-\vec{\mu}\right)^{T}\Sigma^{-1}\left(\vec{x}-\vec{\mu}\right)\right) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)}\sigma_{i}} \exp\left(-\frac{\left(\vec{x}_{i}-\mu_{i}\right)^{2}}{2\sigma_{i}^{2}}\right)
\end{aligned}
$$

Proof

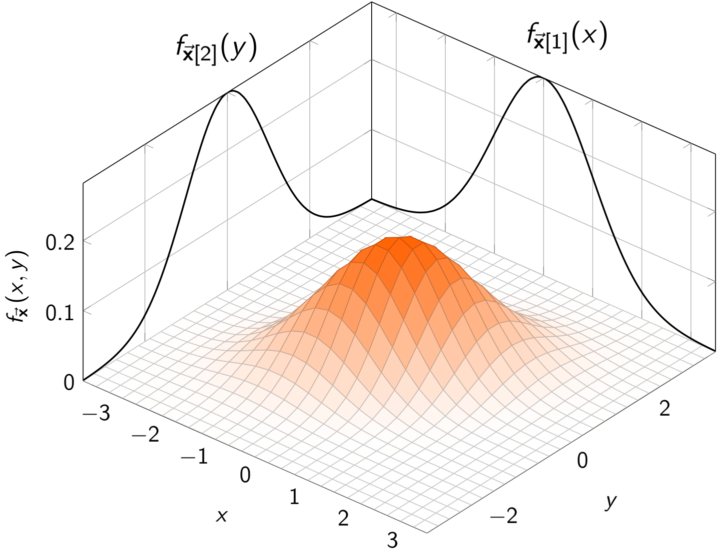
$$\begin{aligned} f_{\vec{\mathbf{x}}}\left(\vec{x}\right) &= \frac{1}{\sqrt{\left(2\pi\right)^{n}\left|\Sigma\right|}} \exp\left(-\frac{1}{2}\left(\vec{x}-\vec{\mu}\right)^{T}\Sigma^{-1}\left(\vec{x}-\vec{\mu}\right)\right) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)}\sigma_i} \exp\left(-\frac{\left(\vec{x}_i-\mu_i\right)^2}{2\sigma_i^2}\right) \\ &= \prod_{i=1}^{n} f_{\vec{\mathbf{x}}_i}\left(\vec{x}_i\right) \end{aligned}$$

Linear transformations

Let $\vec{x}$ be a Gaussian random vector of dimension n with mean $\vec{\mu}$ and covariance matrix Σ

For any matrix $A \in \mathbb{R}^{m \times n}$ and $\vec{b} \in \mathbb{R}^m$ $\vec{Y} = A\vec{x} + \vec{b}$ is Gaussian with mean $A\vec{\mu} + \vec{b}$ and covariance matrix $A \Sigma A^T$

Subvectors are also Gaussian



Direction of iid standard Gaussian vectors

If the covariance matrix of a Gaussian vector $\vec{x}$ is I, then $\vec{x}$ is isotropic

It does not favor any direction

For any orthogonal matrix $U\vec{x}$ has the same distribution
(Gaussian with mean $U\vec{0} = \vec{0}$ and covariance matrix $UIU^T = UU^T = I$)

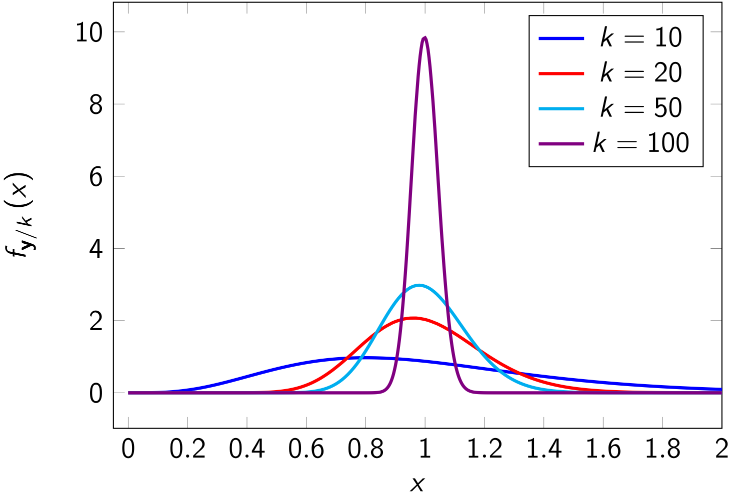
Magnitude of iid standard Gaussian vectors

In low dimensions joint pdf is mostly concentrated around the origin

High dimensions?

$||\vec{x}||_2^2 = \sum_{i=1}^{k} \vec{x}[i]^2$ is a χ^2 (chi squared) random variable with k degrees of freedom

Magnitude of iid standard Gaussian vectors



Mean

$$\mathrm{E}\left(||\vec{\mathbf{x}}||_2^2\right)$$

Mean

$$\mathrm{E}\left(||\vec{x}||_2^2\right) = \mathrm{E}\left(\sum_{i=1}^{k} \vec{x}[i]^2\right)$$

Mean

$$\mathrm{E}\left(||\vec{x}||_2^2\right) = \mathrm{E}\left(\sum_{i=1}^{k} \vec{x}[i]^2\right)$$

$$= \sum_{i=1}^{k} \mathrm{E}\left(\vec{x}[i]^2\right)$$

Mean

$$\begin{aligned} \mathrm{E}\left(\|\vec{x}\|_2^2\right) &= \mathrm{E}\left(\sum_{i=1}^{k} \vec{x}[i]^2\right) \\ &= \sum_{i=1}^{k} \mathrm{E}\left(\vec{x}[i]^2\right) \\ &= k \end{aligned}$$

Variance

$$\mathrm{E}\left(\left(||\vec{\mathbf{x}}||_2^2\right)^2\right)$$

Variance

$$\mathrm{E}\left(\left(||\vec{\mathsf{x}}||_2^2\right)^2\right) = \mathrm{E}\left(\left(\sum_{i=1}^{k} \vec{\mathsf{x}}[i]^2\right)^2\right)$$

Variance

$$
\begin{aligned}
\mathrm{E}\left(\left(||\vec{x}||_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{k} \vec{x}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{k}\sum_{j=1}^{k} \vec{x}[i]^2 \vec{x}[j]^2\right)
\end{aligned}
$$

Variance

$$
\begin{aligned}
\mathrm{E}\left(\left(||\vec{\mathsf{x}}||_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{k} \vec{\mathsf{x}}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{k}\sum_{j=1}^{k} \vec{\mathsf{x}}[i]^2 \vec{\mathsf{x}}[j]^2\right) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{E}\left(\vec{\mathsf{x}}[i]^2 \vec{\mathsf{x}}[j]^2\right)
\end{aligned}
$$

Variance

$$\begin{aligned}
\mathrm{E}\left(\left(||\vec{x}||_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{k} \vec{x}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{k}\sum_{j=1}^{k} \vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{E}\left(\vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{k} \mathrm{E}\left(\vec{x}[i]^4\right) + 2\sum_{i=1}^{k-1}\sum_{j=i}^{k} \mathrm{E}\left(\vec{x}[i]^2\right) \mathrm{E}\left(\vec{x}[j]^2\right)
\end{aligned}$$

Variance

$$
\begin{aligned}
\mathrm{E}\left(\left(||\vec{x}||_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{k} \vec{x}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{k}\sum_{j=1}^{k} \vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{E}\left(\vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{k} \mathrm{E}\left(\vec{x}[i]^4\right) + 2\sum_{i=1}^{k-1}\sum_{j=i}^{k} \mathrm{E}\left(\vec{x}[i]^2\right)\mathrm{E}\left(\vec{x}[j]^2\right) \\
&= 3k + k(k-1) \quad \text{4th moment of standard Gaussian equals 3}
\end{aligned}
$$

Variance

$$
\begin{aligned}
\mathrm{E}\left(\left(||\vec{x}||_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{k} \vec{x}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{k} \sum_{j=1}^{k} \vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} \mathrm{E}\left(\vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{k} \mathrm{E}\left(\vec{x}[i]^4\right) + 2 \sum_{i=1}^{k-1} \sum_{j=i}^{k} \mathrm{E}\left(\vec{x}[i]^2\right) \mathrm{E}\left(\vec{x}[j]^2\right) \\
&= 3k + k(k-1) \quad \text{4th moment of standard Gaussian equals 3} \\
&= k(k+2)
\end{aligned}
$$

Variance

$$\mathrm{Var}\left(||\vec{\mathbf{x}}||_2^2\right) = \mathrm{E}\left(\left(||\vec{\mathbf{x}}||_2^2\right)^2\right) - \mathrm{E}\left(||\vec{\mathbf{x}}||_2^2\right)^2$$
$$= k(k+2) - k^2 = 2k$$

Relative standard deviation around mean scales as $\sqrt{2/k}$

Non-asymptotic tail bound

Let $\vec{x}$ be an iid standard Gaussian random vector of dimension k

For any $\epsilon > 0$

$$P\left(k\left(1-\epsilon\right) < ||\vec{x}||_2^2 < k\left(1+\epsilon\right)\right) \geq 1 - \frac{2}{k\epsilon^2}$$

Markov's inequality

Let x be a nonnegative random variable

For any positive constant $a > 0$,

$$\mathrm{P}(\mathsf{x} \geq a) \leq \frac{\mathrm{E}(\mathsf{x})}{a}$$

Proof

Define the indicator variable $1_{\mathsf{x} \geq a}$

$$\mathsf{x} - a\,1_{\mathsf{x} \geq a} \geq 0$$

Proof

Define the indicator variable $1_{x \geq a}$

$$x - a\, 1_{x \geq a} \geq 0$$

$$\mathrm{E}(x) \geq a\, \mathrm{E}(1_{x \geq a}) = a\, \mathrm{P}(x \geq a)$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$\mathrm{P}\left(|\mathbf{y} - k| \geq k\epsilon\right)$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$\mathrm{P}\left(|\mathbf{y} - k| \geq k\epsilon\right) = \mathrm{P}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2 \geq k^2\epsilon^2\right)$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$\begin{aligned}
\mathrm{P}\left(|\mathbf{y} - k| \geq k\epsilon\right) &= \mathrm{P}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2 \geq k^2\epsilon^2\right) \\
&\leq \frac{\mathrm{E}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2\right)}{k^2\epsilon^2} \qquad \text{by Markov's inequality}
\end{aligned}$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$\begin{aligned}
\mathrm{P}\left(|\mathbf{y} - k| \geq k\epsilon\right) &= \mathrm{P}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2 \geq k^2\epsilon^2\right) \\
&\leq \frac{\mathrm{E}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2\right)}{k^2\epsilon^2} \qquad \text{by Markov's inequality} \\
&= \frac{\mathrm{Var}\left(\mathbf{y}\right)}{k^2\epsilon^2}
\end{aligned}$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$\begin{aligned} \mathrm{P}\left(|\mathbf{y}-k| \geq k\epsilon\right) &= \mathrm{P}\left(\left(\mathbf{y}-\mathrm{E}\left(\mathbf{y}\right)\right)^2 \geq k^2\epsilon^2\right) \\ &\leq \frac{\mathrm{E}\left(\left(\mathbf{y}-\mathrm{E}\left(\mathbf{y}\right)\right)^2\right)}{k^2\epsilon^2} \qquad \text{by Markov's inequality} \\ &= \frac{\mathrm{Var}\left(\mathbf{y}\right)}{k^2\epsilon^2} \\ &= \frac{2}{k\epsilon^2} \end{aligned}$$

Non-asymptotic Chernoff tail bound

Let $\vec{x}$ be an iid standard Gaussian random vector of dimension k

For any $\epsilon > 0$

$$P\left(k\left(1-\epsilon\right) < ||\vec{x}||_2^2 < k\left(1+\epsilon\right)\right) \geq 1 - 2\exp\left(-\frac{k\epsilon^2}{8}\right)$$

Proof

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$. The result is implied by

$$P(\mathbf{y} > k(1+\epsilon)) \leq \exp\left(-\frac{k\epsilon^2}{8}\right)$$

$$P(\mathbf{y} < k(1-\epsilon)) \leq \exp\left(-\frac{k\epsilon^2}{8}\right)$$

Proof

Fix $t > 0$

$P(\mathbf{y} > a)$

Proof

Fix $t > 0$

$$P(\mathbf{y} > a) = P(\exp(t\mathbf{y}) > \exp(at))$$

Proof

Fix $t > 0$

$$
\begin{aligned}
P(\mathbf{y} > a) &= P(\exp(t\mathbf{y}) > \exp(at)) \\
&\leq \exp(-at)\,\mathrm{E}(\exp(t\mathbf{y})) \qquad \text{by Markov's inequality}
\end{aligned}
$$

Proof

Fix $t > 0$

$$\begin{aligned} P(\mathbf{y} > a) &= P(\exp(t\mathbf{y}) > \exp(at)) \\ &\leq \exp(-at)\, \mathrm{E}(\exp(t\mathbf{y})) \qquad \text{by Markov's inequality} \\ &\leq \exp(-at)\, \mathrm{E}\left(\exp\left(\sum_{i=1}^{k} t\mathbf{x_i}^2\right)\right) \end{aligned}$$

Proof

Fix $t > 0$

$$\begin{aligned}
P(\mathbf{y} > a) &= P(\exp(t\mathbf{y}) > \exp(at)) \\
&\leq \exp(-at)\, \mathrm{E}(\exp(t\mathbf{y})) \qquad \text{by Markov's inequality} \\
&\leq \exp(-at)\, \mathrm{E}\left(\exp\left(\sum_{i=1}^{k} t\mathbf{x_i}^2\right)\right) \\
&\leq \exp(-at) \prod_{i=1}^{k} \mathrm{E}\left(\exp\left(t\mathbf{x_i}^2\right)\right) \quad \text{by independence of } \mathbf{x_1}, \ldots, \mathbf{x_k}
\end{aligned}$$

Proof

Lemma (by direct integration)

$$\mathrm{E}\left(\exp\left(t\mathsf{x}^2\right)\right) = \frac{1}{\sqrt{1-2t}}$$

Equivalent to controlling higher-order moments since

$$\begin{aligned}\mathrm{E}\left(\exp\left(t\mathsf{x}^2\right)\right) &= \mathrm{E}\left(\sum_{i=0}^{\infty}\frac{\left(t\mathsf{x}^2\right)^i}{i!}\right)\\ &= \sum_{i=0}^{\infty}\frac{\mathrm{E}\left(t^i\left(\mathsf{x}^{2i}\right)\right)}{i!}.\end{aligned}$$

Proof

Fix $t > 0$

$$\begin{aligned} P(\mathbf{y} > a) &\leq \exp(-at) \prod_{i=1}^{k} \mathrm{E}\left(\exp\left(t\mathbf{x_i}^2\right)\right) \\ &= \frac{\exp(-at)}{(1-2t)^{\frac{k}{2}}} \end{aligned}$$

Proof

Setting $a := k(1+\epsilon)$ and

$$t := \frac{1}{2} - \frac{1}{2(1+\epsilon)},$$

we conclude

$$\begin{aligned} P(\mathbf{y} > k(1+\epsilon)) &\leq (1+\epsilon)^k \, 2 \exp\left(-\frac{k\epsilon}{2}\right) \\ &\leq \exp\left(-\frac{k\epsilon^2}{8}\right) \end{aligned}$$

Projection onto a fixed subspace

$\mathcal{P}_{\mathcal{S}_1}\,\vec{z}$ $\mathcal{P}_{\mathcal{S}_2}\,\vec{z}$

$$0.007 = \frac{||\mathcal{P}_{\mathcal{S}_1}\,\vec{z}||_2}{||\vec{x}||_2} < \frac{||\mathcal{P}_{\mathcal{S}_2}\,\vec{z}||_2}{||\vec{x}||_2} = 0.043$$

$$\frac{0.043}{0.007} = 6.14 \approx \sqrt{\frac{\dim\left(\mathcal{S}_2\right)}{\dim\left(\mathcal{S}_1\right)}}$$ (not a coincidence)

Projection onto a fixed subspace

Let $\mathcal{S}$ be a k-dimensional subspace of $\mathbb{R}^n$ and $\vec{\mathbf{z}} \in \mathbb{R}^n$ a vector of iid standard Gaussian noise

$||\mathcal{P}_{\mathcal{S}}\, \vec{\mathbf{z}}||_2^2$ is a χ^2 random variable with k degrees of freedom

It has the same distribution as

$$\mathbf{y} := \sum_{i=1}^{k} \mathbf{x_i}^2$$

where $\mathbf{x_1}, \ldots, \mathbf{x_k}$ are iid standard Gaussians.

Proof

Let UU^T be a projection matrix for $\mathcal{S}$, where the columns of $U \in \mathbb{R}^{n \times k}$ are orthonormal:

$$||\mathcal{P}_{\mathcal{S}}\, \vec{\mathbf{z}}||_2^2$$

Proof

Let UU^T be a projection matrix for $\mathcal{S}$, where the columns of $U \in \mathbb{R}^{n \times k}$ are orthonormal:

$$||\mathcal{P}_{\mathcal{S}}\,\vec{\mathbf{z}}||_2^2 = \left|\left|UU^T\vec{\mathbf{z}}\right|\right|_2^2$$

Proof

Let UU^T be a projection matrix for $\mathcal{S}$, where the columns of $U \in \mathbb{R}^{n \times k}$ are orthonormal:

$$\begin{aligned} \left|\left|\mathcal{P}_{\mathcal{S}}\,\vec{\mathbf{z}}\right|\right|_2^2 &= \left|\left|UU^T\vec{\mathbf{z}}\right|\right|_2^2 \\ &= \vec{\mathbf{z}}^T UU^T UU^T \vec{\mathbf{z}} \end{aligned}$$

Proof

Let UU^T be a projection matrix for $\mathcal{S}$, where the columns of $U \in \mathbb{R}^{n\times k}$ are orthonormal:

$$\begin{aligned}||\mathcal{P}_{\mathcal{S}}\,\vec{\mathbf{z}}||_2^2 &= \left|\left|UU^T\vec{\mathbf{z}}\right|\right|_2^2 \\ &= \vec{\mathbf{z}}^T UU^T UU^T \vec{\mathbf{z}} \\ &= \vec{\mathbf{z}}^T UU^T \vec{\mathbf{z}}\end{aligned}$$

Proof

Let UU^T be a projection matrix for $\mathcal{S}$, where the columns of $U \in \mathbb{R}^{n \times k}$ are orthonormal:

$$\begin{aligned} ||\mathcal{P}_{\mathcal{S}}\,\vec{\mathbf{z}}||_2^2 &= \left|\left| UU^T \vec{\mathbf{z}} \right|\right|_2^2 \\ &= \vec{\mathbf{z}}^T UU^T UU^T \vec{\mathbf{z}} \\ &= \vec{\mathbf{z}}^T UU^T \vec{\mathbf{z}} \\ &= \vec{\mathbf{w}}^T \vec{\mathbf{w}} \end{aligned}$$

Proof

Let UU^T be a projection matrix for $\mathcal{S}$, where the columns of $U \in \mathbb{R}^{n\times k}$ are orthonormal:

$$
\begin{aligned}
||\mathcal{P}_{\mathcal{S}}\,\vec{\mathbf{z}}||_2^2 &= \left|\left| UU^T\vec{\mathbf{z}} \right|\right|_2^2 \\
&= \vec{\mathbf{z}}^T UU^T UU^T \vec{\mathbf{z}} \\
&= \vec{\mathbf{z}}^T UU^T \vec{\mathbf{z}} \\
&= \vec{\mathbf{w}}^T \vec{\mathbf{w}} \\
&= \sum_{i=1}^{k} \vec{\mathbf{w}}[i]^2
\end{aligned}
$$

Proof

Let UU^T be a projection matrix for $\mathcal{S}$, where the columns of $U \in \mathbb{R}^{n \times k}$ are orthonormal:

$$\begin{aligned}
||\mathcal{P}_{\mathcal{S}}\,\vec{\mathbf{z}}||_2^2 &= \left|\left|UU^T\vec{\mathbf{z}}\right|\right|_2^2 \\
&= \vec{\mathbf{z}}^T UU^T UU^T \vec{\mathbf{z}} \\
&= \vec{\mathbf{z}}^T UU^T \vec{\mathbf{z}} \\
&= \vec{\mathbf{w}}^T \vec{\mathbf{w}} \\
&= \sum_{i=1}^{k} \vec{\mathbf{w}}[i]^2
\end{aligned}$$

$\vec{\mathbf{w}} := U^T\vec{\mathbf{z}}$ is Gaussian with mean zero and covariance matrix

$$\Sigma_{\vec{\mathbf{w}}} = U^T \Sigma_{\vec{\mathbf{z}}} U$$

Proof

Let UU^T be a projection matrix for $\mathcal{S}$, where the columns of $U \in \mathbb{R}^{n \times k}$ are orthonormal:

$$\begin{aligned}
\left|\left|\mathcal{P}_{\mathcal{S}}\,\vec{\mathbf{z}}\right|\right|_2^2 &= \left|\left|UU^T\vec{\mathbf{z}}\right|\right|_2^2 \\
&= \vec{\mathbf{z}}^T UU^T UU^T \vec{\mathbf{z}} \\
&= \vec{\mathbf{z}}^T UU^T \vec{\mathbf{z}} \\
&= \vec{\mathbf{w}}^T \vec{\mathbf{w}} \\
&= \sum_{i=1}^{k} \vec{\mathbf{w}}[i]^2
\end{aligned}$$

$\vec{\mathbf{w}} := U^T\vec{\mathbf{z}}$ is Gaussian with mean zero and covariance matrix

$$\begin{aligned}
\Sigma_{\vec{\mathbf{w}}} &= U^T \Sigma_{\vec{\mathbf{z}}} U \\
&= U^T U = I
\end{aligned}$$

Non-asymptotic Chernoff tail bound

Let $\vec{x}$ be an iid standard Gaussian random vector of dimension k

For any $\epsilon > 0$

$$P\left(k\left(1-\epsilon\right) < ||\vec{x}||_2^2 < k\left(1+\epsilon\right)\right) \geq 1 - 2\exp\left(-\frac{k\epsilon^2}{8}\right)$$

Projection onto a fixed subspace

Let $\mathcal{S}$ be a k-dimensional subspace of $\mathbb{R}^n$ and $\vec{z} \in \mathbb{R}^n$ a vector of iid standard Gaussian noise

For any $\epsilon > 0$

$$P\left(k\left(1-\epsilon\right) < ||\mathcal{P}_{\mathcal{S}}\,\vec{z}||_2 < k\left(1+\epsilon\right)\right) \geq 1 - 2\exp\left(-\frac{k\epsilon^2}{8}\right)$$

Dimensionality reduction

- PCA preserves the most *energy* (ℓ_2 norm)
- Problem 1: Computationally expensive
- Problem 2: Depends on *all* of the data
- (Possible) Solution: Just project randomly!
- For a data set $\vec{x}_1, \vec{x}_2, \ldots \in \mathbb{R}^m$ compute $\mathbf{A}\vec{x}_1, \mathbf{A}\vec{x}_2, \ldots \in \mathbb{R}^m$ where $\mathbf{A} \in \mathbb{R}^{k \times n}$ ($k < n$) has iid standard Gaussian entries

Fixed vector

Let $\mathbf{A}$ be a $a \times b$ matrix with iid standard Gaussian entries

If $\vec{v} \in \mathbb{R}^b$ is a deterministic vector with unit ℓ_2 norm, then $\mathbf{A}\vec{v}$ is an a-dimensional iid standard Gaussian vector

Proof:

Fixed vector

Let $\mathbf{A}$ be a $a \times b$ matrix with iid standard Gaussian entries

If $\vec{v} \in \mathbb{R}^b$ is a deterministic vector with unit ℓ_2 norm, then $\mathbf{A}\vec{v}$ is an a-dimensional iid standard Gaussian vector

Proof:

$(\mathbf{A}\vec{v})[i]$, $1 \leq i \leq a$ is Gaussian with mean zero and variance

$$\begin{aligned}\mathrm{Var}\left(\mathbf{A}_{i,:}^T \vec{v}\right) &= \vec{v}^T \Sigma_{\mathbf{A}_{i,:}} \vec{v} \\ &= \vec{v}^T I \vec{v} \\ &= \|\vec{v}\|_2^2 = 1\end{aligned}$$

Non-asymptotic Chernoff tail bound

Let $\vec{x}$ be an iid standard Gaussian random vector of dimension k

For any $\epsilon > 0$

$$P\left(k\left(1-\epsilon\right) < ||\vec{x}||_2^2 < k\left(1+\epsilon\right)\right) \geq 1 - 2\exp\left(-\frac{k\epsilon^2}{8}\right)$$

Fixed vector

Let $\mathbf{A}$ be a $a \times b$ matrix with iid standard Gaussian entries

For any $\vec{v} \in \mathbb{R}^p$ with unit norm and any $\epsilon \in (0, 1)$

$$\sqrt{a\,(1-\epsilon)} \leq ||\mathbf{A}\vec{v}||_2 \leq \sqrt{a\,(1+\epsilon)}$$

with probability at least $1 - 2\exp\left(-a\epsilon^2/8\right)$

Johnson-Lindenstrauss lemma

Let $\mathbf{A}$ be a $k \times n$ matrix with iid standard Gaussian entries

Let $\vec{x}_1, \ldots, \vec{x}_p \in \mathbb{R}^n$ be any fixed set of p deterministic vectors

For any pair $\vec{x}_i, \vec{x}_j$ and any $\epsilon \in (0, 1)$

$$(1 - \epsilon) \left|\left|\vec{x}_i - \vec{x}_j\right|\right|_2^2 \leq \left|\left|\frac{1}{\sqrt{k}}\mathbf{A}\vec{x}_i - \frac{1}{\sqrt{k}}\mathbf{A}\vec{x}_j\right|\right|_2^2 \leq (1 + \epsilon) \left|\left|\vec{x}_i - \vec{x}_j\right|\right|_2^2$$

with probability at least $\frac{1}{p}$ as long as

$$k \geq \frac{16 \log(p)}{\epsilon^2}$$

Proof

Aim: Control action of **A** the normalized differences

$$\vec{v}_{ij} := \frac{\vec{x}_i - \vec{x}_j}{\left|\left|\vec{x}_i - \vec{x}_j\right|\right|_2}$$

Our event of interest is the intersection of the events

$$\mathcal{E}_{ij} = \left\{ k\left(1-\epsilon\right) < \left|\left|\mathbf{A}\vec{v}_{ij}\right|\right|_2^2 < k\left(1+\epsilon\right) \right\} \quad 1 \leq i < p, \ i < j \leq p$$

Fixed vector

Let $\mathbf{A}$ be a $a \times b$ matrix with iid standard Gaussian entries

For any $\vec{v} \in \mathbb{R}^b$ with unit norm and any $\epsilon \in (0, 1)$

$$\sqrt{a\,(1-\epsilon)} \leq ||\mathbf{A}\vec{v}||_2 \leq \sqrt{a\,(1+\epsilon)}$$

with probability at least $1 - 2\exp\left(-a\epsilon^2/8\right)$

This implies

$$\mathrm{P}\left(\mathcal{E}_{ij}^c\right) \leq \frac{2}{p^2} \quad \text{if } k \geq \frac{16\,\log(p)}{\epsilon^2}$$

Union bound

For any events $S_1, S_2, \ldots, S_n$ in a probability space

$$\mathrm{P}\left(\cup_i S_i\right) \leq \sum_{i=1}^{n} \mathrm{P}\left(S_i\right).$$

Proof

Number of events $\mathcal{E}_{ij}$ equals $\binom{p}{2} = p\,(p-1)\,/2$

By the union bound

$$\mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right)$$

Proof

Number of events $\mathcal{E}_{ij}$ equals $\binom{p}{2} = p\,(p-1)\,/2$

By the union bound

$$\mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right) = 1 - \mathrm{P}\left(\bigcup_{i,j} \mathcal{E}_{ij}^{c}\right)$$

Proof

Number of events $\mathcal{E}_{ij}$ equals $\binom{p}{2} = p\,(p-1)\,/2$

By the union bound

$$\mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right) = 1 - \mathrm{P}\left(\bigcup_{i,j} \mathcal{E}_{ij}^c\right) \geq 1 - \sum_{i,j} \mathrm{P}\left(\mathcal{E}_{ij}^c\right)$$

Proof

Number of events $\mathcal{E}_{ij}$ equals $\binom{p}{2} = p\,(p-1)\,/2$

By the union bound

$$\begin{aligned}
\mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right) &= 1 - \mathrm{P}\left(\bigcup_{i,j} \mathcal{E}_{ij}^c\right) \\
&\geq 1 - \sum_{i,j} \mathrm{P}\left(\mathcal{E}_{ij}^c\right) \\
&\geq 1 - \frac{p\,(p-1)}{2}\,\frac{2}{p^2}
\end{aligned}$$

Proof

Number of events $\mathcal{E}_{ij}$ equals $\binom{p}{2} = p\,(p-1)\,/2$

By the union bound

$$\begin{aligned} \mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right) &= 1 - \mathrm{P}\left(\bigcup_{i,j} \mathcal{E}_{ij}^c\right) \\ &\geq 1 - \sum_{i,j} \mathrm{P}\left(\mathcal{E}_{ij}^c\right) \\ &\geq 1 - \frac{p\,(p-1)}{2}\,\frac{2}{p^2} \\ &\geq \frac{1}{p} \end{aligned}$$

Dimensionality reduction for visualization

Motivation: Visualize high-dimensional features projected onto 2D or 3D

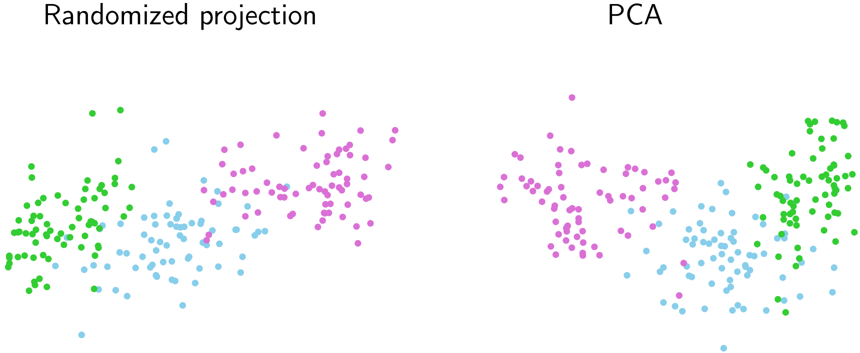
Example:

Seeds from three different varieties of wheat: Kama, Rosa and Canadian

Features:

- Area
- Perimeter
- Compactness
- Length of kernel
- Width of kernel
- Asymmetry coefficient
- Length of kernel groove

Dimensionality reduction for visualization



Nearest neighbors in random subspace

Nearest neighbors classification (Algorithm 4.2 in Lecture Notes 1) computes n distances in $\mathbb{R}^m$ for each new example

Cost: $\mathcal{O}(nmp)$ for p examples

Idea: Use a $k \times m$ iid standard Gaussian matrix to project onto k-dimensional space beforehand

Cost:

- kmn operations to project training set
- kmp operations to project test set
- knp to perform nearest-neighbor classification

Much faster!

Face recognition

Training set: 360 64×64 images from 40 different subjects (9 each)

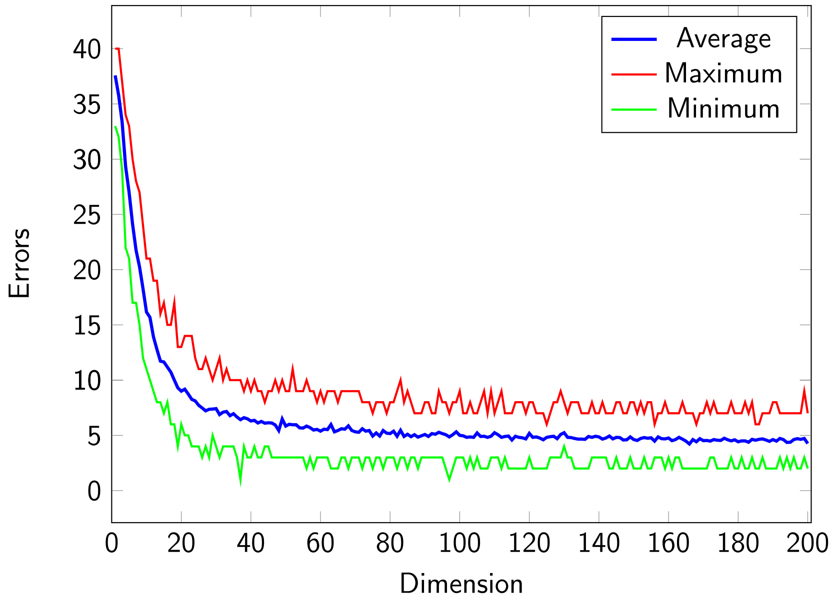
Test set: 1 new image from each subject

We model each image as a vector in $\mathbb{R}^{4096}$ ($m = 4096$)

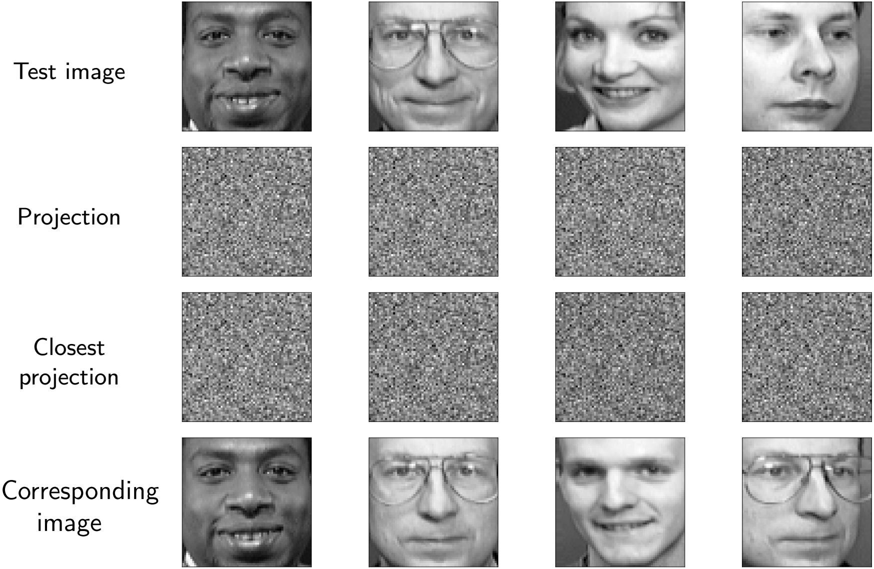
To classify we:

1. Project onto random a k-dimensional subspace
2. Apply nearest-neighbor classification using the ℓ_2-norm distance in $\mathbb{R}^k$

Performance



Nearest neighbor in $\mathbb{R}^{50}$



Gaussian random variables

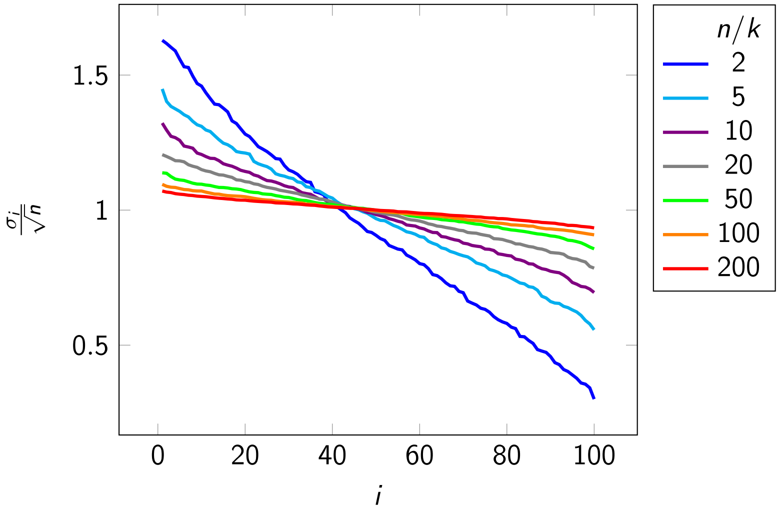
Gaussian random vectors

Randomized projections

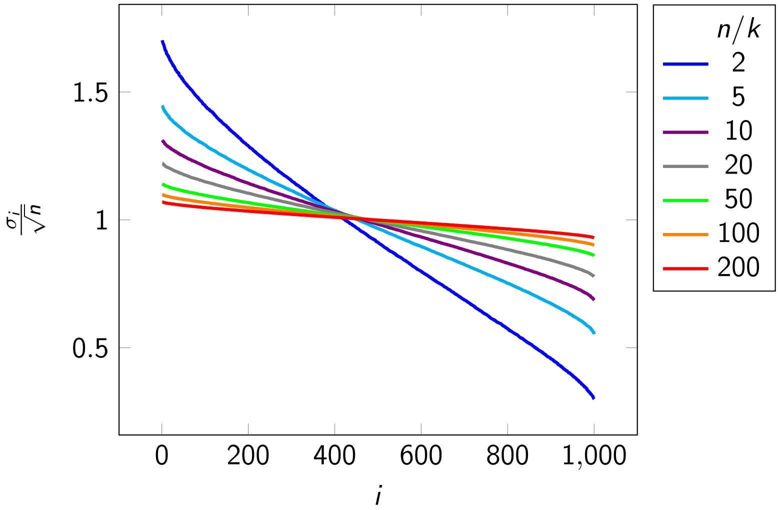
SVD of a random matrix

Randomized SVD

Singular values of $n \times k$ matrix, $k = 100$



Singular values of $n \times k$ matrix, $k = 1000$



Singular values of a Gaussian matrix

Intuitively as n grows

$$\mathbf{A} \approx U\left(\sqrt{n}\, I\right) V^T = \sqrt{n}\, UV^T,$$

iid Gaussian vectors in high dimensions are *almost* orthogonal

Singular values of a Gaussian matrix

Let $\mathbf{A}$ be a $n \times k$ matrix with iid standard Gaussian entries such that $n > k$

For any fixed $\epsilon > 0$, the singular values of $\mathbf{A}$ satisfy

$$\sqrt{n(1-\epsilon)} \leq \boldsymbol{\sigma}_{\mathbf{k}} \leq \boldsymbol{\sigma}_{\mathbf{1}} \leq \sqrt{n(1+\epsilon)}$$

with probability at least $1 - 1/k$ as long as

$$n > \frac{64k}{\epsilon^2} \log \frac{12}{\epsilon}$$

Proof

Recall that

$$\sigma_1 = \max_{\left\{||\vec{x}||_2=1 \mid \vec{x}\in\mathbb{R}^k\right\}} ||\mathbf{A}\vec{x}||_2$$
$$\sigma_k = \min_{\left\{||\vec{x}||_2=1 \mid \vec{x}\in\mathbb{R}^k\right\}} ||\mathbf{A}\vec{x}||_2$$

so the bounds are equivalent to

$$n\left(1-\epsilon\right) < ||\mathbf{A}\vec{v}||_2^2 < n\left(1+\epsilon\right)$$

Proof

Idea: Use union bound over all unit-norm vectors

Problem: They are infinite!

Solution: Use union bound on a finite set, then show that this is enough

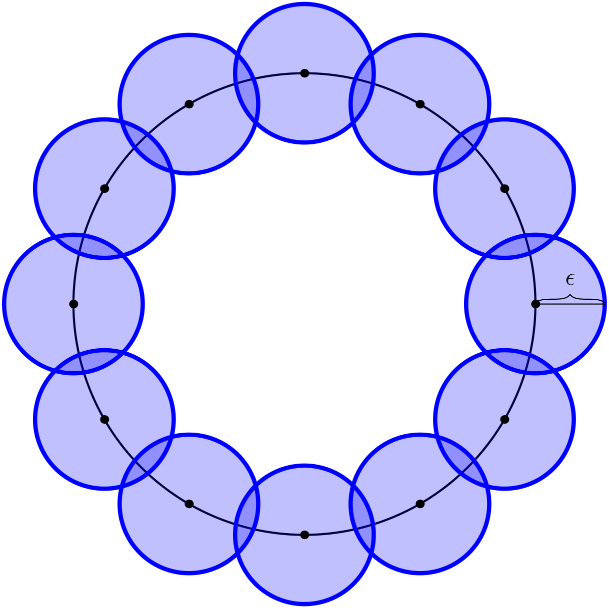
ϵ-net

An ϵ-net of a set $\mathcal{X} \subseteq \mathbb{R}^k$ is a subset $\mathcal{N}_\epsilon \subseteq \mathcal{X}$ such that for every vector $\vec{x} \in \mathcal{X}$ there exists $\vec{y} \in \mathcal{N}_\epsilon$ for which

$$||\vec{x} - \vec{y}||_2 \leq \epsilon.$$

The covering number $\mathcal{N}(\mathcal{X}, \epsilon)$ of a set $\mathcal{X}$ at scale ϵ is the minimal cardinality of an ϵ-net of $\mathcal{X}$

ϵ-net



Covering number of a sphere

The covering number of the *n*-dimensional sphere $\mathcal{S}^{k-1}$ at scale ϵ satisfies

$$\mathcal{N}\left(\mathcal{S}^{k-1}, \epsilon\right) \leq \left(\frac{2+\epsilon}{\epsilon}\right)^{k} \leq \left(\frac{3}{\epsilon}\right)^{k}$$

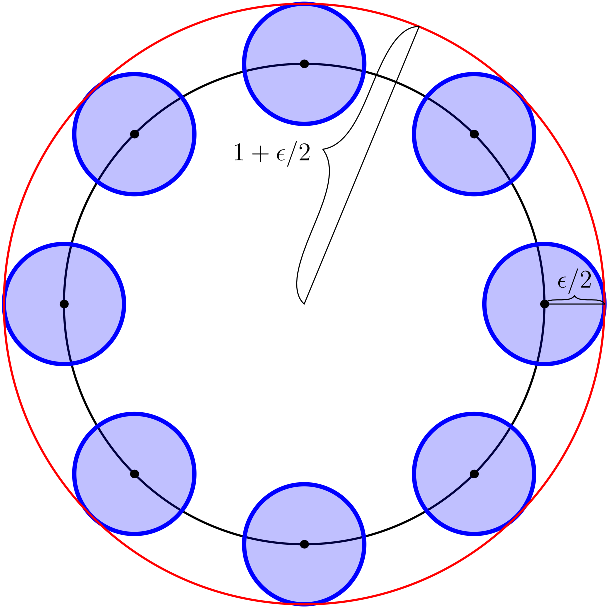
Covering number of a sphere

- Initialize $\mathcal{N}_\epsilon$ to the empty set
- Choose a point $\vec{x} \in \mathcal{S}^{k-1}$ such that

$$||\vec{x} - \vec{y}||_2 > \epsilon \quad \text{for any } \vec{y} \in \mathcal{N}_\epsilon$$

- Add $\vec{x}$ to $\mathcal{N}_\epsilon$ until there are no points in $\mathcal{S}^{k-1}$ that are ϵ away from any point in $\mathcal{N}_\epsilon$

Covering number of a sphere



Covering number of a sphere

$$\mathrm{Vol}\left(\mathcal{B}^{k}_{1+\epsilon/2}\left(\vec{0}\right)\right) \geq \mathrm{Vol}\left(\cup_{\vec{x}\in\mathcal{N}_\epsilon}\mathcal{B}^{k}_{\epsilon/2}\left(\vec{x}\right)\right)$$

Covering number of a sphere

$$\begin{aligned}\text{Vol}\left(\mathcal{B}^k_{1+\epsilon/2}\left(\vec{0}\right)\right) &\geq \text{Vol}\left(\cup_{\vec{x}\in\mathcal{N}_\epsilon}\mathcal{B}^k_{\epsilon/2}\left(\vec{x}\right)\right)\\ &= \left|\mathcal{N}_\epsilon\right| \text{Vol}\left(\mathcal{B}^k_{\epsilon/2}\left(\vec{0}\right)\right)\end{aligned}$$

Covering number of a sphere

$$\begin{aligned}\text{Vol}\left(\mathcal{B}^k_{1+\epsilon/2}\left(\vec{0}\right)\right) &\geq \text{Vol}\left(\cup_{\vec{x}\in\mathcal{N}_\epsilon}\mathcal{B}^k_{\epsilon/2}\left(\vec{x}\right)\right)\\ &= \left|\mathcal{N}_\epsilon\right| \text{Vol}\left(\mathcal{B}^k_{\epsilon/2}\left(\vec{0}\right)\right)\end{aligned}$$

By multivariable calculus

$$\text{Vol}\left(\mathcal{B}^k_r\left(\vec{0}\right)\right) = r^k\,\text{Vol}\left(\mathcal{B}^k_1\left(\vec{0}\right)\right)$$

Covering number of a sphere

$$\begin{aligned}\text{Vol}\left(\mathcal{B}^k_{1+\epsilon/2}\left(\vec{0}\right)\right) &\geq \text{Vol}\left(\cup_{\vec{x}\in\mathcal{N}_\epsilon}\mathcal{B}^k_{\epsilon/2}\left(\vec{x}\right)\right)\\ &= |\mathcal{N}_\epsilon|\,\text{Vol}\left(\mathcal{B}^k_{\epsilon/2}\left(\vec{0}\right)\right)\end{aligned}$$

$$\text{Vol}\left(\mathcal{B}^k_r\left(\vec{0}\right)\right) = r^k\,\text{Vol}\left(\mathcal{B}^k_1\left(\vec{0}\right)\right)$$

so we conclude

$$(1+\epsilon/2)^k \geq |\mathcal{N}_\epsilon|\,(\epsilon/2)^k$$

Proof

1. We prove the bounds

$$n\left(1-\epsilon_2\right) < ||\mathbf{A}\vec{v}||_2^2 < n\left(1+\epsilon_2\right)$$

   where $\epsilon_2 := \epsilon/2$ on an $\epsilon_1 := \epsilon/4$ net of the sphere

2. We show that by the triangle inequality, this implies that the bounds hold on all the sphere

Fixed vector

Let $\mathbf{A}$ be a $a \times b$ matrix with iid standard Gaussian entries

For any $\vec{v} \in \mathbb{R}^b$ with unit norm and any $\epsilon \in (0, 1)$

$$\sqrt{a\,(1-\epsilon)} \leq \|\mathbf{A}\vec{v}\|_2 \leq \sqrt{a\,(1+\epsilon)}$$

with probability at least $1 - 2\exp\left(-a\epsilon^2/8\right)$

Bound on the ϵ_1-net

We define the event

$$\mathcal{E}_{\vec{v},\epsilon_2} := \left\{ n\left(1-\epsilon_2\right)||\vec{v}||_2^2 \leq ||\mathbf{A}\vec{v}||_2^2 \leq n\left(1+\epsilon_2\right)||\vec{v}||_2^2 \right\}$$

$$\mathrm{P}\left(\cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}}\mathcal{E}^c_{\vec{v},\epsilon_2}\right)$$

Bound on the ϵ_1-net

We define the event

$$\mathcal{E}_{\vec{v},\epsilon_2} := \left\{ n\left(1-\epsilon_2\right)||\vec{v}||_2^2 \leq ||\mathbf{A}\vec{v}||_2^2 \leq n\left(1+\epsilon_2\right)||\vec{v}||_2^2 \right\}$$

$$\mathrm{P}\left(\cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}}\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \leq \sum_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right)$$

Bound on the ϵ_1-net

We define the event

$$\mathcal{E}_{\vec{v},\epsilon_2} := \left\{ n\left(1-\epsilon_2\right)||\vec{v}||_2^2 \leq ||\mathbf{A}\vec{v}||_2^2 \leq n\left(1+\epsilon_2\right)||\vec{v}||_2^2 \right\}$$

$$\begin{aligned} \mathrm{P}\left(\cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v},\epsilon_2}\right) &\leq \sum_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \\ &\leq \left|\mathcal{N}_{\epsilon_1}\right| \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \end{aligned}$$

Bound on the ϵ_1-net

We define the event

$$\mathcal{E}_{\vec{v},\epsilon_2} := \left\{ n\left(1-\epsilon_2\right)||\vec{v}||_2^2 \leq ||\mathbf{A}\vec{v}||_2^2 \leq n\left(1+\epsilon_2\right)||\vec{v}||_2^2 \right\}$$

$$\begin{aligned}
\mathrm{P}\left(\cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}}\mathcal{E}^c_{\vec{v},\epsilon_2}\right) &\leq \sum_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \\
&\leq |\mathcal{N}_{\epsilon_1}|\,\mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \\
&\leq 2\left(\frac{12}{\epsilon}\right)^k \exp\left(-\frac{n\epsilon^2}{32}\right)
\end{aligned}$$

Bound on the ϵ_1-net

We define the event

$$\mathcal{E}_{\vec{v},\epsilon_2} := \left\{ n\left(1-\epsilon_2\right)||\vec{v}||_2^2 \leq ||\mathbf{A}\vec{v}||_2^2 \leq n\left(1+\epsilon_2\right)||\vec{v}||_2^2 \right\}$$

$$\begin{aligned}
\mathrm{P}\left(\cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v},\epsilon_2}\right) &\leq \sum_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \\
&\leq \left|\mathcal{N}_{\epsilon_1}\right| \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \\
&\leq 2\left(\frac{12}{\epsilon}\right)^k \exp\left(-\frac{n\epsilon^2}{32}\right) \\
&\leq \frac{1}{k} \qquad \text{if } n > \frac{64k}{\epsilon^2}\log\frac{12}{\epsilon}
\end{aligned}$$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{k-1}$

There exists $\vec{v} \in \mathcal{N}\left(\mathcal{X}, \epsilon_1\right)$ such that $\left|\left|\vec{x} - \vec{v}\right|\right|_2 \leq \epsilon/4$

$\left|\left|\mathbf{A}\vec{x}\right|\right|_2$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{k-1}$

There exists $\vec{v} \in \mathcal{N}(\mathcal{X}, \epsilon_1)$ such that $||\vec{x} - \vec{v}||_2 \leq \epsilon/4$

$$||\mathbf{A}\vec{x}||_2 \leq ||\mathbf{A}\vec{v}||_2 + ||\mathbf{A}(\vec{x} - \vec{v})||_2$$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{k-1}$

There exists $\vec{v} \in \mathcal{N}(\mathcal{X}, \epsilon_1)$ such that $||\vec{x} - \vec{v}||_2 \leq \epsilon/4$

$$\begin{aligned} ||\mathbf{A}\vec{x}||_2 &\leq ||\mathbf{A}\vec{v}||_2 + ||\mathbf{A}(\vec{x} - \vec{v})||_2 \\ &\leq \sqrt{n}\left(1 + \frac{\epsilon}{2}\right) + ||\mathbf{A}(\vec{x} - \vec{v})||_2 \qquad \text{assuming } \cup_{\vec{v} \in \mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v}, \epsilon_2} \text{ holds} \end{aligned}$$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{k-1}$

There exists $\vec{v} \in \mathcal{N}(\mathcal{X}, \epsilon_1)$ such that $||\vec{x} - \vec{v}||_2 \leq \epsilon/4$

$$
\begin{aligned}
||\mathbf{A}\vec{x}||_2 &\leq ||\mathbf{A}\vec{v}||_2 + ||\mathbf{A}(\vec{x} - \vec{v})||_2 \\
&\leq \sqrt{n}\left(1 + \frac{\epsilon}{2}\right) + ||\mathbf{A}(\vec{x} - \vec{v})||_2 \qquad \text{assuming } \cup_{\vec{v} \in \mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v},\epsilon_2} \text{ holds} \\
&\leq \sqrt{n}\left(1 + \frac{\epsilon}{2}\right) + \boldsymbol{\sigma_1} ||\vec{x} - \vec{v}||_2
\end{aligned}
$$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{k-1}$

There exists $\vec{v} \in \mathcal{N}\left(\mathcal{X}, \epsilon_1\right)$ such that $\left|\left|\vec{x} - \vec{v}\right|\right|_2 \leq \epsilon/4$

$$\begin{aligned}
\left|\left|\mathbf{A}\vec{x}\right|\right|_2 &\leq \left|\left|\mathbf{A}\vec{v}\right|\right|_2 + \left|\left|\mathbf{A}\left(\vec{x} - \vec{v}\right)\right|\right|_2 \\
&\leq \sqrt{n}\left(1 + \frac{\epsilon}{2}\right) + \left|\left|\mathbf{A}\left(\vec{x} - \vec{v}\right)\right|\right|_2 \qquad \text{assuming } \cup_{\vec{v} \in \mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v}, \epsilon_2} \text{ holds} \\
&\leq \sqrt{n}\left(1 + \frac{\epsilon}{2}\right) + \boldsymbol{\sigma_1}\left|\left|\vec{x} - \vec{v}\right|\right|_2 \\
&\leq \sqrt{n}\left(1 + \frac{\epsilon}{2}\right) + \frac{\boldsymbol{\sigma_1}\epsilon}{4}
\end{aligned}$$

Upper bound on the sphere

$$\boldsymbol{\sigma_1} \leq \sqrt{n}\left(1 + \frac{\epsilon}{2}\right) + \frac{\boldsymbol{\sigma_1}\epsilon}{4}$$

$$\begin{aligned}\boldsymbol{\sigma_1} &\leq \sqrt{n}\left(\frac{1+\epsilon/2}{1-\epsilon/4}\right)\\ &= \sqrt{n}\left(1+\epsilon-\frac{\epsilon\,(1-\epsilon)}{4-\epsilon}\right)\\ &\leq \sqrt{n}\,(1+\epsilon)\end{aligned}$$

Lower bound on the sphere

$$||\mathbf{A}\vec{x}||_2$$

Lower bound on the sphere

$$||\mathbf{A}\vec{x}||_2 \geq ||\mathbf{A}\vec{v}||_2 - ||\mathbf{A}\left(\vec{x} - \vec{v}\right)||_2$$

Lower bound on the sphere

$$\begin{aligned} ||\mathbf{A}\vec{x}||_2 &\geq ||\mathbf{A}\vec{v}||_2 - ||\mathbf{A}(\vec{x}-\vec{v})||_2 \\ &\geq \sqrt{n}\left(1-\frac{\epsilon}{2}\right) - ||A(\vec{x}-\vec{v})||_2 \qquad \text{assuming } \cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}}\mathcal{E}^c_{\vec{v},\epsilon_2} \text{ holds} \end{aligned}$$

Lower bound on the sphere

$$
\begin{aligned}
||\mathbf{A}\vec{x}||_2 &\geq ||\mathbf{A}\vec{v}||_2 - ||\mathbf{A}\left(\vec{x}-\vec{v}\right)||_2 \\
&\geq \sqrt{n}\left(1-\frac{\epsilon}{2}\right) - ||A\left(\vec{x}-\vec{v}\right)||_2 \qquad \text{assuming } \cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}}\mathcal{E}^{c}_{\vec{v},\epsilon_2} \text{ holds} \\
&\geq \sqrt{n}\left(1-\frac{\epsilon}{2}\right) - \boldsymbol{\sigma_1}\,||\vec{x}-\vec{v}||_2
\end{aligned}
$$

Lower bound on the sphere

$$\begin{aligned}
||\mathbf{A}\vec{x}||_2 &\geq ||\mathbf{A}\vec{v}||_2 - ||\mathbf{A}\left(\vec{x} - \vec{v}\right)||_2 \\
&\geq \sqrt{n}\left(1 - \frac{\epsilon}{2}\right) - ||A\left(\vec{x} - \vec{v}\right)||_2 \qquad \text{assuming } \cup_{\vec{v} \in \mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v},\epsilon_2} \text{ holds} \\
&\geq \sqrt{n}\left(1 - \frac{\epsilon}{2}\right) - \boldsymbol{\sigma_1}\,||\vec{x} - \vec{v}||_2 \\
&\geq \sqrt{n}\left(1 - \frac{\epsilon}{2}\right) - \frac{\epsilon}{4}\sqrt{n}\left(1 + \epsilon\right)
\end{aligned}$$

Lower bound on the sphere

$$
\begin{aligned}
||\mathbf{A}\vec{x}||_2 &\geq ||\mathbf{A}\vec{v}||_2 - ||\mathbf{A}(\vec{x}-\vec{v})||_2 \\
&\geq \sqrt{n}\left(1-\frac{\epsilon}{2}\right) - ||A(\vec{x}-\vec{v})||_2 \qquad \text{assuming } \cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathcal{E}^{c}_{\vec{v},\epsilon_2} \text{ holds} \\
&\geq \sqrt{n}\left(1-\frac{\epsilon}{2}\right) - \boldsymbol{\sigma_1}\,||\vec{x}-\vec{v}||_2 \\
&\geq \sqrt{n}\left(1-\frac{\epsilon}{2}\right) - \frac{\epsilon}{4}\sqrt{n}\,(1+\epsilon) \\
&= \sqrt{n}\,(1-\epsilon)
\end{aligned}
$$

Gaussian random variables

Gaussian random vectors

Randomized projections

SVD of a random matrix

Randomized SVD

Fast SVD

For a matrix $M \in \mathbb{R}^{m \times n}$ which is approximately rank k:

1. Choose a small oversampling parameter p (usually 5 or slightly larger).
2. Find a matrix $\widetilde{U} \in \mathbb{R}^{m \times (k+p)}$ with $k + p$ orthonormal columns that approximately span the column space of M
3. Compute $W \in \mathbb{R}^{(k+p) \times n}$ defined by $W := \widetilde{U}^T M$
4. Compute the SVD of $W = U_W S_W V_W^T$
5. Output $U := (\widetilde{U} U_W)_{:,1:k}$, $S := (S_W)_{1:k,1:k}$ and $V := (V_W)_{:,1:k}$ as the SVD of M

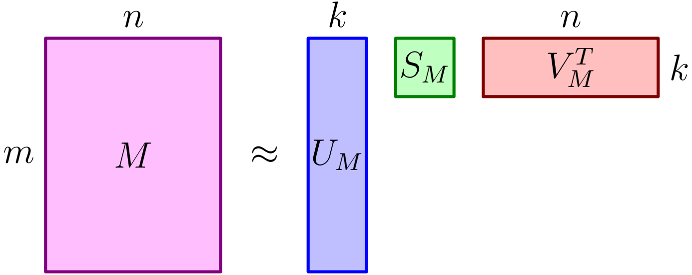
Fast SVD

For a matrix $M \in \mathbb{R}^{m \times n}$ which is approximately rank k:

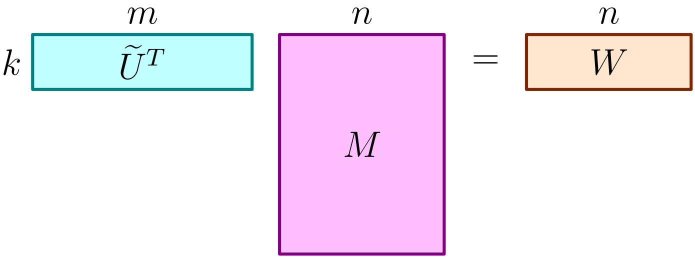
1. Choose a small oversampling parameter p (usually 5 or slightly larger).
2. Find a matrix $\widetilde{U} \in \mathbb{R}^{m \times (k+p)}$ with $k+p$ orthonormal columns that approximately span the column space of M
3. Compute $W \in \mathbb{R}^{(k+p) \times n}$ defined by $W := \widetilde{U}^T M$ $\mathcal{O}(kmn)$
4. Compute the SVD of $W = U_W S_W V_W^T$ $\mathcal{O}(k^2 n)$
5. Output $U := (\widetilde{U} U_W)_{:,1:k}$, $S := (S_W)_{1:k,1:k}$ and $V := (V_W)_{:,1:k}$ as the SVD of M

Complexity of regular SVD is $\mathcal{O}(mn \min\{m, n\})$

Fast SVD



Fast SVD



Fast SVD

The method works if (1) M is rank k and (2) $\widetilde{U}$ spans the column space

M

Fast SVD

The method works if (1) M is rank k and (2) $\widetilde{U}$ spans the column space

$$M = \widetilde{U}\widetilde{U}^T M$$

Fast SVD

The method works if (1) M is rank k and (2) $\widetilde{U}$ spans the column space

$$
\begin{aligned}
M &= \widetilde{U}\widetilde{U}^T M \\
&= \widetilde{U} W
\end{aligned}
$$

Fast SVD

The method works if (1) M is rank k and (2) $\widetilde{U}$ spans the column space

$$\begin{aligned} M &= \widetilde{U}\widetilde{U}^T M \\ &= \widetilde{U} W \\ &= \widetilde{U} U_W S_W V_W^T \end{aligned}$$

Fast SVD

The method works if (1) M is rank k and (2) $\widetilde{U}$ spans the column space

$$\begin{aligned} M &= \widetilde{U}\widetilde{U}^T M \\ &= \widetilde{U} W \\ &= \widetilde{U} U_W S_W V_W^T \end{aligned}$$

where $U := \widetilde{U} U_W$ is an $m \times k$ matrix with orthonormal columns

Fast SVD

The method works if (1) M is rank k and (2) $\widetilde{U}$ spans the column space

$$\begin{aligned} M &= \widetilde{U}\widetilde{U}^T M \\ &= \widetilde{U} W \\ &= \widetilde{U} U_W S_W V_W^T \end{aligned}$$

where $U := \widetilde{U} U_W$ is an $m \times k$ matrix with orthonormal columns

$$U^T U = U_W^T \widetilde{U}^T \widetilde{U} U_W$$

Fast SVD

The method works if (1) M is rank k and (2) $\widetilde{U}$ spans the column space

$$
\begin{aligned}
M &= \widetilde{U}\widetilde{U}^T M \\
&= \widetilde{U} W \\
&= \widetilde{U} U_W S_W V_W^T
\end{aligned}
$$

where $U := \widetilde{U} U_W$ is an $m \times k$ matrix with orthonormal columns

$$
\begin{aligned}
U^T U &= U_W^T \widetilde{U}^T \widetilde{U} U_W \\
&= U_W^T U_W = I
\end{aligned}
$$

Power iterations

For approximately low-rank matrices performance depends on gap between σ_k and σ_{k+1}

The gap can be increased by power iterations

This method is only used when computing $\widetilde{U}$

The input is

$$\widetilde{M} := \left(MM^T\right)^q M$$

Power iterations

For approximately low-rank matrices performance depends on gap between σ_k and σ_{k+1}

The gap can be increased by power iterations

This method is only used when computing $\widetilde{U}$

The input is

$$\begin{aligned}\widetilde{M} &:= \left(MM^T\right)^q M \\ &= \left(U_M S_M^2 U_M^T\right)^q U_M S_M V_M^T\end{aligned}$$

Power iterations

For approximately low-rank matrices performance depends on gap between σ_k and σ_{k+1}

The gap can be increased by power iterations

This method is only used when computing $\widetilde{U}$

The input is

$$\begin{aligned}
\widetilde{M} &:= \left(MM^T\right)^q M \\
&= \left(U_M S_M^2 U_M^T\right)^q U_M S_M V_M^T \\
&= U_M S_M^2 U_M^T U_M S_M^2 U_M^T \cdots U_M S_M^2 U_M^T U_M V_M^T \\
&= U_M S_M^{2q+1} V_M^T
\end{aligned}$$

Problem

How do we estimate the column space of a low-rank matrix?

- Project onto random subspace with slightly larger dimension
- Select random columns

Randomized column-space approximation

For a matrix $M \in \mathbb{R}^{m \times n}$ which is approximately rank k:

1. Create an $n \times (k+p)$ iid standard Gaussian matrix $\mathbf{A}$, where p is a small integer (e.g. 5)
2. Compute the $m \times (k+p)$ matrix $\mathbf{B} = M\mathbf{A}$
3. Orthonormalize the columns of $\mathbf{B}$ and output them as a matrix $\widetilde{\mathbf{U}} \in \mathbb{R}^{m \times (k+p)}$.
4. Apply power iterations if necessary.

Randomized column-space approximation

$$\begin{aligned}\mathbf{B} &= M\mathbf{A} \\ &= U_M S_M V_M^T \mathbf{A} \\ &= U_M S_M \mathbf{C}\end{aligned}$$

Randomized column-space approximation

$$\begin{aligned}\mathbf{B} &= M\mathbf{A} \\ &= U_M S_M V_M^T \mathbf{A} \\ &= U_M S_M \mathbf{C}\end{aligned}$$

- If M is low rank $\mathbf{C}$ is a $k \times (k + p)$ iid standard Gaussian matrix

Randomized column-space approximation

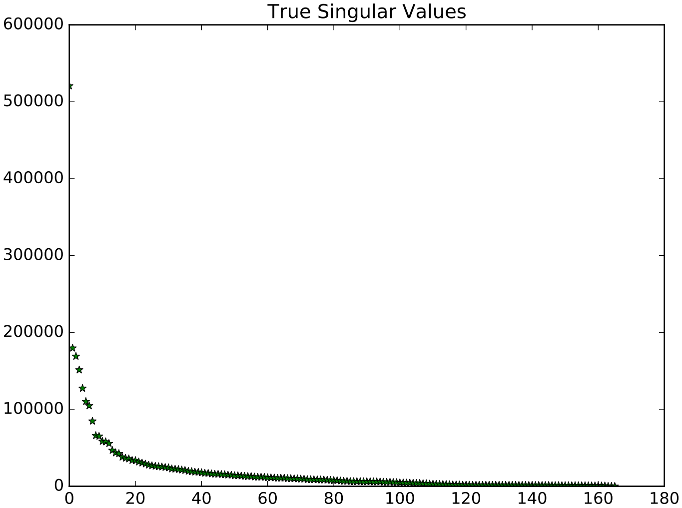
$$
\begin{aligned}
\mathbf{B} &= M\mathbf{A} \\
&= U_M S_M V_M^T \mathbf{A} \\
&= U_M S_M \mathbf{C}
\end{aligned}
$$

- If M is low rank $\mathbf{C}$ is a $k \times (k+p)$ iid standard Gaussian matrix
- Otherwise, $\mathbf{C}$ is a $\min\{m, n\} \times (k+p)$ iid standard Gaussian matrix

Randomized SVD of a video

- Video with 160 1080×1920 frames
- We interpret each frame as a vector in $\mathbb{R}^{20,736,000}$
- Matrix formed by these vectors is approximately low rank
- Regular SVD takes 12 seconds (281.1 seconds if we take 691 frames)
- Fast SVD with randomized-column-space estimate takes 5.8 seconds (10.4 seconds for 691 frames) to obtain a rank-10 approximation ($q = 2$, $p = 7$)

True singular values



Left singular vector approximation

Random column selection

For a matrix $M \in \mathbb{R}^{m \times n}$ which is approximately rank k:

1. Select a random subset of column indices $\mathcal{I} := \{\mathbf{i_1}, \mathbf{i_2}, \ldots, \mathbf{i_{k'}}\}$ with $k' \geq k$

2. Orthonormalize the submatrix corresponding to $\mathcal{I}$:

$$M_{\mathcal{I}} := \begin{bmatrix} M_{:,\mathbf{i_1}} & M_{:,\mathbf{i_2}} & \cdots & M_{:,\mathbf{i_{k'}}} \end{bmatrix}$$

   and output them as a matrix $\widetilde{\mathbf{U}} \in \mathbb{R}^{m \times k'}$

Random column selection

(Possible) Problem: If right singular vectors are sparse, this will not work

$$\boldsymbol{M_{\mathcal{I}}} = U_M S_M (\boldsymbol{V_M})_{\mathcal{I}}$$

Example

$$M := \begin{bmatrix} -3 & 2 & 2 & 2 \\ 3 & 2 & 2 & 2 \\ -3 & 2 & 2 & 2 \\ 3 & 2 & 2 & 2 \end{bmatrix}$$

Example

$$M = U_M S_M V_M^T = \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \\ 0.5 & -0.5 \\ 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 6.9282 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 0 & 0.577 & 0.577 & 0.577 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Example, $\mathcal{I} = \{2, 3\}$

$$M_{\mathcal{I}} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \\ 2 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} 6.2982 \begin{bmatrix} 0.577 & 0.577 \end{bmatrix}.$$

Randomized SVD of a video

- Video with 160 1080×1920 frames
- We interpret each frame as a vector in $\mathbb{R}^{20,736,000}$
- Matrix formed by these vectors is approximately low rank
- Regular SVD takes 12 seconds (281.1 seconds if we take 691 frames)
- Fast SVD with random-column-selection estimate takes 5.2 seconds to obtain a rank-10 approximation ($k' = 17$)

Left singular vector approximation

Singular value approximation