Lecture 7 Concerntration

IEMS 402 Statistical Learning

Northwestern

Asymptotic VS Non-Asymptotic

Drawback of Asymptotic Theory

Concerntration

First sense of Concerntration

inequalities of the form

Randomly sampled data

$$\mathbb{P}(X \geq t) \leq \phi(t)$$

where ϕ goes to zero (quickly) as $t \to \infty$

Error/risk

First examples

Proposition (Markov's inequality)

If $X \geq 0$, *then* $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$ *for all* $t \geq 0$.

First examples

Proposition (Markov's inequality)

If $X \geq 0$*, then* $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$ *for all* $t \geq 0$*.*

Proposition (Chebyshev's inequality)

For any $t \geq 0$, $\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$

First examples

Proposition (Markov's inequality)

If $X \geq 0$*, then* $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$ *for all* $t \geq 0$.

Proposition (Chebyshev's inequality)

For any $t \geq 0$, $\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$

Should be $O(e^{-t})$?

Moment Generating Function

moment generating function

$$M_X(t) = \mathbb{E}[e^{tX}]$$

A function

Argument of moment generating function

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Moment Generating Function

moment generating function

$$M_X(t) = \mathbb{E}[e^{tX}]$$

A function

Argument of moment generating function

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

Sum of Independent Random Variables:

Suppose X_1, X_2, ..., X_n are n independent random variables, and the random variable Y is defined as

$$Y = X_1 + X_2 + \cdots + X_n.$$

Then,

$$\begin{aligned} M_Y(s) &= E[e^{sY}] \\ &= E[e^{s(X_1+X_2+\cdots+X_n)}] \\ &= E[e^{sX_1} e^{sX_2} \cdots e^{sX_n}] \\ &= E[e^{sX_1}]E[e^{sX_2}] \cdots E[e^{sX_n}] \quad \text{(since the } X_i\text{'s are independent)} \\ &= M_{X_1}(s)M_{X_2}(s) \cdots M_{X_n}(s). \end{aligned}$$

Chernoff bound

$$P(X \geq a) \leq \inf_{t>0} M(t) e^{-ta}$$

Chernoff bound

$$P(X \geq a) \leq \inf_{t>0} M(t)e^{-ta}$$

sub-Gaussian random variable

A mean-zero random variable X is σ^2-*sub-Gaussian* if

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \quad \text{for all } \lambda \in \mathbb{R}.$$

Example

Exercise

If $X \in [a, b]$, then

$$\mathbb{E}[\exp(\lambda(X - \mathbb{E}[X]))] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

Chernoff bound

$$P(X \geq a) \leq \inf_{t>0} M(t)e^{-ta}$$

<u>sub-Gaussian random variable</u>

A mean-zero random variable X is σ^2-*sub-Gaussian* if

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \quad \text{for all } \lambda \in \mathbb{R}.$$

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

Hoeffding Inequality

Corollary (Hoeffding bounds)

If X_i are independent σ_i^2-sub-Gaussian random variables,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right) \leq \exp\left(-\frac{nt^2}{\frac{2}{n}\sum_{i=1}^{n}\sigma_i^2}\right).$$

- usually stated as $X_i \in [a, b]$, so bound is $\exp(-\frac{2nt^2}{(b-a)^2})$

Should be $O(1/\sqrt{n})$?

Moment Generating Function is Powerful

Not Required

Bernstein's Inequality

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq t\right) \vee \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \leq -t\right) \leq \exp\left(-\frac{nt^2}{2\sigma^2 + 2ct/3}\right),$$

σ^2 : variance $|X_i| \leq c$

Homework 5, Question 3

Moment Generating Function is Powerful

Not Required

Proposition

Let $\{Z_i\}_{i=1}^N$ be σ^2-sub-Gaussian (not necessarily independent). Then

$$\mathbb{E}\left[\max_i Z_i\right] \leq \sqrt{2\sigma^2 \log N}.$$

Application

Johnson-Lindenstrauss Lemma

Lemma For any $0 < \epsilon < 1$ and any interger n let k be a possitive interger such that

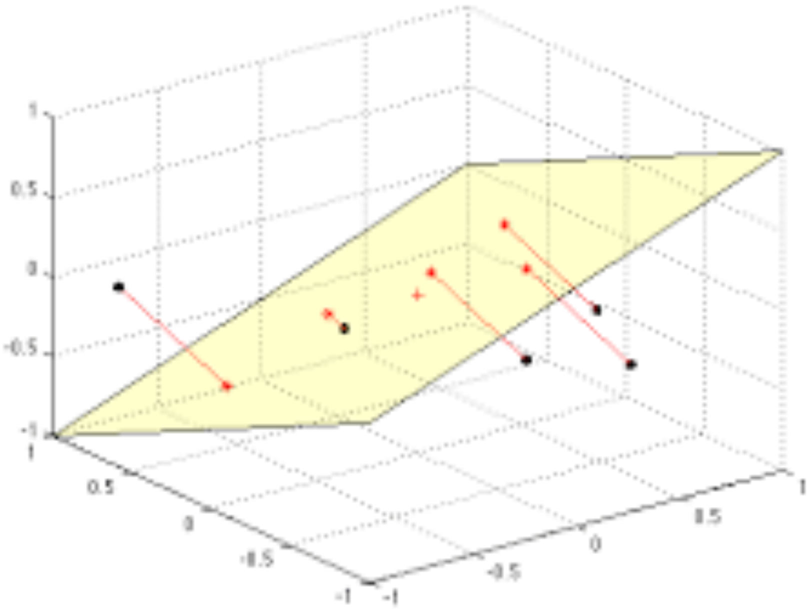
$$k \geq \frac{24}{3\epsilon^2 - 2\epsilon^3} \log n \tag{2}$$

then for any set A of n points $\in \Re^d$ there exists a map $f : \Re^d \rightarrow \Re^k$ such that for all $x_i, x_j \in A$

$$(1-\epsilon)||x_i - x_j||^2 \leq ||f(x_i) - f(x_j)||^2 \leq (1+\epsilon)||x_i - x_j||^2 \tag{3}$$

https://cs.stanford.edu/people/mmahoney/cs369m/Lectures/lecture1.pdf

Why it's important

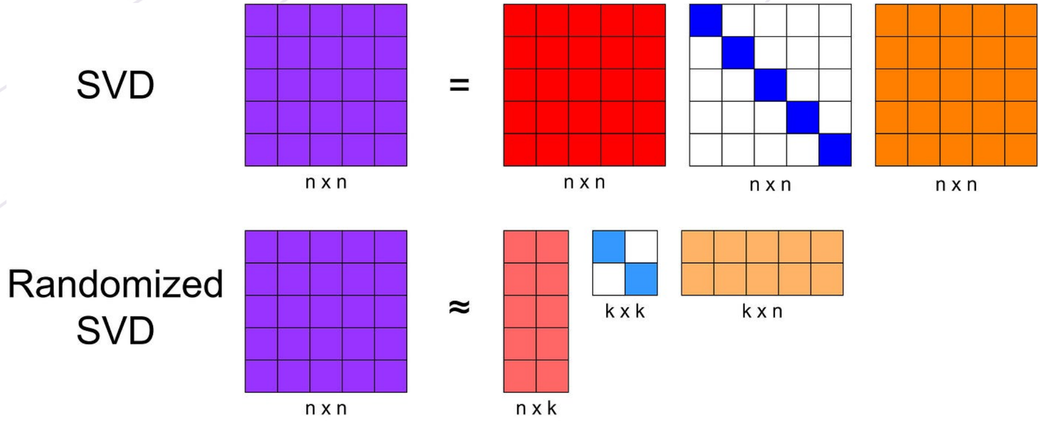


CIFAR 100: 6000 32x32images,

Idea: random projection

Definition Let R be a random matrix of order $k \times d$ i.e $R_{ij} \overset{i.i.d}{\sim} N(0,1)$ and u be any fixed vector $\in \Re^d$. Define $v = \frac{1}{\sqrt{k}} R \cdot u$. Thus $v \in \Re^k$ and $v_i = \frac{1}{\sqrt{k}} \sum_j R_{ij} u_j$

Why it’s important



Halko, Nathan, Per-Gunnar Martinsson, and Joel A. Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions." SIAM review

Idea: random projection

Definition Let R be a random matrix of order $k \times d$ i.e $R_{ij} \overset{i.i.d}{\sim} N(0,1)$ and u be any fixed vector $\in \Re^d$. Define $v = \frac{1}{\sqrt{k}} R \cdot u$. Thus $v \in \Re^k$ and $v_i = \frac{1}{\sqrt{k}} \sum_j R_{ij} u_j$

Fact 1. $\mathbb{E}[\|v\|^2] = \|u\|^2$

Question. $\mathbb{P}(\|v\|^2 \geq (1+\epsilon)\|u\|^2)$

Assume $\|u\| = 1$

Random projection

Question. $\mathbb{P}(\|v\|^2 \geq (1+\epsilon)\|u\|^2)$

$x_i = R_i^\top \cdot u$

Means $\dfrac{\sum_{i=1}^{k} x_i^2}{k} \geq (1+\epsilon)$

Random projection

Question. $\mathbb{P}(\|v\|^2 \geq (1+\epsilon)\|u\|^2)$

$$x_i = R_i^\top \cdot u$$

Means $\frac{\sum_{i=1}^{k} x_i^2}{k} \geq (1+\epsilon) \to e^{\lambda x} \geq e^{\lambda(1+\epsilon)k}$

$$\mathbb{E}[e^{\lambda x}] = \prod_{i=1}^{k} \mathbb{E}[e^{\lambda x_i}] = \left(\mathbb{E}[e^{\lambda x_i}]\right)^k$$

Random projection

Question. $\mathbb{P}(\|v\|^2 \geq (1+\epsilon)\|u\|^2)$

$x_i = R_i^\top \cdot u$

Means $\frac{\sum_{i=1}^k x_i^2}{k} \geq (1+\epsilon) \rightarrow e^{\lambda x} \geq e^{\lambda(1+\epsilon)k}$

$$\mathbb{E}[e^{\lambda x}] = \prod_{i=1}^{k} \mathbb{E}[e^{\lambda x_i}] = \left(\mathbb{E}[e^{\lambda x_i}]\right)^k$$

Thus $\mathbb{P}[e^{\lambda(1+\epsilon)k}] \leq (\frac{1}{\sqrt{1-2\lambda}})^k \cdot \frac{1}{e^{\lambda(1+\epsilon)k}}$

Random projection

Question. $\mathbb{P}(\|v\|^2 \geq (1+\epsilon)\|u\|^2)$

$x_i = R_i^\top \cdot u$

Means $\frac{\sum_{i=1}^k x_i^2}{k} \geq (1+\epsilon) \rightarrow e^{\lambda x} \geq e^{\lambda(1+\epsilon)k}$

$$\mathbb{E}[e^{\lambda x}] = \prod_{i=1}^{k} \mathbb{E}[e^{\lambda x_i}] = \left(\mathbb{E}[e^{\lambda x_i}]\right)^k$$

Thus $\mathbb{P}[e^{\lambda(1+\epsilon)k}] \leq \left(\frac{1}{\sqrt{1-2\lambda}}\right)^k \cdot \frac{1}{e^{\lambda(1+\epsilon)k}}$ → set $\lambda = \frac{\epsilon}{2(1+\epsilon)}$ → $\leq e^{-(\epsilon^2/2-\epsilon^3)k/2} \leq n^{-2}$

Why?

Uniform bound!

Note

Not Required

another proof using epsilon-net: **Theorem 8.**
https://www.cs.princeton.edu/~smattw/Teaching/Fa19Lectures/lec9/lec9.pdf