A Maldição da Dimensão

Thiago Ramos

IMPA

💀 Sexta-Feira 13 💀

Fatos Básicos

Definição

Um espaço de probabilidade é uma tripla $(\Omega, \mathcal{F}, \mathbb{P})$ *onde:*

- Ω *é um conjunto base:*
- $\mathcal{F}$ *é uma família de eventos possíveis para* Ω. *Vamos assumir que* $\mathcal{F}$ *é uma* σ*-álgebra.*
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ *uma função que retorna a probabilidade de um evento ocorrer.*

Fatos Básicos

Os eventos em $\mathcal{F}$ satisfazem:

- Se $A \in \mathcal{F}$ então $A^c \in \mathcal{F}$.
- Se $(A_i)_{i \in \mathbb{N}} \subset \mathcal{F}$ então $\cup_i A_i \in \mathcal{F}$.

Fatos Básicos

Os eventos em $\mathcal{F}$ satisfazem:

- Se $A \in \mathcal{F}$ então $A^c \in \mathcal{F}$.
- Se $(A_i)_{i\in\mathbb{N}} \subset \mathcal{F}$ então $\cup_i A_i \in \mathcal{F}$.

E a função $\mathbb{P}$ satisfaz:

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(\Omega) = 1$.
- Se $(A_i)_{i\in\mathbb{N}} \subset \mathcal{F}$ é formada por eventos disjuntos então

$$\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i).$$

Fatos Básicos

Definição

Dizemos que $X : \Omega \to \mathbb{R}$ *é uma variável aleatória se para todo conjunto de Borel* B *de* $\mathbb{R}$*, temos que*

$$X^{-1}(B) = \{w \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

Fatos Básicos

Definição

Dizemos que $X : \Omega \to \mathbb{R}$ *é uma variável aleatória se para todo conjunto de Borel* B *de* $\mathbb{R}$*, temos que*

$$X^{-1}(B) = \{w \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

Vamos denotar o conjunto $X^{-1}(B)$ *como*

$$X^{-1}(B) := \{X \in B\}.$$

Fatos Básicos

Definição

Dado uma variável aleatória X, se existir uma função f_X tal que

$$\mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(y)dy,$$

dizemos que f_X é **função de densidade** *para X.*

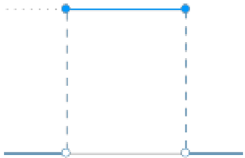
Fatos Básicos

Exemplo

Dizemos que X *tem distribuição uniforme no intervalo* $(0,1)$ *se*

$$\mathbb{P}(X \leq t) = \int_{-\infty}^{t} 1_{(0,1)}(y)dy.$$

Onde $1_{(0,1)}(y) = 1$ *se* $y \in (0,1)$ *e* $1_{(0,1)}(y) = 0$ *se* $y \notin (0,1)$.



Fatos Básicos

Exemplo

Dizemos que X tem distribuição normal com média 0 e variância 1 se

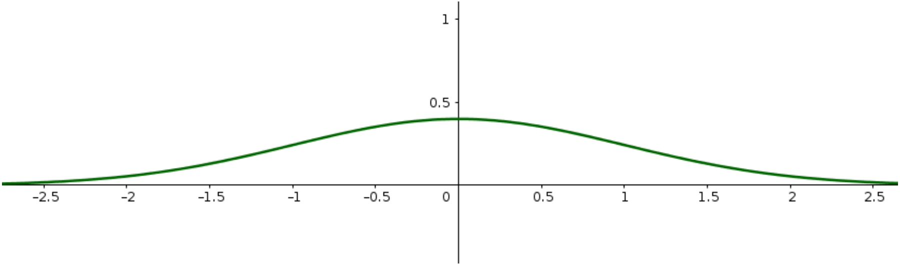
$$\mathbb{P}(X \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-y^2/2} dy.$$

Fatos Básicos

Exemplo

Dizemos que X tem distribuição normal com média 0 e variância 1 se

$$\mathbb{P}(X \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-y^2/2} dy.$$



Fatos Básicos

Exemplo

Dizemos que X tem distribuição normal com média μ e variância σ^2 se

$$\mathbb{P}(X \leq t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-(y-\mu)^2/(2\sigma^2)} dy.$$

Exemplo

Dizemos que X *tem distribuição normal com média* μ *e variância* σ^2 *se*

$$\mathbb{P}(X \leq t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-(y-\mu)^2/(2\sigma^2)} dy.$$

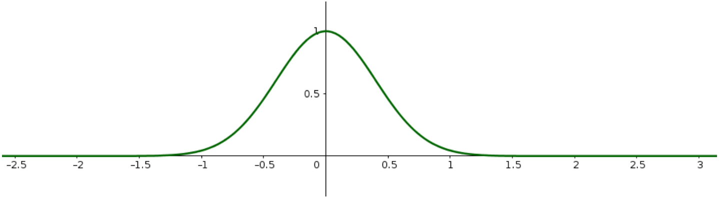


Figura: Média 0 e variância 0.4

Fatos Básicos

Exemplo

Dizemos que X tem distribuição Rademacher se

$$\mathbb{P}(X = 1) = 1/2$$
$$\mathbb{P}(X = -1) = 1/2.$$

Fatos Básicos

Exemplo

Dizemos que X tem distribuição Rademacher se

$$\mathbb{P}(X = 1) = 1/2$$
$$\mathbb{P}(X = -1) = 1/2.$$

Isto é, jogamos uma moeda, se esta cair como cara, então $X = 1$, caso contrário, $X = -1$.

Fatos Básicos

Definição

Dada uma variável aleatória X, definimos o seu valor esperado como

$$\mathbb{E}(X) = \int_{\omega\in\Omega} X(\omega) d\mathbb{P}(\omega).$$

Se X possuir uma função de distribuição, também podemos escrever a expressão acima como

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

Fatos Básicos

Mais geralmente, temos o seguinte resultado:

Lema

Seja X uma variável aleatória com função de densidade f_X. Seja g uma função mensurável qualquer, então

$$\mathbb{E}(g(X)) = \int g(x) f_X(x) dx.$$

Fatos Básicos

Definição

Dada uma variável aleatória X, definimos a sua variância como

$$\mathbb{V}ar(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

Fatos Básicos

Definição

Dada uma variável aleatória X, definimos a sua variância como

$$\mathbb{V}ar(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

Ou seja, pelo lema anterior, temos que

$$\mathbb{V}ar(X) = \mathbb{E}(g(X)),$$

onde $g(x) = (x - \mathbb{E}(X))^2$.

Fatos Básicos

Definição

Dada uma variável aleatória X, definimos a sua variância como

$$\mathbb{V}ar(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

Ou seja, pelo lema anterior, temos que

$$\mathbb{V}ar(X) = \mathbb{E}(g(X)),$$

onde $g(x) = (x - \mathbb{E}(X))^2$.
É fácil mostrar que

$$\mathbb{V}ar(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Fatos Básicos

Exemplo

Se X tem distribuição Rademacher, então $\mathbb{E}(X) =$

Fatos Básicos

Exemplo

Se X tem distribuição Rademacher, então $\mathbb{E}(X) = 0$ *e* $\mathbb{V}ar(X) = 1$.

Fatos Básicos

Exemplo

Se X *tem distribuição Rademacher, então* $\mathbb{E}(X) = 0$ *e* $\mathbb{V}ar(X) = 1$.

Exemplo

Se X *tem distribuição uniforme no intervalo* $(0, 1)$, *então* $\mathbb{E}(X) =$

Fatos Básicos

Exemplo

Se X tem distribuição Rademacher, então $\mathbb{E}(X) = 0$ e $\mathbb{V}ar(X) = 1$.

Exemplo

Se X tem distribuição uniforme no intervalo $(0,1)$, então $\mathbb{E}(X) = 1/2$ e $\mathbb{V}ar(X) = 1/12$.

Fatos Básicos

Exemplo

Se X tem distribuição Rademacher, então $\mathbb{E}(X) = 0$ *e* $\mathbb{V}ar(X) = 1$.

Exemplo

Se X tem distribuição uniforme no intervalo $(0,1)$, *então* $\mathbb{E}(X) = 1/2$ *e* $\mathbb{V}ar(X) = 1/12$.

Exemplo

Se X tem distribuição normal com média μ e variância σ^2, então $\mathbb{E}(X) =$

Fatos Básicos

Exemplo

Se X tem distribuição Rademacher, então $\mathbb{E}(X) = 0$ e $\mathbb{V}ar(X) = 1$.

Exemplo

Se X tem distribuição uniforme no intervalo $(0, 1)$, então $\mathbb{E}(X) = 1/2$ e $\mathbb{V}ar(X) = 1/12$.

Exemplo

Se X tem distribuição normal com média μ e variância σ^2, então $\mathbb{E}(X) = \mu$ e $\mathbb{V}ar(X) = \sigma^2$.

Nota para os Analistas

I had a not atypical functional analyst's suspicion of probability theory as nothing more than a subset of functional analysis with strange names whose purpose is to confuse those who have not bothered to join the club. [...] But this somehow misses the point. Like any other foreign language, probability theory is structured around its own natural thought patterns and is critical to a mode of thinking;

Barry Simon

Independência

Definição

Dizemos que duas variáveis aleatórias X, Y são independentes se

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Independência

Definição

Dizemos que duas variáveis aleatórias X, Y são independentes se

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Lema

Se X, Y são independentes, então

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Independência

Definição

Dizemos que duas variáveis aleatórias X, Y são independentes se

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Lema

Se X, Y são independentes, então

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Lema

Se X, Y são independentes, então

$$\mathbb{V}ar(X + Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y).$$

Uma Aplicação

Teorema

Considere $v_1, \ldots, v_n \in \mathbb{R}^n$ *tal que* $|v_i| = 1$, *para todo* $i = 1, \ldots, n$. *Então, existem* $\epsilon_i \in \{1, -1\}$ *tal que*

$$|\epsilon_1 v_1 + \cdots + \epsilon_n v_n| \leq \sqrt{n}.$$

Uma Aplicação

Demonstração.

Considere $\epsilon_1, \ldots, \epsilon_n$ variáveis Rademachers independentes e defina

$$X = |\epsilon_1 v_1 + \cdots + \epsilon_n v_n|^2.$$

Uma Aplicação

Demonstração.

Considere $\epsilon_1, \ldots, \epsilon_n$ variáveis Rademachers independentes e defina

$$X = |\epsilon_1 v_1 + \cdots + \epsilon_n v_n|^2.$$

Então

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle v_i, v_j \rangle \mathbb{E}(\epsilon_i \epsilon_j) \\ &= \sum_{i=1}^{n} |v_1| = n.\end{aligned}$$

Uma Aplicação

Demonstração.

Considere $\epsilon_1, \ldots, \epsilon_n$ variáveis Rademachers independentes e defina

$$X = |\epsilon_1 v_1 + \cdots + \epsilon_n v_n|^2.$$

Então

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle v_i, v_j \rangle \mathbb{E}(\epsilon_i \epsilon_j) \\ &= \sum_{i=1}^{n} |v_1| = n.\end{aligned}$$

Já que as variáveis são independentes e portanto, se $i \neq j$, temos que $\mathbb{E}(\epsilon_i \epsilon_j) = \mathbb{E}(\epsilon_i)\mathbb{E}(\epsilon_j) = 0$, e se $i = j$, temos que

$$\mathbb{E}(\epsilon_i \epsilon_j) = \mathbb{E}(\epsilon_i^2) = \mathbb{V}\text{ar}(\varepsilon_i) = 1.$$

Demonstração.

Como $X = |\epsilon_1 v_1 + \cdots + \epsilon_n v_n|^2$, e

$$\mathbb{E}(X) = n,$$

deve existir $\epsilon_1, \ldots, \epsilon_n \in \{1, -1\}$ específicos tal que $X \leq n$, já que caso contrário, teríamos que

$$\mathbb{E}(X) > n.$$

Lei dos Grandes Números

Teorema

Seja $X_1, X_2, \ldots$ *uma sequência de variáveis aleatórias independentes e com a mesma distribuição. Suponha que* $\mathbb{E}(X_1) = \mu$ *e* $\mathbb{E}(|X_1|) < \infty$. *Então,*

$$\frac{S_n}{n} := \frac{X_1 + \cdots + X_n}{n} \longrightarrow_n \mu, \;\; q.t.p..$$

Lei dos Grandes Números

Teorema

Seja $X_1, X_2, \ldots$ uma sequência de variáveis aleatórias independentes e com a mesma distribuição. Suponha que $\mathbb{E}(X_1) = \mu$ e $\mathbb{E}(|X_1|) < \infty$. Então,

$$\frac{S_n}{n} := \frac{X_1 + \cdots + X_n}{n} \longrightarrow_n \mu, \quad q.t.p..$$

(Monte Carlo): Suponha que temos uma moeda com probabilidade $p \in [0, 1]$ de ser cara.

Lei dos Grandes Números

Teorema

Seja $X_1, X_2, \ldots$ uma sequência de variáveis aleatórias independentes e com a mesma distribuição. Suponha que $\mathbb{E}(X_1) = \mu$ e $\mathbb{E}(|X_1|) < \infty$. Então,

$$\frac{S_n}{n} := \frac{X_1 + \cdots + X_n}{n} \longrightarrow_n \mu, \ \ q.t.p..$$

(Monte Carlo): Suponha que temos uma moeda com probabilidade $p \in [0,1]$ de ser cara. Uma forma de estimar a probabilidade p é jogar essa moeda uma quantidade n (muito grande) de vezes, contar quantos lançamentos foram cara, e dividir por n.

Problema

Vamos estimar (utilizando Python) o valor de π através do cálculo do volume de uma esfera unitária de dimensão d utilizando Monte Carlo.

Problema

Vamos estimar (utilizando Python) o valor de π através do cálculo do volume de uma esfera unitária de dimensão d utilizando Monte Carlo. Sabemos que o volume da esfera unitária de dimensão d é dado pela fórmula:

$$V(d) = \frac{2\pi^{d/2}}{d\Gamma(d/2)}.$$

Para isso, vamos simular uma sequência de vetores aleatórios

$$X_n = (X_1^n, \dots, X_d^n)$$

onde as coordenadas X_i são variáveis aleatórias independentes e com distribuição uniforme em $(0, 1)$.

A Maldição da Dimensão: Capítulo 1

Para isso, vamos simular uma sequência de vetores aleatórios

$$X_n = (X_1^n, \dots, X_d^n)$$

onde as coordenadas X_i são variáveis aleatórias independentes e com distribuição uniforme em $(0, 1)$.
Então vamos contar quantos desses vetores estão na esfera unitária, e depois vamos dividir essa soma por n e utilizar a fórmula do volume da esfera para estimar π.

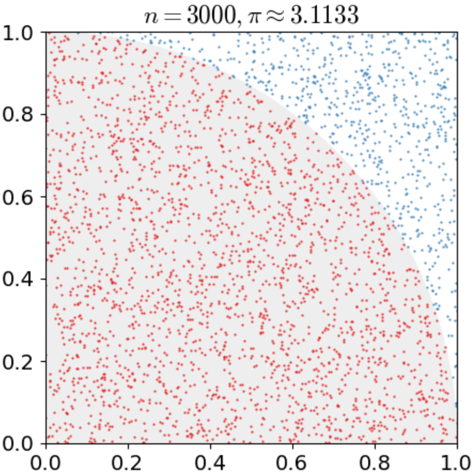
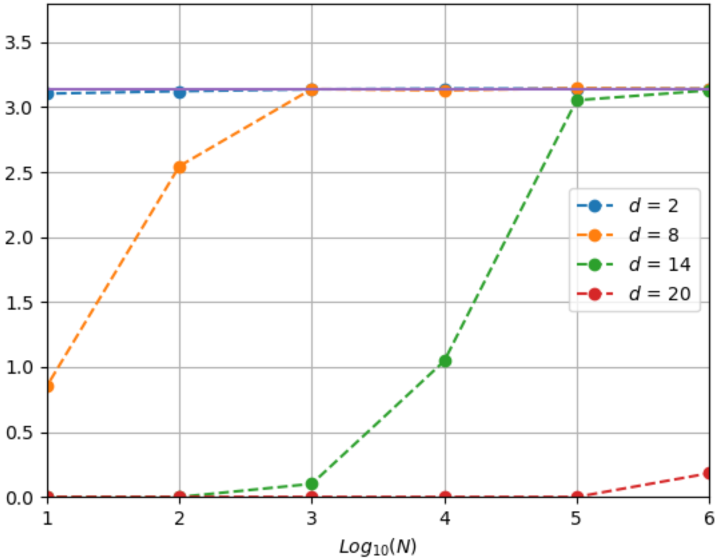


Figura: Exemplo para $d = 2$.

A Maldição da Dimensão: Capítulo 1



Problema

Dado um vetor aleatório,

$$X = (X_1, \dots, X_d)$$

com coordenadas independentes, onde cada X_i tem distribuição normal de média 0 e variância 1 *qual o valor esperado de*

$$\|X\|_2 = \sqrt{X_1^2 + \cdots + X_d^2} \text{ ?}$$

- **Intuição:** Note que

$$\mathbb{E}(\|X\|_2^2) = \sum_{i=1}^{d} \mathbb{E}(X_i^2) = d.$$

- **Intuição:** Note que

$$\mathbb{E}(\|X\|_2^2) = \sum_{i=1}^{d} \mathbb{E}(X_i^2) = d.$$

Logo, um chute natural seria dizer que

$$\|X\|_2 \approx \sqrt{d}.$$

- **Intuição:** Além disso, decompondo X como

$$X = \|X\|_2 \cdot \frac{X}{\|X\|_2}.$$

- **Intuição:** Além disso, decompondo X como

$$X = \|X\|_2 \cdot \frac{X}{\|X\|_2}.$$

Dessa forma, $\dfrac{X}{\|X\|_2}$ nada mais é que um ponto uniformemente distribuído na esfera unitária de dimensão $d - 1$.

- **Intuição:** Além disso, decompondo X como

$$X = \|X\|_2 \cdot \frac{X}{\|X\|_2}.$$

Dessa forma, $\frac{X}{\|X\|_2}$ nada mais é que um ponto uniformemente distribuído na esfera unitária de dimensão $d-1$.

Como $\|X\|_2 \approx \sqrt{d}$, é esperado que X se comporte como um ponto uniformemente distribuído na esfera de dimensão $d-1$ com raio $\sqrt{d}$.

Figura: Ilustração do caso $d = 2$ (esquerda) e d muito grande (direita).

Problema

Dados vetores aleatórios, independentes

$$X = (X_1, \ldots, X_d)$$

$$Y = (Y_1, \ldots, Y_d)$$

com coordenadas independentes, onde cada coordenada tem distribuição normal de média 0 e variância 1 *qual o valor esperado de*

$$\langle X, Y \rangle ?$$

Lema

Seja

$$X = (X_1, \ldots, X_d)$$

um vetor aleatório com entradas independentes e com distribuição normal com média 0 *e variância* 1. *Então*

$$\mathbb{E}\langle X, x\rangle^2 = \|x\|_2^2.$$

Demonstração.

É fácil ver que $\mathbb{E}(XX^t) = I_d$, já que a entrada (i, j) da matriz $\mathbb{E}(XX^t)$ é da forma $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j), \ i \neq j$.
Logo,

$$\mathbb{E}(\langle X, x\rangle^2) = x^t \mathbb{E}(XX^t) x = x^t x = \|x\|_2^2.$$

Como X, Y são independentes, temos que

$$\begin{aligned}\mathbb{E}(\langle X, Y\rangle^2) &= \mathbb{E}_Y\mathbb{E}_X(\langle X, Y\rangle^2 | Y)\\ &= \mathbb{E}_Y\|Y\|_2^2 = d.\end{aligned}$$

A Maldição da Dimensão: Capítulo 3

Como X, Y são independentes, temos que

$$\begin{aligned}\mathbb{E}(\langle X, Y\rangle^2) &= \mathbb{E}_Y\mathbb{E}_X(\langle X, Y\rangle^2|Y) \\ &= \mathbb{E}_Y\|Y\|_2^2 = d.\end{aligned}$$

Note que então, o valor do ângulo esperado entre os vetores aleatórios X, Y, normalizados, é da forma

$$\left|\langle \frac{X}{\|X\|_2}, \frac{Y}{\|Y\|_2}\rangle\right| \approx \frac{\sqrt{d}}{\sqrt{d}\cdot\sqrt{d}} = \frac{1}{\sqrt{d}}.$$

Desigualdades de Concentração

Os resultados anteriores só funcionam se as variáveis aleatórias se concentram em suas respectivas médias.

Desigualdades de Concentração

Os resultados anteriores só funcionam se as variáveis aleatórias se concentram em suas respectivas médias.

Definição

Dizemos que uma variável aleatória X com média finita é sub-gaussiana, se existe uma constante K tal que

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq 2e^{-t^2/K^2}.$$

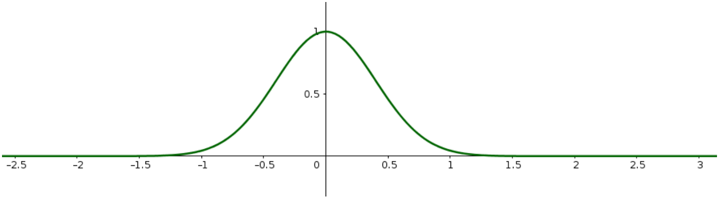


Figura: Decaimento sub-gaussiano

Desigualdades de Concentração

Lema (Markov)

Seja X variável aleatória positiva com média finita. Então

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Lema (Markov)

Seja X variável aleatória positiva com média finita. Então

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Demonstração.

Temos que

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X1_{X\geq t}) + \mathbb{E}(X1_{X<t}) \\ &\geq t\mathbb{E}(1_{X\geq t}) + 0 \\ &= t\mathbb{P}(X \geq t).\end{aligned}$$

Desigualdades de Concentração

Lema (Chebyshev)

Seja X variável aleatória média e variância finita. Então

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathbb{V}ar(X)}{t^2}.$$

Desigualdades de Concentração

Lema (Chebyshev)

Seja X variável aleatória média e variância finita. Então

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathbb{V}ar(X)}{t^2}.$$

Demonstração.

Basta aplicar a desigualdade de Markov para $Y = (X - \mathbb{E}(X))^2$.

Desigualdades de Concentração

Lema (Chernoff)

Seja X variável aleatória tal que $\mathbb{E}(e^{sX})$ seja finito para todo $s \in \mathbb{R}$. Então

$$\mathbb{P}(X - \mathbb{E}(X) \geq t) \leq \frac{\mathbb{E}(e^{s(X-\mathbb{E}(X))})}{e^{st}}.$$

Desigualdades de Concentração

Lema (Chernoff)

Seja X variável aleatória tal que $\mathbb{E}(e^{sX})$ seja finito para todo $s \in \mathbb{R}$. Então

$$\mathbb{P}(X - \mathbb{E}(X) \geq t) \leq \frac{\mathbb{E}(e^{s(X-\mathbb{E}(X))})}{e^{st}}.$$

Demonstração.

Basta aplicar a desigualdade de Markov para $Y = e^{s(X-\mathbb{E}(X))}$.

Proposição

Para provar que X é sub-gaussiana, é suficiente mostrar que existe uma constante $K_1 > 0$ tal que

$$\mathbb{E}(e^{s(X-\mathbb{E}(X))}) \leq e^{K_1^2 s^2/2}.$$

Desigualdades de Concentração

Proposição

Para provar que X é sub-gaussiana, é suficiente mostrar que existe uma constante $K_1 > 0$ tal que

$$\mathbb{E}(e^{s(X-\mathbb{E}(X))}) \leq e^{K_1^2 s^2/2}.$$

Demonstração.

Substituindo o limitante acima na expressão, temos

$$\mathbb{P}(X - \mathbb{E}(X) \geq t) \leq \frac{\mathbb{E}(e^{s(X-\mathbb{E}(X))})}{e^{st}} \leq e^{K_1^2 s^2/2} e^{-st}.$$

Minimizando em $s \geq 0$, temos algo do tipo

$$\mathbb{P}(X - \mathbb{E}(X) \geq t) \leq e^{-t^2/K^2}.$$

Exemplo

Se X tem distribuição normal com média 0 e variância σ^2, então X é sub-gaussiana, já que

$$\mathbb{E}(e^{sX}) = e^{\sigma^2 s^2/2}.$$

Desigualdades de Concentração

Exemplo

Se X tem distribuição Rademacher, então X é sub-gaussiana.

Desigualdades de Concentração

Exemplo

Se X tem distribuição Rademacher, então X é sub-gaussiana.

Demonstração.

$$\begin{aligned}
\mathbb{E}(e^{sX}) =& \mathbb{P}(X = 1)e^{1\cdot s} + \mathbb{P}(X = -1)e^{-1\cdot s} \\
=& \frac{1}{2}(e^{-s} + e^{s}) \\
=& \sum_{i=0}^{\infty} \frac{s^{2i}}{(2i)!} \\
\leq& 1 + \sum_{i=0}^{\infty} \frac{s^{2i}}{2^k \cdot k!} \\
=& e^{s^2/2}.
\end{aligned}$$

Desigualdades de Concentração

Exemplo

Seja X variável aleatória com média 0 e tal que $X \in [a, b]$, então X é sub-gaussiana.

Exemplo

Seja X variável aleatória com média 0 e tal que $X \in [a, b]$, então X é sub-gaussiana.

Demonstração.

Seja X' uma cópia independente de X. É claro que $\mathbb{E}(X') = 0$. Então

$$\begin{aligned} \mathbb{E}_X(e^{sX}) &= \mathbb{E}_X(e^{s(X - \mathbb{E}_{X'}(X'))}) = \mathbb{E}_X(e^{\mathbb{E}_{X'}((s(X-X'))}) \\ &\leq \mathbb{E}_X \mathbb{E}_{X'}(e^{s(X-X')}) \end{aligned}$$

Exemplo

Seja X variável aleatória com média 0 e tal que $X \in [a, b]$, então X é sub-gaussiana.

Demonstração.

Seja X' uma cópia independente de X. É claro que $\mathbb{E}(X') = 0$. Então

$$\begin{aligned}\mathbb{E}_X(e^{sX}) &= \mathbb{E}_X(e^{s(X-\mathbb{E}_{X'}(X')}) = \mathbb{E}_X(e^{\mathbb{E}_{X'}((s(X-X'))}) \\ &\leq \mathbb{E}_X\mathbb{E}_{X'}(e^{s(X-X')})\end{aligned}$$

Seja Y uma variável aleatória com distribuição Rademacher, independente de X e X', então

$$\mathbb{E}_X\mathbb{E}_{X'}(e^{s(X-X')}) = \mathbb{E}_X\mathbb{E}_{X'}\mathbb{E}_Y(e^{sY(X-X')})$$

Demonstração.

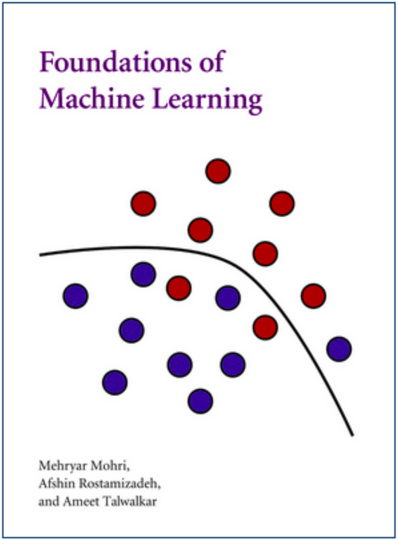
Temos então que

$$\mathbb{E}_X(e^{sX}) = \mathbb{E}_X \mathbb{E}_{X'} \mathbb{E}_Y(e^{sY(X-X')})$$

mas pelo exemplo anterior e o fato de $X \in [a, b]$, temos que

$$\begin{aligned} \mathbb{E}_X \mathbb{E}_{X'} \mathbb{E}_Y(e^{s(X-X')Y}) &\leq \mathbb{E}_X \mathbb{E}_{X'}(e^{s^2(X-X')^2/2}) \\ &= e^{s^2(b-a)^2/2}. \end{aligned}$$

Matemática em prol da Skynet



Matemática em prol da Skynet

Aplicações em Machine Learning:

1. Support Vector Machines
2. Métodos de Kernel
3. Boosting
4. Lasso/Ridge Regression
5. Redes Neurais

Matemática em prol da Skynet: Classificando Carros

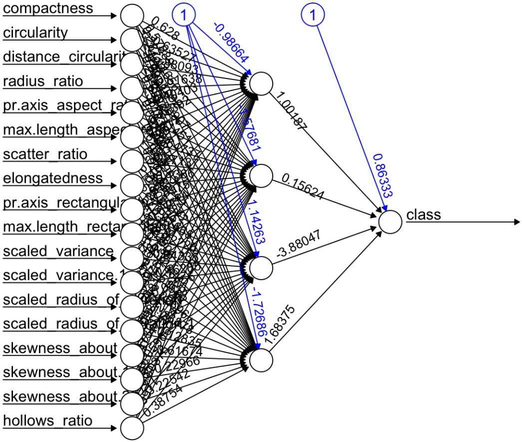


Figura: Informações (18) sobre medidas de silhuetas de carros. Saída é um carro normal ou um ônibus. Taxa de acerto por volta de 50%.

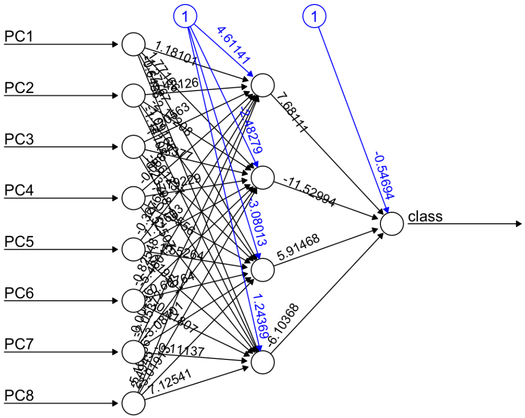


Figura: Utilizando uma técnica de redução de dimensão (PCA), conseguimos uma taxa de acerto por volta de 95%.

Referências

- High-Dimensional Probability: An Introduction with Applications in Data Science - Vershynin, R.
- High-Dimensional Statistics: A Non-Asymptotic Viewpoint - Wainwright, M.
- The Probabilistic Method - Alon, N; Spencer, J.
- https://medium.com/@amoldeshmukh137
- https://towardsdatascience.com