Métodos geométricos aplicados a ciências da informação

Fábio C. C. Meneghetti

Instituto de Matemática, Estatística e Computação Científica
Universidade Estadual de Campinas

22 de fevereiro de 2022

Parte I

Motivacao

História

- Claude Shannon (1916–2001): entropia como medida de informação de uma variável aleatória: $H(X) = \sum_i p_i \log p_i$.

História

- Claude Shannon (1916–2001): entropia como medida de informação de uma variável aleatória: $H(X) = \sum_i p_i \log p_i$.
- Fisher (1890–1962): medida de informação sobre um parâmetro que uma variável aleatória carrega: $I(\ ) = \int_X p\ (x)\ \log \frac{\mathrm{d}p}{\mathrm{d}}(x)\ ^2 \mathrm{d}x$

História

- Claude Shannon (1916–2001): entropia como medida de informação de uma variável aleatória: $H(X) = \sum_i p_i \log p_i$.
- Fisher (1890–1962): medida de informação sobre um parâmetro que uma variável aleatória carrega: $I(\ ) = \int_X p\ (x)\ \log \frac{\mathrm{d}p}{\mathrm{d}}(x)\ ^2 \mathrm{d}x$
- C. R. Rao (1920–): a informação de Fisher sobre múltiplos parâmetros $(\ _1; \ldots; \ _d)$ é uma métrica Riemanniana!

História

- Claude Shannon (1916–2001): entropia como medida de informação de uma variável aleatória: $H(X) = \sum_i p_i \log p_i$.
- Fisher (1890–1962): medida de informação sobre um parâmetro que uma variável aleatória carrega: $I(\ ) = \int_X p\ (x)\ \log \frac{\mathrm{d}p}{\mathrm{d}}(x)\ ^2 \mathrm{d}x$
- C. R. Rao (1920–): a informação de Fisher sobre múltiplos parâmetros $(\ _1; \ldots; \ _d)$ é uma métrica Riemanniana!
  - *metrica da informacao de Fisher*

História

- Claude Shannon (1916–2001): entropia como medida de informação de uma variável aleatória: $H(X) = \sum_i p_i \log p_i$.
- Fisher (1890–1962): medida de informação sobre um parâmetro que uma variável aleatória carrega: $I(\ ) = \int_X p\ (x)\ \log \frac{\mathrm{d}p}{\mathrm{d}}(x)\ ^2 \mathrm{d}x$
- C. R. Rao (1920–): a informação de Fisher sobre múltiplos parâmetros $(\ _1; \ldots; \ _d)$ é uma métrica Riemanniana!
  - *metrica da informacao de Fisher*
  - ela induz uma geometria sobre as distribuições de probabilidade

História

- Claude Shannon (1916–2001): entropia como medida de informação de uma variável aleatória: $H(X) = \sum_i p_i \log p_i$.
- Fisher (1890–1962): medida de informação sobre um parâmetro que uma variável aleatória carrega: $I(\ ) = \int_X p\ (x)\ \log \frac{\mathrm{d}p}{\mathrm{d}}(x)\ ^2 \mathrm{d}x$
- C. R. Rao (1920–): a informação de Fisher sobre múltiplos parâmetros $(\ _1; \ldots; \ _d)$ é uma métrica Riemanniana!
  - *metrica da informacao de Fisher*
  - ela induz uma geometria sobre as distribuições de probabilidade
  - isso nos permite falar sobre distâncias e curvaturas no espaço das distribuições

História

- Claude Shannon (1916–2001): entropia como medida de informação de uma variável aleatória: $H(X) = \sum_i p_i \log p_i$.
- Fisher (1890–1962): medida de informação sobre um parâmetro que uma variável aleatória carrega: $I(\ ) = \int_X p\ (x)\ \log \frac{\mathrm{d}p}{\mathrm{d}}(x)\ ^2 \mathrm{d}x$
- C. R. Rao (1920–): a informação de Fisher sobre múltiplos parâmetros $(\ _1; \ldots; \ _d)$ é uma métrica Riemanniana!
  - *metrica da informacao de Fisher*
  - ela induz uma geometria sobre as distribuições de probabilidade
  - isso nos permite falar sobre distâncias e curvaturas no espaço das distribuições
  - área de pesquisa: *geometria da informacao*

Por exemplo:

- no caso das distribuições gaussianas univariadas...

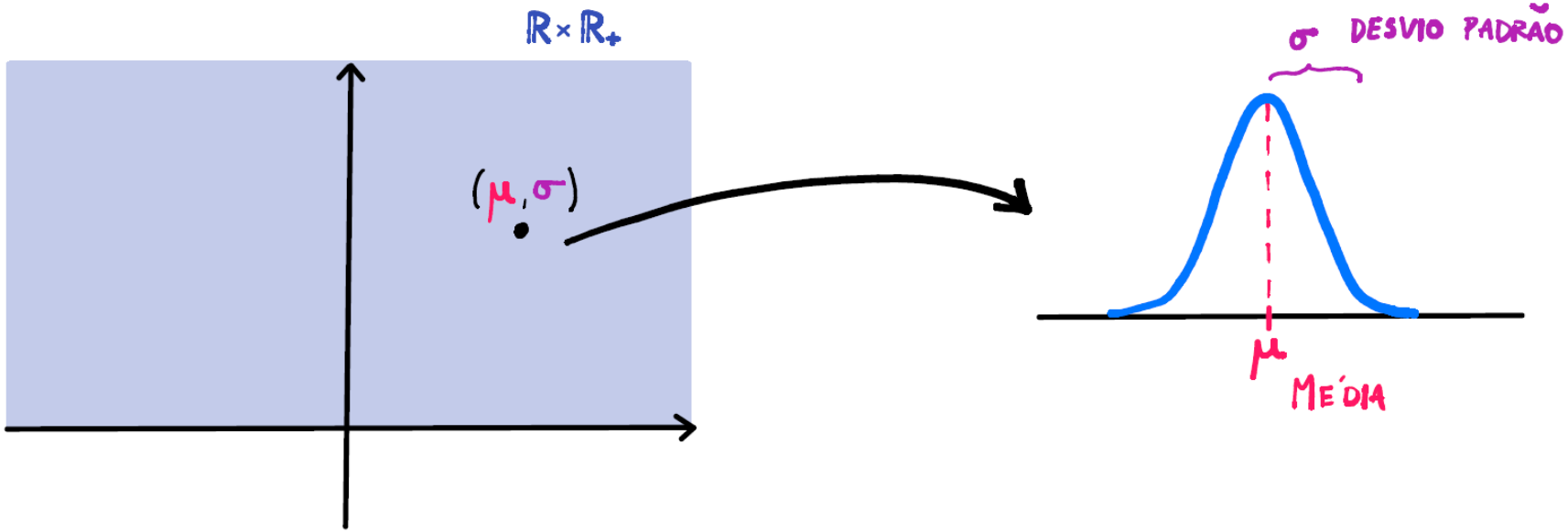
$$p(x; \ ; \ ) = \frac{1}{\sqrt{2}} \exp \left(\frac{jx \quad j^2}{2 \ ^2} \right) ;$$

Por exemplo:

- no caso das distribuições gaussianas univariadas...

$$p(x; \quad ; \quad) = \frac{1}{2} \exp \left(\frac{jx \qquad j^2}{2 \quad ^2} \right)^{!} \quad ;$$

o espaço de parâmetros é



- com a métrica da informação de Fisher, obtemos uma geometria hiperbólica! (versão deformada do meio-plano de Poincaré)

- com a métrica da informação de Fisher, obtemos uma geometria hiperbólica! (versão deformada do meio-plano de Poincaré)

Parte II

Teoria

Distribuicoes de probabilidade

distribuicoes de probabilidade

Distribuicoes de probabilidade

distribuicoes de probabilidade

discretas: $X \in \{x_1, \ldots, x_n\}$, $P[X = x_i] = p_i$, $\sum_i p_i = 1$.

Distribuicoes de probabilidade

distribuicoes de probabilidade

discretas: $X \in \{x_1, \ldots, x_n\}$, $P[X = x_i] = p_i$, $\sum_i p_i = 1$.

contnuas: $X \in \mathcal{X}$, $P[X \in A] = \int_A p(x)\,dx$.

Distribuicoes de probabilidade

distribuicoes de probabilidade

discretas: $X \in \{x_1, \ldots, x_n\}$, $P[X = x_i] = p_i$, $\sum_i p_i = 1$.

contnuas: $X \in \mathcal{X}$, $P[X \in A] = \int_A p(x)\,dx$.

de forma mais geral, temos um espaco de probabilidade $(\mathcal{X}, \mathcal{F}, P)$, e uma medida - nita dominante .

Distribuic-oes de probabilidade

distribuic-oes de probabilidade

discretas: $X \in \{x_1, \ldots, x_n\}$, $P[X = x_i] = p_i$, $\sum_i p_i = 1$.

contnuas: $X \in \mathcal{X}$, $P[X \in A] = \int_A p(x)\,dx$.

de forma mais geral, temos um espaco de probabilidade $(\mathcal{X}, \mathcal{F}, P)$, e uma medida - nita dominante .

A funcao densidade e a derivada de Radon-Nikodym $p(x) = \frac{dP}{d}(x)$, $p : X \to \mathbb{R}_+$ que satisfaz $P(A) = \int_A p(x)\,d(x)$.

Distribuic-oes de probabilidade

distribuic-oes de probabilidade

discretas: $X \in \{x_1, \ldots, x_n\}$, $P[X = x_i] = p_i$, $\sum_i p_i = 1$.

cont nuas: $X \in \mathcal{X}$, $P[X \in A] = \int_A p(x)\,dx$.

de forma mais geral, temos um espac o de probabilidade $(\mathcal{X}, \mathcal{F}, P)$, e uma medida - nita dominante .

A func-ao densidade e a derivada de Radon-Nikodym $p(x) = \frac{dP}{d}(x)$, $p : \mathcal{X} \to \mathbb{R}_+$ que satisfaz $P(A) = \int_A p(x)\,d(x)$.

$\mathcal{X}$ enumer avel e c medida de contagem $\Rightarrow$ distribuic-ao discreta, p = func-ao massa, $\int_{x \in A} = \sum_{x \in A}$

Distribuic-oes de probabilidade

distribuic-oes de probabilidade

discretas: $X \in \{x_1, \ldots, x_n\}$, $P[X = x_i] = p_i$, $\sum_i p_i = 1$.

contnuas: $X \in \mathcal{X}$, $P[X \in A] = \int_A p(x)\,dx$.

de forma mais geral, temos um espaco de probabilidade $(\mathcal{X}, \mathcal{F}, P)$, e uma medida - nita dominante .

A func-ao densidade e a derivada de Radon-Nikodym $p(x) = \frac{dP}{d}(x)$, $p : X \to \mathbb{R}_+$ que satisfaz $P(A) = \int_A p(x)\,d(x)$.

X enumeravel e c medida de contagem $) =$ distribuic-ao discreta, $p =$ func-ao massa, $\int_{x \in A} = \sum_{x \in A}$

$X \subseteq \mathbb{R}^n$ e L medida de Lebesgue $) =$ distribuic-ao contnua, $p =$ func-ao densidade

Modelos estatsticos

um modelo estatstico e uma famlia parametrizada de distribuicoes $P = \{P_\theta : \theta \in \Theta\}$ no espaco $(X, \mathcal{F})$.

Modelos estatsticos

um modelo estatstico e uma famlia parametrizada de distribuicoes $P = \{P_\theta : \theta \in \Theta\}$ no espaco (X, F).

R^d conjunto aberto de parâmetros.

Modelos estatsticos

um modelo estatstico e uma famlia parametrizada de distribuicoes $P = \{P_\theta : \theta \in \Theta\}$ no espaco (X, F).

$\Theta \subseteq R^d$ conjunto aberto de parâmetros.

na vida real costuma-se tomar como modelo a famlia de funcoes densidade $p_\theta = \frac{dP_\theta}{d\mu} : \theta \in \Theta$.

Modelos estatsticos

um modelo estatstico e uma famlia parametrizada de distribuicoes $P = \{P_\theta : \theta \in \Theta\}$ no espaco $(X, \mathcal{F})$.

$\Theta \subset \mathbb{R}^d$ conjunto aberto de parâmetros.

na vida real costuma-se tomar como modelo a famlia de funcoes densidade $p_\theta = \frac{dP_\theta}{d\mu} : \theta \in \Theta$.

vamos considerar modelos estatsticos regulares

Modelos estatsticos

um modelo estatstico e uma famlia parametrizada de distribuicoes $P = \{P_\theta : \theta \in \Theta\}$ no espaco $(X, \mathcal{F})$.

$\Theta \subset \mathbb{R}^d$ conjunto aberto de parâmetros.

na vida real costuma-se tomar como modelo a famlia de funcoes densidade $p_\theta = \frac{dP_\theta}{d\mu} : \theta \in \Theta$.

vamos considerar modelos estatsticos regulares

a funcao $\theta \mapsto p_\theta$ e C^1

Modelos estatsticos

um modelo estatstico e uma famlia parametrizada de distribuic-oes $P = \{P_\theta : \theta \in \Theta\}$ no espaco $(X, \mathcal{F})$.

$\Theta \subset \mathbb{R}^d$ conjunto aberto de parâmetros.

na vida real costuma-se tomar como modelo a famlia de func-oes densidade $p_\theta = \frac{dP_\theta}{d\mu} : \theta \in \Theta$.

vamos considerar modelos estatsticos regulares

a func-ao $\theta \mapsto p_\theta$ e C^1

$p_\theta(x) > 0$ para todo x;

Metrica Riemanniana

Vamos agora adicionar uma estrutura Riemanniana ao espaco P .

Metrica Riemanniana

Vamos agora adicionar uma estrutura Riemanniana ao espacod P .

denote por` (x) := log p (x) a funcao log-probabilidade

Metrica Riemanniana

Vamos agora adicionar uma estrutura Riemanniana ao espacodP .

denote por` (x) := log p (x) a funcao log-probabilidade

a metrica de Fisher e a metrica Riemanniana (produto interno em T) dada por

$$g(V;W) := \int_X p(x) \frac{\partial \ell(x)}{\partial V} \frac{\partial \ell(x)}{\partial W} d(x):$$

Metrica Riemanniana

Vamos agora adicionar uma estrutura Riemanniana ao espacod P .

denote por` (x) := log p (x) a funcao log-probabilidade

a metrica de Fisher e a metrica Riemanniana (produto interno em T) dada por

$$g(V;W) := \int_X p(x) \frac{@\grave{}(x)}{@V} \frac{@\grave{}(x)}{@W} d(x):$$

na base coordenada local $e_i = \frac{@}{@_i}$, a matriz da metrica e chamada dematriz de Fisher I( ), com elementos

$$g_{ij}( ) := g(e_i;e_j) = \int_X p(x) \frac{@\grave{}(x)}{@_i} \frac{@\grave{}(x)}{@_j} d(x)$$

Metrica Riemanniana

Vamos agora adicionar uma estrutura Riemanniana ao espac-o P .

denote por` (x) := log p (x) a func-ao log-probabilidade

a metrica de Fisher e a metrica Riemanniana (produto interno em T) dada por

$$g(V;W) := \int_X p(x) \frac{@`(x)}{@V} \frac{@`(x)}{@W} d(x):$$

na base coordenada local $e_i = \frac{@}{@_i}$, a matriz da metrica e chamada de matriz de Fisher I( ), com elementos

$$g_{ij}( ) := g(e_i;e_j) = \int_X p(x) \frac{@`(x)}{@_i} \frac{@`(x)}{@_j} d(x)$$

I( ) e simetrica e positiva-de nida.

Interpretacoes da metrica de Fisher

Em estatstica

Interpretac-oes da metrica de Fisher

Em estatstica

O escorede em x e o gradiente da func-ao log-probabilidade:

$$s(x) = r \log p(x) = \left(\frac{@`}{@_1}; : : : ; \frac{@`}{@_d} \right)^{>} (x)$$

mede sensibilidade a mudancas nos parâmetros da func-ao log-probabilidade

Interpretac-oes da m-etrica de Fisher

Em estatstica

O escorede em x e o gradiente da func-ao log-probabilidade:

$$s(x) = r \log p(x) = \left(\frac{@`}{@_1}; : : : ; \frac{@`}{@_d} \right)^{>} (x)$$

mede sensibilidade a mudanc-as nos par^ametros da func-ao log-probabilidade

A matriz de Fisher e a matriz de covari^ancia do escore:

$$I( ) = \operatorname{cov}(s ; s) = E[s \quad s^{>}]$$

e um limitante inferior para a covari^ancia de um estimador n-ao-enviesado^ (Limitante de Cram-er-Rao):

$$\operatorname{cov}(\hat{}) \quad I( )^{1}$$

em teoria da informacao

em teoria da informac-ao

entropia relativa (divergência de Kullback-Leibler):

$$D_{KL}(p\|q) = \int_X p(x) \log \frac{p(x)}{q(x)} d\ (x)$$

em teoria da informac-ao

entropia relativa (divergência de Kullback-Leibler):

$$D_{KL}(p\|q) = \int_X p(x) \log \frac{p(x)}{q(x)} d\ (x)$$

diz quantos bits (na verdade nats), em media, s-ao necessarios para codi car p usando um codigo otimizado para codi carq

- em teoria da informação
  - entropia relativa (divergência de Kullback-Leibler):
    $D_{\mathrm{KL}}(pkq) = \int_X p(x) \log \frac{p(x)}{q(x)} \mathrm{d}\ (x)$
    - diz quantos bits (na verdade nats), em media, sao necessarios para codi car p usando um codigo otimizado para codi car q
  - a matriz de Fisher é o Hessiano diagonal da entropia relativa:

$$g_{ij}(\ _0) = \frac{@^2}{@\ _i @\ _j} D_{\mathrm{KL}}(p\ _0 kp\ ) \Big|_{\ = \ _0}$$

Unicidade

uma estatsticae um mapa mensuravel : X ! Y

Unicidade

uma estatsticae um mapa mensuravel : X ! Y

induz uma famlia de distribuicoes de probabilidade empurradas P em Y: P (A) := P (^{1}(A))

Unicidade

uma estatsticae um mapa mensuravel : X ! Y

induz uma famlia de distribuic-oes de probabilidade empurradas P em Y: P (A) := P (1(A))

por sua vez, obtemos novas func-oes densidade = $\frac{d \quad P}{d}$

Unicidade

uma estatsticae um mapa mensuravel : X ! Y

induz uma famlia de distribuic-oes de probabilidade empurradas P em Y: P (A) := P (1(A))

por sua vez, obtemos novas func-oes densidade = d P / d

a estatstica e dita su ciente se p (x) = p ((x)) h(x) para alguma func-ao h independente de.

Unicidade

uma estatsticae um mapa mensuravel : X ! Y

induz uma famlia de distribuic-oes de probabilidade empurradas P em Y: P (A) := P (1(A))

por sua vez, obtemos novas func-oes densidade = d P / d

a estatstica e dita su ciente se p (x) = p ((x)) h(x) para alguma funcao h independente de .

Teorema (Chentsov)

A metrica de Fisher e a unica metrica Riemanniana, a menos de uma constante, invariante por estatsticas su cientes.

Geodesicas

geodesicas sao \linhas retas" nas variedades Riemannianas

Geodesicas

geodesicas sao "linhas retas" nas variedades Riemannianas

dados dois pontos p_1, p_0, a curva $\gamma : [0,1] \to P$ ligando-os, que minimiza comprimento, e um segumento de geodesica

Geodesicas

geodesicas s-ao \linhas retas" nas variedades Riemannianas

dados dois pontos p ; p_0, a curva $: [0; 1] ! P$ ligando-os, que minimiza comprimento, e um segumento de geodesica

o comprimento e dado por $`( ) = \int_0^1 \sqrt{g_{(t)}(\dot{}(t); \dot{}(t))}\, dt$

Geodesicas

geodesicas s-ao \linhas retas" nas variedades Riemannianas

dados dois pontos $p ; p_0$, a curva $: [0; 1] ! P$ ligando-os, que minimiza comprimento, e um segumento de geodesica

o comprimento e dado por $`( ) = \int_0^1 \sqrt{g_{(t)}(_(t); _(t))}\, dt$

essa noc-ao de ne uma distância geodesica, chamada distância de Fisher-Rao[1] na geometria da informac-ao:

$$d_{FR}(p ; p^0) = \min \{ `( ) : (0) = p ; (1) = p^0 \}$$

[1] quando a variedade e completa e conexa por caminhos

Geodesicas

geodesicas s-ao \linhas retas" nas variedades Riemannianas

dados dois pontos $p ; p_0$, a curva $: [0; 1] ! P$ ligando-os, que minimiza comprimento, e um segumento de geodesica

o comprimento e dado por $`( ) = \int_0^1 \sqrt{g_{(t)}(_(t); _(t))}\, dt$

essa noc-ao de ne uma distância geodesica, chamada distância de Fisher-Rao[1] na geometria da informac-ao:

$$d_{FR}(p ; p^0) = \min \{ `( ) : (0) = p ; (1) = p^0 \}$$

formalmente, s-ao curvas que têm derivada covariante zero: $\_ \_ \quad 0$

[1] quando a variedade e completa e conexa por caminhos

Geodesicas

geodesicas s-ao \linhas retas" nas variedades Riemannianas

dados dois pontos$p$; p_0, a curva $: [0; 1] ! P$ ligando-os, que minimiza comprimento, e um segumento de geodesica

o comprimento e dado por$`( ) = \int_0^1 \sqrt{g_{(t)}(\dot{}(t); \dot{}(t))}\, dt$

essa noc-ao de ne uma distância geodesica, chamada distância de Fisher-Rao na geometria da informac-ao:[1]

$$d_{FR}(p ; p^0) = \min \{ `( ) : (0) = p ; (1) = p^0 \}$$

formalmente, s-ao curvas que t^em derivada covariante zero: $\dot{}\dot{} \equiv 0$

r e a conex-ao dada pelos smbolos de Christo el

$_{ij;k}( ) = E_p [(@_i@_j` + \frac{1}{2}@_i` @_j` ) @_k`]$

[1] quando a variedade e completa e conexa por caminhos

Famlias exponenciais

uma famlia exponencialf p : 2 g com parâmetros naturais
2 R^d e dada por

$$p(x) := \exp[t(x); \ i \quad F( ) + k(x) \ ; \quad x \ 2 \ X$$

Famlias exponenciais

uma famlia exponencial $f_p : \in \Theta g$ com parâmetros naturais $\in \Theta \subset R^d$ e dada por

$$p_\theta(x) := \exp[\langle t(x), \theta \rangle - F(\theta) + k(x)] \quad ; \quad x \in X$$

$t(x)$ estatstica su ciente

Famlias exponenciais

uma famlia exponencial $f p : 2 g$ com parâmetros naturais $2 R^d$ e dada por

$$p (x) := \exp [t(x); i F( ) + k(x) ; x 2 X$$

$t(x)$ estatstica su ciente
$F( )$ func-ao estritamente convexa

Famlias exponenciais

uma famlia exponencial $f p : 2 g$ com parâmetros naturais $2 R^d$ e dada por

$$p (x) := \exp [t(x); i F( ) + k(x) ; x 2 X$$

$t(x)$ estatstica su ciente
$F( )$ func-ao estritamente convexa
$k(x)$ qualquer

Famlias exponenciais

uma famlia exponencial $f_p : \in \Theta g$ com parâmetros naturais $\in \Theta \subseteq R^d$ e dada por

$$p_\theta(x) := \exp[\langle t(x), \theta \rangle - F(\theta) + k(x)], \quad x \in X$$

$t(x)$ estatstica su ciente
$F(\theta)$ func-ao estritamente convexa
$k(x)$ qualquer

ha uma express-ao simples para a matriz de Fisher $g_{ij}(\theta) = \frac{\partial^2 F(\theta)}{\partial \theta_i \partial \theta_j}$

Exemplos

Famlias exponenciais englobam muitos casos

Exemplos

Famlias exponenciais englobam muitos casos

distribuicoes normais $t(x) = (x; x^2)$, $(\ _1; \ _2) = (\ _2; \frac{1}{2 \ _2^2})$,

$F = \frac{\ _1^2}{4 \ _2} + \frac{1}{2} \log \ _2$, $k(x) = 0$

Exemplos

Famlias exponenciais englobam muitos casos

distribuicoes normais $t(x) = (x; x^2)$, $(\ _1; \ _2) = (\ _2; \frac{1}{2 \ _2^2})$,

$F = \frac{\ _1^2}{4 \ _2} + \frac{1}{2} \log \ _2$, $k(x) = 0$

distribuicoes poisson $t(x) = x$, $k(x) = x!$, $\quad = \log \quad$, $F(\ ) = \quad = e$

Exemplos

Famlias exponenciais englobam muitos casos

distribuicoes normais $t(x) = (x; x^2)$, $(\ _1; \ _2) = (\ _2; \frac{1}{2 \ _2^2})$,

$F = \frac{\ _1^2}{4 \ _2} + \frac{1}{2} \log \ _2$, $k(x) = 0$

distribuicoes poisson $t(x) = x$, $k(x) = x!$, $\quad = \log \quad$, $F(\ ) = \quad = e$

gama, beta, exponencial, etc.

Parâmetros duais

famlias exponenciais têm parametrizac-oes duais: = r F( )

Parâmetros duais

famlias exponenciais têm parametrizac-oes duais: = r F( )
e possvel voltar para os parâmetros naturais via = r F ( )

Parâmetros duais

famlias exponenciais têm parametrizac-oes duais $\eta := r \quad F(\ )$
e possvel voltar para os parâmetros naturais via $= r \quad F\ (\ )$
onde $F\ (\ ) := h\ ;\ \ i \quad F(\ )$ e a transformada de Legendre

Parâmetros duais

famlias exponenciais têm parametrizac-oes duais $\eta := r\ F(\ )$
e possvel voltar para os parâmetros naturais via $= r\ F\ (\ )$
onde $F\ (_R) := h\ ;\ \ i\ \ F(\ )$ e a transformada de Legendre
$F\ (\ ) = \ _X\ p\ (x)\grave{}\ (x)\, d\ (x)$ e a entropia de Shannon negativa

Parâmetros duais

famlias exponenciais têm parametrizac-oes duais $\eta := r\ F(\ )$
e possvel voltar para os parâmetros naturais via $= r\ F\ (\ )$
onde $F\ (_R) := h\ ;\ \ i\ \ F(\ )$ e a transformada de Legendre
$F\ (\ ) = \ _X\ p\ (x)\grave{}\ (x)\, d\ (x)$ e a entropia de Shannon negativa

as parametrizac-oes e de fato s-ao duais, no sentido que:

$$e_i = \frac{@}{@_i}; \; e^j = \frac{@}{@_j} \; =) \quad g(e_i; e^j) = \; _{ij}.$$

as parametrizac-oes e de fato s-ao duais, no sentido que:

$$e_i = \frac{@}{@_i}; \; e^j = \frac{@}{@_j} \; =) \quad g(e_i; e^j) = \; _{ij}.$$

as parametrizac-oes e de fato s-ao duais, no sentido que:

$$e_i = \frac{@}{@_i};\ e^j = \frac{@}{@_j} \Rightarrow g(e_i; e^j) = {}_{ij}.$$

temos que $g_{ij}(\ ) = \frac{@_i}{@_j}$, e $g^{ij}(\ ) = \frac{@_i}{@_j}$.

as parametrizac-oes e de fato s-ao duais, no sentido que:

$$e_i = \frac{@}{@_i};\ e^j = \frac{@}{@_j} \Rightarrow g(e_i; e^j) = {}_{ij}.$$

temos que $g_{ij}(\ ) = \frac{@_i}{@_j}$, e $g^{ij}(\ ) = \frac{@_i}{@_j}$.

Observac-ao: outras famlias, como as misturas, tambem têm parâmetros duais

Geodesicas duais

as parametrizacoes duais induzem duas geometrias duais
a (e)-geometria e a (m)-geometria

sao simplesmente as geometrias planas das parametrizacoes

Geodesicas duais

as parametrizacoes duais induzem duas geometrias duais
a (e)-geometria e a (m)-geometria sao simplesmente planas:

- sao simplesmente as geometrias planas das parametrizacoes
- uma (e)-geodesica e uma reta nos parâmetros

Geodesicas duais

as parametrizacoes duais induzem duas geometrias duais localmente planas a (e)-geometria e a (m)-geometria

- sao simplesmente as geometrias planas das parametrizacoes
- uma (e)-geodesica e uma reta nos parâmetros
- uma (m)-geodesica e uma reta nos parâmetros

Geodesicas duais

as parametrizac-oes duais induzem duas geometrias duais
a (e)-geometria e a (m)-geometria

- s-ao simplesmente as geometrias planas das parametrizac-oes
- uma (e)-geodesica e uma reta nos parâmetros
- uma (m)-geodesica e uma reta nos parâmetros

as geometrias (e portanto as geodesicas) duais s-ao descritas por u
famlia conex-oes r , 2 R, com coe cientes

$$\Gamma^{(\alpha)}_{ij,k}(\theta) = E_p\left[\left(\frac{\partial}{\partial\theta^i}\frac{\partial}{\partial\theta^j} + \frac{1-\alpha}{2}\frac{\partial}{\partial\theta^i}\frac{\partial}{\partial\theta^j}\right)\frac{\partial}{\partial\theta^k}\right]$$

Geodesicas duais

as parametrizac-oes duais induzem duas geometrias duais localmente planas a (e)-geometria e a (m)-geometria

- s-ao simplesmente as geometrias planas das parametrizac-oes
- uma (e)-geodesica e uma reta nos parâmetros
- uma (m)-geodesica e uma reta nos parâmetros

as geometrias (e portanto as geodesicas) duais s-ao descritas por u famlia conex-oes r , 2 R, com coe cientes

$$\Gamma^{(\alpha)}_{ij,k}(\theta) = E_p\left[\left(\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}\right)\left(\frac{\partial}{\partial\theta_k}\right) + \frac{1-\alpha}{2}\left(\frac{\partial}{\partial\theta_i}\right)\left(\frac{\partial}{\partial\theta_j}\right)\left(\frac{\partial}{\partial\theta_k}\right)\right]$$

= 0 =) geodesicas usuais
= 1 =) (e)-geodesicas
= 1 =) (m)-geodesicas

Projecoes de informacao

as (e)-projecoes e (m)-projecoes ortogonais em uma subvariedade S minimizam as entropias relativas duais $D_{KL}(p\|q)$ e $D_{KL}(q\|p)$

Projecoes de informacao

as (e)-projecoes e (m)-projecoes ortogonais em uma subvariedade S minimizam as entropias relativas duais $D_{KL}(p\|q)$ e $D_{KL}(q\|p)$

Distribuic-oes discretas

os pontos s-ao distribuic-oes discretas se X ! $[0; 1]$

$X = f x_1; : : : ; x_{d+1} g$ P
$p(x_i) = p_i \; 2 \; (0; 1), \quad {}_i p_i = 1$

Distribuic-oes discretas

os pontos s-ao distribuic-oes discretas $p : X \to [0,1]$

$$X = \{x_1, \ldots, x_{d+1}\} \quad \sum_i p_i = 1$$
$$p(x_i) = p_i \in (0,1),$$

essa variedade pode ser identi cada com o interior simplexo padr-ao

$$^d = \left\{ p \in R^{d+1} \quad 0 < p_j < 1; \quad \sum_i p_i = 1 \right\}$$

Distribuic-oes discretas

os pontos s-ao distribuic-oes discretas $p : X \to [0,1]$

$$X = \{x_1, \ldots, x_{d+1}\} \qquad \sum_i p_i = 1$$
$$p(x_i) = p_i \in (0,1),$$

essa variedade pode ser identi cada com o interior do simplexo padr-ao

$$^d = \left\{ p \in R^{d+1} \quad 0 < p_j < 1; \quad \sum_i p_i = 1 \right\}^o$$

parametrizac-ao

$$\text{domnio} = \{(p_1, \ldots, p_d) \in R^d_+ \quad \sum_i p_i < 1\}$$

Distribuic-oes discretas

os pontos s-ao distribuic-oes discretas $p : X \to [0,1]$

$$X = \{x_1, \ldots, x_{d+1}\} \qquad p(x_i) = p_i \in (0,1), \quad \sum_i p_i = 1$$

essa variedade pode ser identificada com o interior do simplexo padr-ao

$$\Delta^d = \left\{ p \in \mathbb{R}^{d+1} \;\middle|\; 0 < p_j < 1; \quad \sum_i p_i = 1 \right\}^o$$

parametrizac-ao

$$\text{domnio} = \left\{ (p_1, \ldots, p_d) \in \mathbb{R}^d_+ \;\middle|\; \sum_i p_i < 1 \right\}$$

$$(p_1, \ldots, p_d) = (p_1, \ldots, p_d, p_{d+1}), \quad \text{com } p_{d+1} = 1 - \sum_{i=1}^{d} p_i$$

matriz de Fisher:

$$g_{ij}(p) = \frac{1}{p_{d+1}} + \frac{ij}{p_i}$$

matriz de Fisher:

$$g_{ij}(p) = \frac{1}{p_{d+1}} + \frac{ij}{p_i}$$

para calcular adistância de Fisher-Rao, fazemos uma reparametrizacao $z_i = 2\sqrt{p_i}$,

matriz de Fisher:

$$g_{ij}(p) = \frac{1}{p_{d+1}} + \frac{ij}{p_i}$$

para calcular adistância de Fisher-Rao, fazemos uma reparametrizacao $z_i = 2\sqrt{p_i}$,

que leva pontos $p = (p_1; : : : ; p_{d+1}) \; 2 \; ^{d}$ em pontos z no setor positivo da esfera

$$S^d_{2;+} = \left\{ z \; 2 \; R^{d+1}_+ \;\; P_{i=1}^{d+1} z_i^2 = 4 \right\}$$

matriz de Fisher:

$$g_{ij}(p) = \frac{1}{p_{d+1}} + \frac{ij}{p_i}$$

para calcular adistância de Fisher-Rao, fazemos uma reparametrizacao $z_i = 2\sqrt{p_i}$,

que leva pontos $p = (p_1; : : : ; p_{d+1}) \; 2 \;^{d}$ em pontos z no setor positivo da esfera

$$S^d_{2;+} = \left\{ z \; 2 \; R^{d+1}_+ \;\middle|\; \sum_{i=1}^{d+1} z_i^2 = 4 \right\}$$

nessa nova parametrizacao, a metrica de Fisher e a metrica esferica usual de $S^d_{2;+} \quad R^{d+1}$: $g_{ij}(z) = \frac{@z}{@p_i}; \frac{@z}{@p_j}$

matriz de Fisher:

$$g_{ij}(p) = \frac{1}{p_{d+1}} + \frac{ij}{p_i}$$

para calcular a distância de Fisher-Rao, fazemos uma reparametrizacao $z_i = 2\sqrt{p_i}$,

que leva pontos $p = (p_1; : : : ; p_{d+1}) \; 2 \quad ^d$ em pontos z no setor positivo da esfera

$$S^d_{2;+} = \left\{ z \; 2 \; R^{d+1}_+ \quad P \; {}^{d+1}_{i=1} \; z_i^2 = 4 \right\}$$

nessa nova parametrizacao, a metrica de Fisher e a metrica esferica usual de $S^d_{2;+} \quad R^{d+1}$: $g_{ij}(z) = \frac{@z}{@p_i}; \frac{@z}{@p_j}$

portanto a distância de Fisher-Rao entre p e q pode facilmente ser calculada como o comprimento do arco ligando z_p e z_q, que equivale a

matriz de Fisher:

$$g_{ij}(p) = \frac{1}{p_{d+1}} + \frac{ij}{p_i}$$

para calcular a distância de Fisher-Rao, fazemos uma reparametrização $z_i = 2\sqrt{p_i}$,

que leva pontos $p = (p_1; : : : ; p_{d+1}) \; 2 \;^{d}$ em pontos z no setor positivo da esfera

$$S^d_{2;+} = \left\{ z \; 2 \; R^{d+1}_+ \;\middle|\; \sum_{i=1}^{d+1} z_i^2 = 4 \right\}$$

nessa nova parametrização, a métrica de Fisher é a métrica esférica usual de $S^d_{2;+} \quad R^{d+1}$: $g_{ij}(z) = \frac{@z}{@p_i}; \frac{@z}{@p_j}$

portanto a distância de Fisher-Rao entre p e q pode facilmente ser calculada como o comprimento do arco ligando z_p e z_q, que equivale a

$$d_{FR}(p; q) = 2 \arccos\left(\sum_{i=1}^{d+1} \sqrt{p_i q_i} \right)$$

um fato interessante e que o comprimento da corda ligando z_p e z_q fornece uma boa aproximacao:

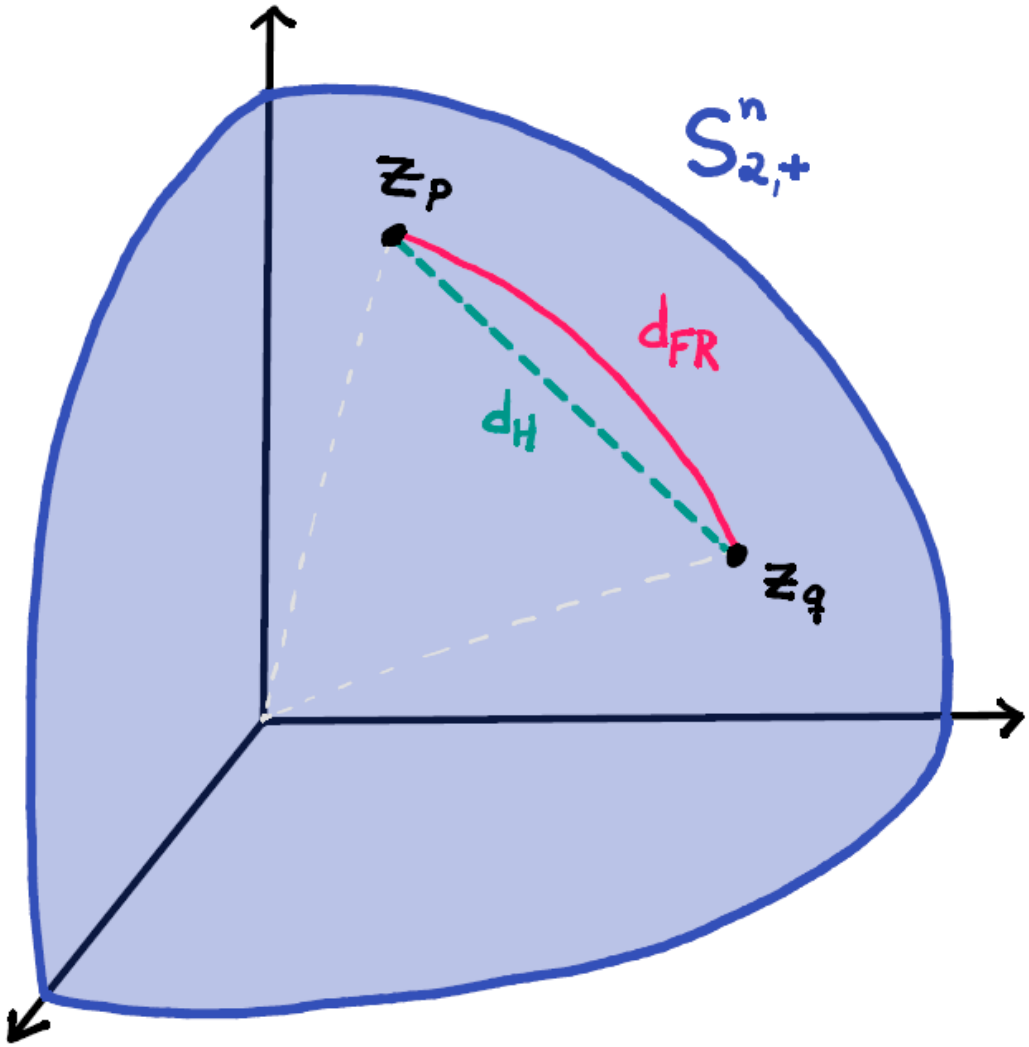
um fato interessante e que o comprimento da corda ligando z_p e z_q fornece uma boa aproximacao:

$$z_p \quad z_q = 2 \sum_{i=1}^{k+1} \left(p \frac{}{p_i} \quad p \frac{}{q_i} \right)^2 !^{1=2}$$

- um fato interessante é que o *comprimento da corda* ligando z_p e z_q fornece uma boa aproximação:

$$z_p \quad z_q \quad = 2 \quad \sum_{i=1}^{d+1} \quad p\overline{p_i} \quad p\overline{q_i}^{\,2} \quad !^{1=2}$$

- essa distância, sem o fator 2, é chamada *distância de Hellinger* d_{H}


$S^n_{2,+}$
z_p
d_{FR}
d_H
z_q

Parte III

Aplicacoes

Funcoes perda para classificadores

problemas de classificacao em aprendizado de maquina:

Funcoes perda para classificadores

problemas de classificacao em aprendizado de maquina:

temos uma familia parametrizada de funcoes $\{f : X \to \mathbb{R}^K\}$

Funcoes perda para classificadores

problemas de classificacao em aprendizado de maquina:

temos uma familia parametrizada de funcoes $f : X \to \mathbb{R}^K$

$X \subset \mathbb{R}^n$ e o espaco dos vetores de caracteristica (ex: imagens) $\mathbb{R}^{n^2}$

Funcoes perda para classificadores

problemas de classificacao em aprendizado de maquina:

temos uma familia parametrizada de funcoes $f : X \to \mathbb{R}^K$

$X \subset \mathbb{R}^n$ e o espaco dos vetores de caracteristica (ex: imagens)

K e o numero de classes (ex: cachorro, gato, etc.)

Funcoes perda para classificadores

problemas de classificacao em aprendizado de maquina:

temos uma familia parametrizada de funcoes $f : X \to \mathbb{R}^K$

- $X \subset \mathbb{R}^n$ e o espaco dos vetores de caracteristica (ex: imagens)
- K e o numero de classes (ex: cachorro, gato, etc.)
- $z = f(x)$ e chamado vetor escore

Funcoes perda para classificadores

problemas de classificacao em aprendizado de maquina:

temos uma familia parametrizada de funcoes $f : X \to \mathbb{R}^K$

$X \subset \mathbb{R}^n$ e o espaco dos vetores de caracteristica (ex: imagens)

K e o numero de classes (ex: cachorro, gato, etc.)

$z = f(x)$ e chamado vetor escore

α sao o parâmetros da maquina (geralmente dados por uma rede neural)

Funcoes perda para classificadores

problemas de classificacao em aprendizado de maquina:

temos uma familia parametrizada de funcoes $f : X \to \mathbb{R}^K$

$X \subset \mathbb{R}^n$ e o espaco dos vetores de caracteristica (ex: imagens)

K e o numero de classes (ex: cachorro, gato, etc.)

$z = f(x)$ e chamado vetor escore

θ sao o parâmetros da maquina (geralmente dados por uma rede neural)

transformamos o vetor escore em um vetor de probabilidades atraves da funcao softmax $\sigma : \mathbb{R}^K \to \Delta^{K-1}$ dada em coordenadas por

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{i=1}^{K} e^{z_i}}$$

Funcoes perda para classificadores

problemas de classificacao em aprendizado de maquina:

temos uma familia parametrizada de funcoes $f : X \to \mathbb{R}^K$

$X \subset \mathbb{R}^n$ e o espaco dos vetores de caracteristica (ex: imagens)

K e o numero de classes (ex: cachorro, gato, etc.)

$z = f(x)$ e chamado vetor escore

θ sao o parâmetros da maquina (geralmente dados por uma rede neural)

transformamos o vetor escore em um vetor de probabilidades atraves da funcao softmax $\sigma : \mathbb{R}^K \to \Delta^{K-1}$ dada em coordenadas por

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{i=1}^{K} e^{z_i}}$$

podemos interpretar $\sigma(z)_i$ como a probabilidade do vetor pertencer a classe i

treinamento supervisionado: somos fornecidos com um conjunto de treinamento $(x_i; y_i)_{i=1}^{m}$ X f $1; : : : ; K g$

treinamento supervisionado: somos fornecidos com um conjunto de treinamento $(x_i; y_i)_{i=1}^{m} \quad X \quad f \quad 1; : : : ; K g$

tomamos uma funcao perda $a : \quad ^{K \quad 1} \quad ^{K \quad 1} ! \quad R_+$

treinamento supervisionado: somos fornecidos com um conjunto de treinamento $(x_i, y_i)_{i=1}^{m} \quad X \quad f \quad 1, \ldots, K\}$

tomamos uma funcao perda $L : \quad ^{K \quad 1} \quad ^{K \quad 1} \; ! \quad R_+$

o problema de aprendizado de maquina consiste em minimizar a per
media do conjunto de treinamento:

$$\min_{2} \quad \frac{1}{m} \sum_{i=1}^{m} \quad L \quad f(x_i), e_{y_i}$$

treinamento supervisionado: somos fornecidos com um conjunto de treinamento $(x_i, y_i)_{i=1}^{m} \quad X \quad f \quad 1, \ldots, K g$

tomamos uma funcao perda $L : \quad ^{K \quad 1} \quad ^{K \quad 1} ! \quad R_+$

o problema de aprendizado de maquina consiste em minimizar a per
media do conjunto de treinamento:

$$\min_{2} \quad \frac{1}{m} \sum_{i=1}^{m} \quad L \quad f(x_i); e_{y_i}$$

isso costuma ser feito atraves do metodo do gradiente

funcoes perda mais usadas:

entropia cruzada: $h(p, q) = \sum_i p_i \log \frac{1}{q_i}$

funções perda mais usadas:

entropia cruzada: $h(p;q) = \sum_i p_i \log \frac{1}{q_i}$

perda quadrática: $L(p;q) = kp - qk_2^2$

funções perda mais usadas:

entropia cruzada: $h(p,q) = \sum_i p_i \log \frac{1}{q_i}$

perda quadrática: $L(p,q) = \|p - q\|_2^2$

nossa proposta: usar as perdas geométrico-informacionais no simplexo, dadas pelo quadrado das distâncias apresentadas:

funções perda mais usadas:

entropia cruzada: $h(p;q) = \sum_i p_i \log \frac{1}{q_i}$

perda quadrática: $L(p;q) = \|p - q\|_2^2$

nossa proposta: usar as perdas geométrico-informacionais no simplexo, dadas pelo quadrado das distâncias apresentadas:

$$4L_{SFR} = d^2_{FR}(p;q) = 4 \arccos \sum_{i=1}^{K} \sqrt{p_i q_i}^{\,2}$$

$$L_{SH} = d^2_H(p;q) = \sum_{i=1}^{K} \left(\sqrt{p_i} - \sqrt{q_i}\right)^2$$

funções perda mais usadas:

entropia cruzada: $h(p;q) = \sum_i p_i \log \frac{1}{q_i}$

perda quadrática: $L(p;q) = \|p - q\|_2^2$

nossa proposta: usar as perdas geométrico-informacionais no simplexo, dadas pelo quadrado das distâncias apresentadas:

$$4L_{SFR} = d^2_{FR}(p;q) = 4 \arccos \sum_{i=1}^{K} \sqrt{p_i q_i}^2$$

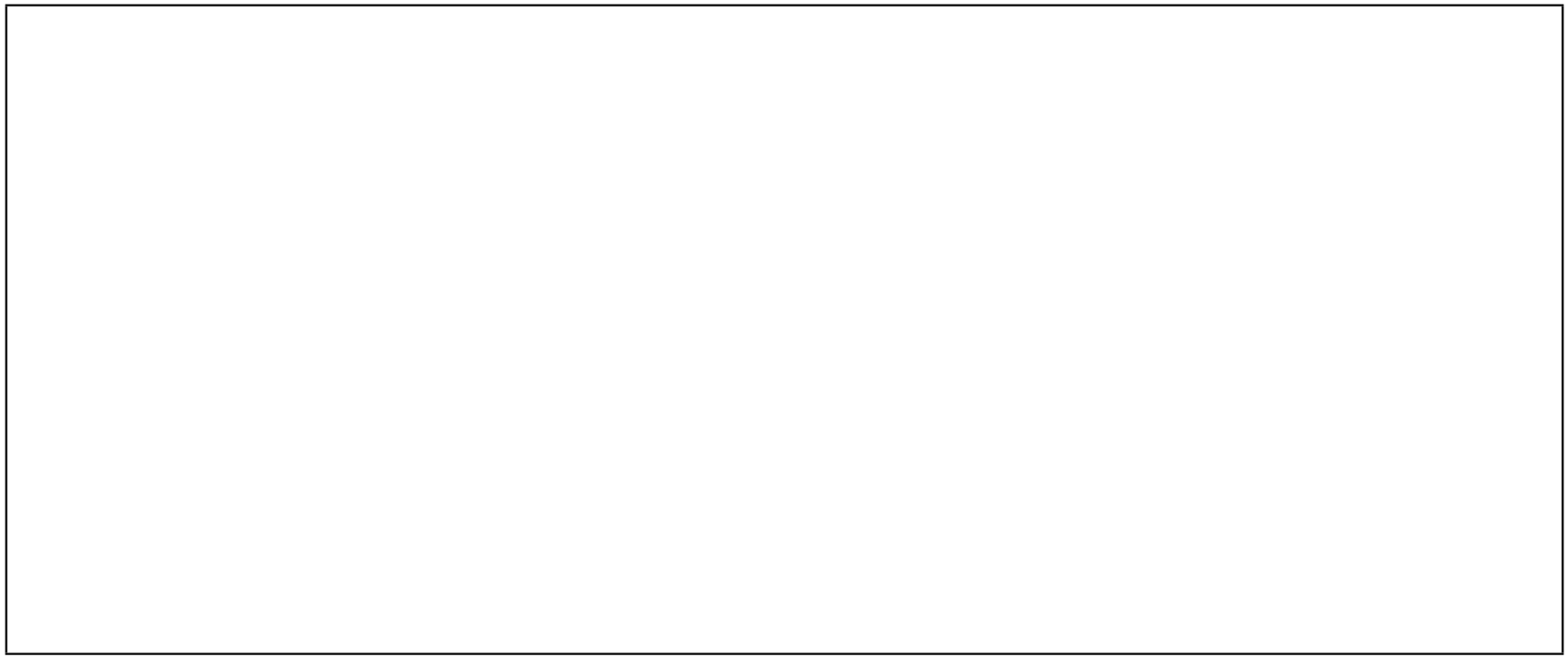
$$L_{SH} = d^2_H(p;q) = \sum_{i=1}^{K} \left(\sqrt{p_i} - \sqrt{q_i}\right)^2$$

este é um trabalho em conjunto com H.K. Miyamoto e S.I.R. Costa, submetido para o ISIT 2022 (International Symposium on Information Theory)[2]

[2] Henrique K. Miyamoto, Fabio C. C. Meneghetti e Sueli I. R. Costa. "Information-Geometric Loss Functions for Learning". Em: ISIT 2022. 2022.

observamos que existem relacoes assintoticas e desigualdades entr
perdas que introduzimos e a perda da entropia cruzada

observamos que existem relacoes assintoticas e desigualdades entr
perdas que introduzimos e a perda da entropia cruzada



Resultados

(a) MNIST (b) Banco de dados sintetico

Figura: Acuracia dos aprendizados com diferentes funcoes perda.

Resultados com rudo

(alguns rotulos do conjunto de treinamento recebem a classe errada)

(a) MNIST, = 0:3

(b) MNIST, = 0:5

(a) Sintetico, = 0:3

(b) Sintetico, = 0:5

possivelmente uma das razoes das perdas de Fisher-Rao e Hellinger terem boa performance no caso ruidoso seja por elas serem limitad

possivelmente uma das razoes das perdas de Fisher-Rao e Hellinger terem boa performance no caso ruidoso seja por elas serem limitadas

este e um trabalho em andamento

Distribuicoes de probabilidade enroladas no toro

um reticulado posto-completo e um subconjunto de $\mathbb{R}^n$ gerado por combinacoes lineares inteiras de uma base $\{b_1, \ldots, b_n\}$

Distribuicões de probabilidade enroladas no toro

um reticulado posto-completo e um subconjunto de $\mathbb{R}^n$ gerado por combinacões lineares inteiras de uma base $\{b_1, \ldots, b_n\}$

de nimos o toro enrolado por como

$$T := \mathbb{R}^n = \quad [\![x]\!] = x + \qquad x \in \mathbb{R}^n$$

Distribuições de probabilidade enroladas no toro

um reticulado posto-completo e um subconjunto de $\mathbb{R}^n$ gerado por combinações lineares inteiras de uma base $\{b_1, \ldots, b_n\}$

de nimos o toro enrolado por como

$$T := \mathbb{R}^n / = \quad [\![x]\!] = x + \quad x \in \mathbb{R}^n$$

temos uma projeção canônica $: \mathbb{R}^n \to T$, $(x) = [\![x]\!]$

Distribuicoes de probabilidade enroladas no toro

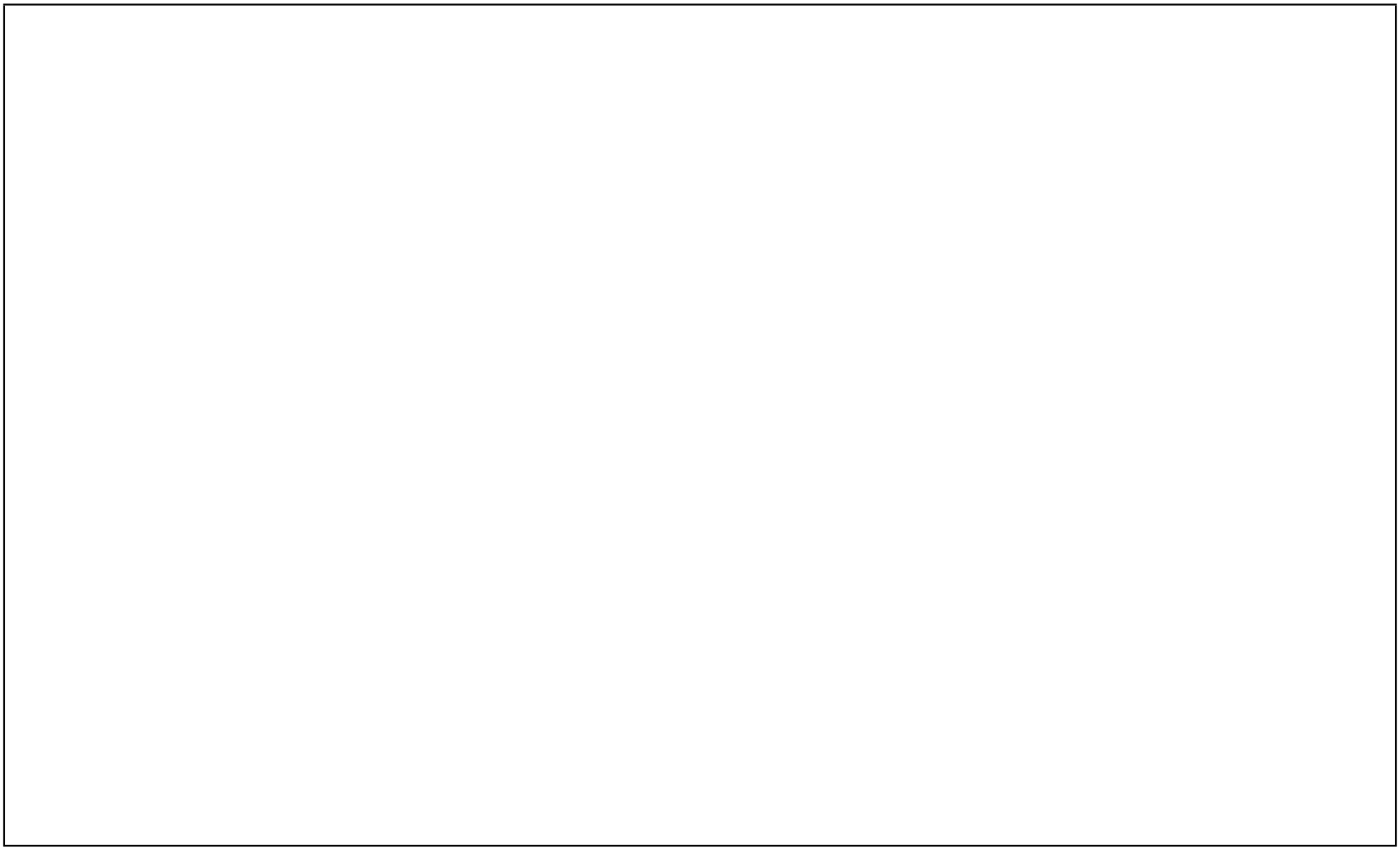
um reticulado posto-completo e um subconjunto de $\mathbb{R}^n$ gerado por combinacoes lineares inteiras de uma base $\{b_1, \ldots, b_n\}$

de nimos o toro enrolado por como

$$T := \mathbb{R}^n / = \quad [\![x]\!] = x + \quad x \in \mathbb{R}^n$$

temos uma projecao canônica $: \mathbb{R}^n \to T$, $(x) = [\![x]\!]$

dada uma distribuicao de probabilidade em $\mathbb{R}^n$, podemos enrola-la no toro com um empurro via , isto e, $P := (\quad) P$



se uma distribuicao P em R^n tem densidade $p(x)$, $x \in R^n$, entao a distribuicao enrolada tem densidade

$$p_{\ }(x) = \sum_{2} p(x + \ )$$

sobre T ou uma regiao fundamental (ex: regiao de Voronoi)

se uma distribuicao P em R^n tem densidade $p(x)$, $x \in R^n$, entao a distribuicao enrolada tem densidade

$$p_{\ }(x) = \sum_{\ \in \ _2} p(x + \ \ )$$

sobre T ou uma regiao fundamental (ex: regiao de Voronoi)

assim, a partir de um modelo estatstico $p \ g_{\ _2}$ em R^n obtemos um modelo estatstico $p_{\ ; \ \ _2}$ em T .

se uma distribuicao P em $\mathbb{R}^n$ tem densidade $p(x)$, $x \in \mathbb{R}^n$, entao a distribuicao enrolada tem densidade

$$p_{\mathbb{Z}^n}(x) = \sum_{k \in \mathbb{Z}^n} p(x + k)$$

sobre T ou uma regiao fundamental (ex: regiao de Voronoi)

assim, a partir de um modelo estatstico p_{μ,σ^2} em $\mathbb{R}^n$ obtemos um modelo estatstico p_{μ,σ^2} em T .

ex: gaussianas multivariadas

$$p_{\mu,K}(x) = \frac{1}{\sqrt{(2\pi)^d |\det K|}} \sum_{k \in \mathbb{Z}^n} e^{-\frac{1}{2}(x+k-\mu)^T K^{-1}(x+k-\mu)}$$

uma propriedade central dessas distribuicoes e que para variância crescente elas se aproximam da distribuicao uniforme $U(x) = \frac{1}{|\det J|}$

uma propriedade central dessas distribuicoes e que para variância crescente elas se aproximam da distribuicao uniforme $U(x) = \frac{1}{|\det J|}$

gostaramos de estudar a geometria dessas distribuicoes

gostaramos de estudar a geometria dessas distribuicoes

em termos da geometria de Fisher-Rao

gostaramos de estudar a geometria dessas distribuições

em termos da geometria de Fisher-Rao
e em termos de medidas de divergência (Kulback-Leibler, f -divergências, normas L^p, etc.)

gostaramos de estudar a geometria dessas distribuicões

- em termos da geometria de Fisher-Rao
- e em termos de medidas de divergência (Kulback-Leibler, f -divergências, normas L^p, etc.)

já existe alguma pesquisa sobre distribuicões desse tipo em termos geometria de Wasserstein[3]

[3] Anton Mallasto e Aasa Feragen. \Optimal Transport Distance between Wrapped Gaussian Distributions". Em: 38th MaxEnt. 2018

gostaríamos de estudar a geometria dessas distribuições

- em termos da geometria de Fisher-Rao
- e em termos de medidas de divergência (Kulback-Leibler, f-divergências, normas L^p, etc.)

já existe alguma pesquisa sobre distribuições desse tipo em termos geometria de Wasserstein[3]

nossa motivação: o fator de achatamento (flatness factor) e a distância L^1 entre uma distribuição P e a uniforme

[3] Anton Mallasto e Aasa Feragen. "Optimal Transport Distance between Wrapped Gaussian Distributions". Em: 38th MaxEnt. 2018

gostaramos de estudar a geometria dessas distribuicoes
- em termos da geometria de Fisher-Rao
- e em termos de medidas de divergência (Kulback-Leibler, f-divergências, normas L^p, etc.)

já existe alguma pesquisa sobre distribuicoes desse tipo em termos da geometria de Wasserstein[3]

nossa motivacao: o fator de achatamento (flatness factor) e a distância L^1 entre uma distribuicao P e a uniforme
- ele e um parâmetro importante para construir codigos que atingem a capacidade no canal AWGN e para garantir segredo no canal Wiretap

[3] Anton Mallasto e Aasa Feragen. "Optimal Transport Distance between Wrapped Gaussian Distributions". Em: 38th MaxEnt. 2018

- gostaríamos de estudar a geometria dessas distribuições
  - em termos da geometria de Fisher-Rao
  - e em termos de medidas de divergência (Kulback-Leibler, f-divergências, normas L^p, etc.)
- já existe alguma pesquisa sobre distribuições desse tipo em termos da geometria de Wasserstein[3]
- nossa motivação: o fator de achatamento (flatness factor) e a distância L^1 entre uma distribuição P e a uniforme
  - ele é um parâmetro importante para construir códigos que atingem a capacidade no canal AWGN e para garantir segredo no canal Wiretap
  - queremos entender se o fator de achatamento medido com outras divergências também tem comportamento interessante

[3] Anton Mallasto e Aasa Feragen. "Optimal Transport Distance between Wrapped Gaussian Distributions". Em: 38th MaxEnt. 2018

- gostaramos de estudar a geometria dessas distribuicões
  - em termos da geometria de Fisher-Rao
  - e em termos de medidas de divergência (Kulback-Leibler, f-divergências, normas L^p, etc.)
- já existe alguma pesquisa sobre distribuicões desse tipo em termos da geometria de Wasserstein[3]
- nossa motivacão: o fator de achatamento (flatness factor) e a distância L^1 entre uma distribuicão P e a uniforme
  - ele e um parâmetro importante para construir códigos que atingem capacidade no canal AWGN e para garantir segredo no canal Wiretap
  - queremos entender se o fator de achatamento medido com outras divergências também tem comportamento interessante
  - este tema esta diretamente conectado ao mestrado do aluno[4]

[3] Anton Mallasto e Aasa Feragen. "Optimal Transport Distance between Wrapped Gaussian Distributions". Em: 38th MaxEnt. 2018

[4] Fabio C. C. Meneghetti. "Reticulados: um estudo de alguns parâmetros relevantes para aplicacões em criptografia". 2020

Parte IV

Futuro

Futuro

queremos continuar estudando alguns aspectos teóricos

Futuro

queremos continuar estudando alguns aspectos teóricos

- a relação entre a estrutura dualmente plana de Amari $(M; g; \nabla; \nabla^*)$ e as geometrias simplética e Kähler

Futuro

queremos continuar estudando alguns aspectos teóricos

- a relação entre a estrutura dualmente plana de Amari $(M; g; r; r^*)$ e as geometrias simplética e Kähler
- extensões da geometria da informação para espaços de dimensão infinita (ex: estrutura de Pistone-Sempi)

Futuro

queremos continuar estudando alguns aspectos teóricos

- a relação entre a estrutura dualmente plana de Amari $M = (g, r, r^*)$ e as geometrias simplética e Kähler
- extensões da geometria da informação para espaços de dimensão infinita (ex: estrutura de Pistone-Sempi)

entender se há relação entre nossa proposta de funções perda, e as -Divergências, e também com o método do gradiente natural

Futuro

queremos continuar estudando alguns aspectos teoricos

- a relacao entre a estrutura dualmente plana de Amari $(M; g; r; r^*)$ e as geometrias simpletica e Kahler
- extensoes da geometria da informacao para espacos de dimensao infinita (ex: estrutura de Pistone-Sempi)

entender se ha relacao entre nossa proposta de funcoes perda, e as -Divergências, e tambem com o metodo do gradiente natural

formalizar a teoria das distribuicoes gaussianas enroladas no toro

Futuro

queremos continuar estudando alguns aspectos teoricos

- a relacao entre a estrutura dualmente plana de Amari $M = (g, r, r^*)$ e as geometrias simpletica e Kahler
- extensoes da geometria da informacao para espacos de dimensao infinita (ex: estrutura de Pistone-Sempi)

entender se ha relacao entre nossa proposta de funcoes perda, e as α-Divergências, e tambem com o metodo do gradiente natural

formalizar a teoria das distribuicoes gaussianas enroladas no toro

- procurar relacoes entre reticulados diferentes (ex: se $B = Z^n$, entao $p_{\mu, K}(x) = \frac{1}{\det} p_{\tilde{\mu}, \tilde{K}; Z^n}(B^{-1}x)$, $\tilde{\mu} = B^{-1}$, $\tilde{K} = B^{-1} K B^{-t}$)

Futuro

queremos continuar estudando alguns aspectos teoricos

- a relacao entre a estrutura dualmente plana de Amari $M = (g, r, r^*)$ e as geometrias simpletica e Kahler
- extensoes da geometria da informacao para espacos de dimensao infinita (ex: estrutura de Pistone-Sempi)

entender se ha relacao entre nossa proposta de funcoes perda, e as -Divergências, e tambem com o metodo do gradiente natural

formalizar a teoria das distribuicoes gaussianas enroladas no toro

- procurar relacoes entre reticulados diferentes (ex: se $B = Z^n$, entao $p_{\mu, K}(x) = \frac{1}{\det} p_{\tilde{\mu}, \tilde{K}; Z^n}(B^{-1}x)$, $\tilde{\mu} = B^{-1}$, $\tilde{K} = B^{-1} K B^{-t}$)
- reticulados duais parecem ter relacao com a transformada de Fourier distribuicao

Livros

[1] Shun'ichi Amari e Hiroshi Nagaoka. *Methods of information geometry*. Trad. por Daishi Harada. Translations of mathematical monographs. American Mathematical Society, 2007.

[3] Nihat Ay, Jürgen Jost, Hông Vân Lê e Lorenz Schwachhöfer. *Information Geometry*. Springer International Publishing, 2017.

[4] Ovidiu Calin e Constantin Udrişte. *Geometric Modeling in Probability and Statistics*. Springer International Publishing, 2014.

Artigos

[2] Colin Atkinson e Ann F. S. Mitchell. \Rao's Distance Measure". Em: Sankhya: The Indian Journal of Statistics, Series (1981).

[6] Ahmet Demirkaya, Jiasi Chen e Samet Oymak. \Exploring the Role of Loss Functions in Multiclass Classi cation". Em: 54th CISS. 2020

[7] Anton Mallasto e Aasa Feragen. \Optimal Transport Distance between Wrapped Gaussian Distributions". Em: 38th MaxEnt. 2018

[8] Fabio C. C. Meneghetti. \Reticulados: um estudo de alguns parâmetros relevantes para aplicações em criptogra a". 2020.

[10] Frank Nielsen. \An Elementary Introduction to Information Geometry". Em:Entropy (2020).

[12] Julianna Pinele, Joao E. Strapasson e Sueli I. R. Costa. \The Fisher-Rao Distance between Multivariate Normal Distributions: Special Cases, Bounds and Applications". Em:Entropy (2020).

[5] Alberto Cena e Giovanni Pistone. “Exponential statistical manifold”. Em: *Annals of the Institute of Statistical Mathematics* (2006).

[11] Tomonori Noda. “Sympletic Structures on Statistical Manifolds”. Em: *J. Aust. Math. Soc.* (2011).

[13] Rui F. Vigelis, Luiza H. F. De Andrade e Charles C. Cavalcante. “Properties of a Generalized Divergence Related to Tsallis Generalized Divergence”. Em: *IEEE Transactions on Information Theory* (2020).