



**UNIVERSIDADE DE ITAÚNA
PRÓ-REITORIA DE ENSINO
COORDENAÇÃO DE CIÊNCIA DA COMPUTACAO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

Eugênio Cunha

**APRENDIZADO DE MÁQUINA APLICADO À VALORAÇÃO DE
REDAÇÕES**

**ITAÚNA
2017**

EUGÊNIO CUNHA

APRENDIZADO DE MÁQUINA APLICADO À VALORAÇÃO DE REDAÇÕES

Projeto submetido à Coordenadoria do Curso de Bacharelado em Ciência da Computação da Universidade de Itaúna - Campus Verde, como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Área de pesquisa: Aprendizagem de Máquina

Orientador: Prof. Dr. Marco Túlio Alves N Rodrigues

Itaúna
2017



UNIVERSIDADE DE ITAÚNA
COORDENAÇÃO DE CIÊNCIA DA COMPUTAÇÃO

EUGÊNIO CUNHA

Este projeto foi julgada adequada para a obtenção do Grau de Bacharel em Ciência da Computação, sendo aprovada pela coordenação de ciência da computação do curso de Bacharelado em Ciência da Computação do Campus verde da Universidade de Itaúna e pela banca examinadora:

Orientador: Prof. Dr. Marco Túlio Alves N
Rodrigues
Universidade de Itaúna- UIT

Avaliador: Prof. Me. Felipe Domingos da
Cunha
Universidade de Itaúna- UIT

Itaúna, 19 de Junho de 2017

Dedico este trabalho ao meu filho Davi, sempre preocupado em proporcionar um minuto de brincadeira durante minhas horas de trabalho, “meus melhores minutos”!

“Os computadores são incrivelmente rápidos, precisos e burros; os homens são incrivelmente lentos, imprecisos e brilhantes; juntos, seus poderes ultrapassam os limites da imaginação”.

Albert Einstein

Resumo

Este trabalho baseou-se no estudo da avaliação de uma redação que é a soma de um conjunto de competências exigidas em um texto do tipo dissertativo-argumentativo com temas diversificados de ordem social, científica, cultural ou política. Fundamentou-se no estudo das técnicas de aprendizado de máquina supervisionado que provê uma gama diversificada de algoritmos poderosos para classificações de textos.

O objetivo deste trabalho é classificar as competências exigidas em um texto de redação a partir do treinamento de uma algoritmo de aprendizado de máquina com base em um corpus de redações avaliadas seguindo as competências exigidas em um texto do tipo dissertativo-argumentativo.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Símbolos

Lista de Abreviaco es

1	Introdução	12
1.1	Motivação e objetivos	13
1.2	Contribuições	13
1.3	Organização do trabalho	13
2	Trabalhos Relacionados	14
3	Fundamentação Teórica	15
3.1	Métodos de Kernel	15
3.2	Kernel em análise de padrões	15
3.2.1	Exemplo de uma equação mais complexa	16
3.3	Tabelas	16
4	Método Proposto	19
5	Desenvolvimento	20
6	Resultados Experimentais	21

7 Conclusão e Trabalhos Futuros	22
Referências Bibliográficas	23
Apêndice A – Título do Apêndice	24
Apêndice B – Exemplo do pacote Algorithm	25

Lista de Figuras

1	Curvas de funções de probabilidade: (a) exemplo 1, (b) exemplo 2. . . .	18
---	---	----

Lista de Tabelas

1	Modelos estatísticos e suas relações.	17
---	---	----

Lista de Símbolos

Z	variavel aleatoria
\mathbb{R}	conjunto dos números reais
t	tempo contínuo
n	tempo discreto
$f(z)$	função densidade de probabilidade
$F(z)$	função de distribuição acumulada
σ	desvio padrão
μ	média ou esperança matemática
$ \cdot $	operador magnitude
∇	operador gradiente

Lista de Abreviacoess

fdp Função densidade de probabilidade

fda Função de distribuição acumulada

EMQ Erro médio quadrático

INTRODUÇÃO

Este documento consiste de um modelo básico para a produção de documentos academicos, seguindo as normas ABNT.

Não e abordado o estudo do LaTeX neste template. Sugerimos a leitura do texto em Oetiker *et al.* (2007). O uso do LaTeX e aconselhavel devido a sua qualidade grafica, facil referenciacao, criacao de listas, indices, referencias bibliograficas e escrita matematica profissional. Porem, nao e obrigatorio o uso deste template, apenas as orientacoes de formatação segundo as normas ABNT devem ser obrigatoriamente seguidas.

Uma observação em particular é a de que, no pacote ABNTex, as referências diretas devem utilizar o comando “citeonline”. Referências indiretas utilizam o comando “cite”.

Exemplo de citacao direta: Uma otima fonte de estudo para compreender o LaTeX e apresentada por Oetiker *et al.* (2007).

Exemplo de citação indireta: Existem boas fontes de pesquisa para entendimento do LaTeX (OETIKER *et al.*, 2007; RADKE *et al.*, 2005), estas incluem documentação online disponível na web.

Uma citação da Internet (UNIVERSITY, 2016; JOSLIN, 1993).

Uma dissertação de mestrado (HARWOOD, 1993) ou Harwood (1993) de acordo com Oetiker *et al.* (2007).

Uma tese de doutorado (JOSLIN, 1993) ou Joslin (1993).

1.1 Motivação e objetivos

1.2 Contribuicoes

1.3 Organizacao do trabalho

Capitulo 2: descricao...

Capitulo 4: descricao...

Capitulo 5: descricao...

Capitulo 6: descricao...

CAPÍTULO 2

TRABALHOS RELACIONADOS

FUNDAMENTAÇÃO TEÓRICA

3.1 Métodos de Kernel

Este capítulo tem como objetivo

3.2 Kernel em análise de padrões

Em análise de padrões, temos como objetivo detectar automaticamente padrões em um conjunto de dados de um determinado problema. Por padrões, podemos entender qualquer relação ou regularidades inerentes à alguma estrutura em uma fonte de dados. Essa análise geralmente é feita a partir dos valores de entrada e suas respectivas saídas (no caso da aprendizagem supervisionada) fornecidas no problema. Essas informações podem formar padrões em que se torna possível verificar o valor de uma saída dada uma nova entrada fornecida pelo usuário.

Diversos problemas podem ser resolvidos utilizando esta abordagem, categorização de textos, análise de sequências de DNA, reconhecimento de escrita, por exemplo.

A abordagem de análise de padrões utilizando métodos de kernel se baseia em adaptar os dados de entrada em um espaço característico adequado e nos algoritmos usados para descobrir os padrões do problema. Levando em conta isso, podemos pensar que qualquer solução com métodos de kernel é composta por estas duas partes: uma em que é feito o mapeamento nesse espaço característico e a outra em que é executado o algoritmo de aprendizagem para detectar os padrões neste espaço. A ideia por trás desta abordagem é poder mapear os dados em um espaço em que possamos

Uma das principais características desses métodos é o atalho computacional que pode ser utilizado, tal atalho é conhecido como função de kernel.

O Kernel é uma função de mapeamento de dados em dimensões superiores com a motivação de torná-los mais fáceis de separar ou estruturá-los de maneira mais adequada. Essas funções podem ser utilizadas nas tarefas de reconhecimento de padrões.

$$Z = X \cdot Y, \quad (3.1)$$

em que Z , X e Y são variáveis complexas. A referência à Equação (3.1) é feita por meio do comando “ref”. O mesmo vale para outros tipos de elementos.

3.2.1 Exemplo de uma equação mais complexa

Equações mais complexas podem ser mais facilmente escritas com uso do programa TexAide. Como, por exemplo,

$$f_{\Gamma^{1/2}}(x; \alpha, \lambda) = \frac{2\lambda^\alpha}{\Gamma(\alpha)} x^{2\alpha-1} \exp(-\lambda x^2). \quad (3.2)$$

$$\alpha, \lambda > 0.$$

em que $\Gamma(\cdot)$ é a função Gama. O programa TexAide é semelhante ao *MathType* do Office, porém ao copiar e colar a equação em um arquivo tex, é gerado o código em LaTeX referente a esta equação.

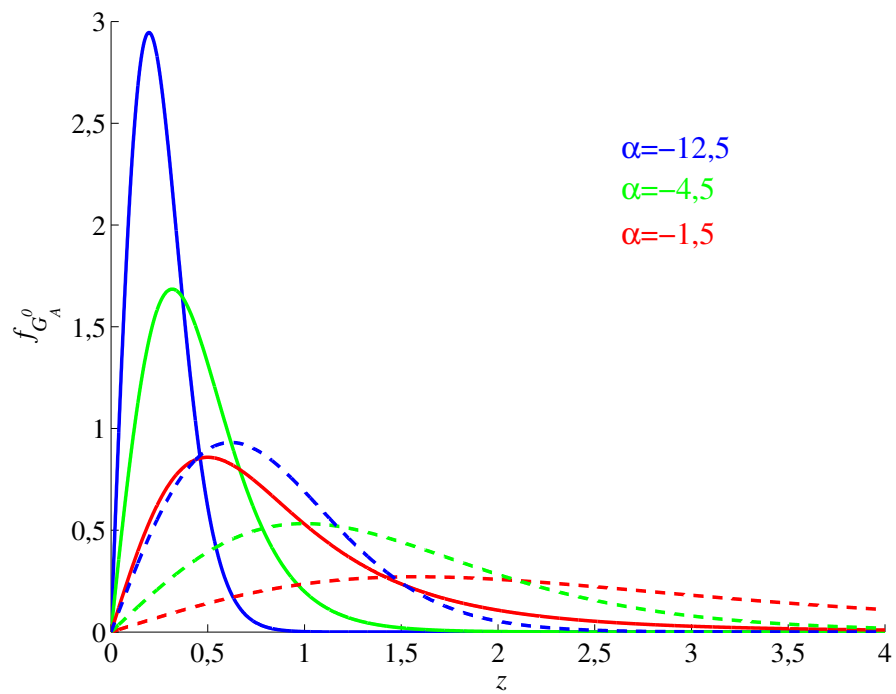
3.3 Tabelas

Tabelas são essenciais na apresentação de dados. A Tabela 1 mostra um exemplo do uso deste tipo de elemento. Vale ressaltar que não é aconselhável o uso de linhas verticais em trabalhos acadêmicos e de pesquisa.

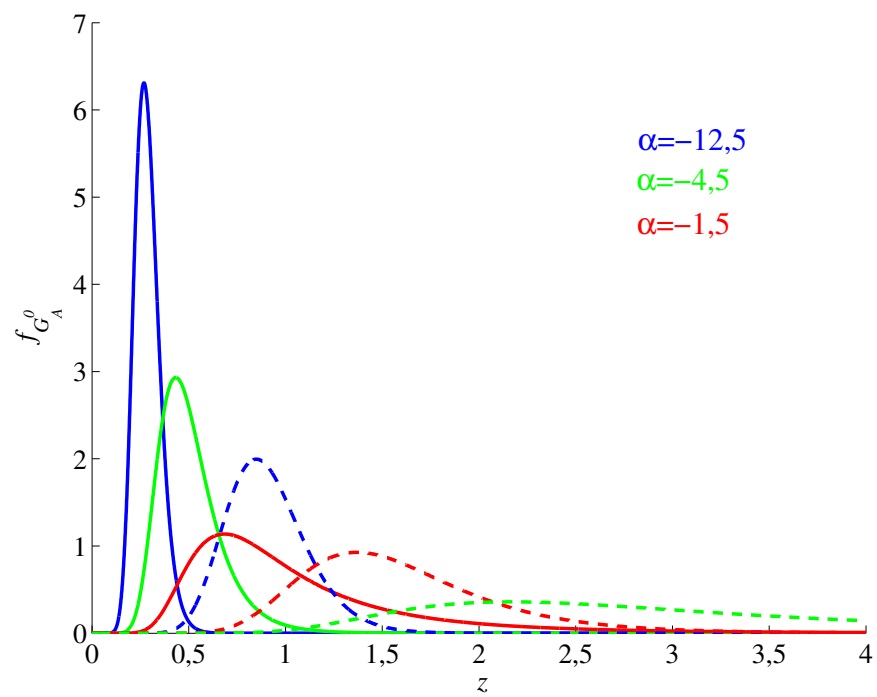
A Figura 1 mostra o exemplo do uso do comando “subfigure”. Apesar de aceitar diferentes tipos de imagens. É preferível que as imagens estejam no formato .eps. Isso garante que a imagem impressa seja exatamente aquela visualizada, como acontece com arquivos pdf.

Tabela 1: Modelos estatísticos e suas relações.

$\mathcal{N}^{-1/2}(x; \alpha, \gamma, \lambda)$	$\alpha, \lambda > 0$		$\alpha, \lambda \rightarrow \infty$	Homogêneo
	$\gamma \rightarrow 0$	Heterogêneo	$\alpha/\lambda \rightarrow \beta$	
	\xrightarrow{D}	$\sqrt{\Gamma}(\alpha, \lambda)$	\xrightarrow{P}	$\sqrt{\beta}$
	\xrightarrow{D}	$\Gamma^{-1/2}(\alpha, \gamma)$	\xrightarrow{P}	$\sqrt{\zeta^{-1}}$
	$\lambda \rightarrow 0$	Extremamente	$-\alpha/\gamma \rightarrow \zeta^{-1}$	
	$-\alpha, \gamma > 0$	Heterogêneo	$-\alpha, \gamma \rightarrow \infty$	Homogêneo
$\mathcal{G}_A(z; \alpha, \gamma, \lambda, n)$	$\alpha, \lambda > 0$		$\alpha, \lambda \rightarrow \infty$	Homogêneo
	$\gamma \rightarrow 0$	Heterogêneo	$\alpha/\lambda \rightarrow \beta$	
	\xrightarrow{D}	$\mathcal{K}_A(\alpha, \lambda, n)$	\xrightarrow{P}	$\sqrt{\Gamma}(n, n/\beta)$
	\xrightarrow{D}	$\mathcal{G}_A^0(\alpha, \gamma, n)$	\xrightarrow{P}	$\sqrt{\Gamma}(n, n\zeta)$
	$\lambda \rightarrow 0$	Extremamente	$-\alpha/\gamma \rightarrow \zeta$	
	$-\alpha, \gamma > 0$	Heterogêneo	$-\alpha, \gamma \rightarrow \infty$	Homogêneo



(a)



(b)

Figura 1: Curvas de funções de probabilidade: (a) exemplo 1, (b) exemplo 2.

CAPÍTULO 4

MÉTODO PROPOSTO

CAPÍTULO 5

DESENVOLVIMENTO

CAPÍTULO 6

RESULTADOS EXPERIMENTAIS

CAPÍTULO 7

CONCLUSÃO E TRABALHOS FUTUROS

Referências Bibliográficas

HARWOOD, Peter. **The title of the work**. Dissertação (Mestrado) — The school of the thesis, The address of the publisher, 7 1993.

JOSLIN, Peter. **The title of the work**. Tese (Doutorado) — The school of the thesis, The address of the publisher, 7 1993. An optional note.

OETIKER, Tobias; PARTL, Hubert; HYNA, Irene; SCHLEGL, Elisabeth. **Uma não tão pequena introdução ao LaTeX**. [S.l.]: Free Software Foundation, 2007. (, v. 4.20.1).

RADKE, Richard J; ANDRA, Srinivas; AL-KOFAHI, Omar; ROYSAM, Badrinath. Image change detection algorithms: a systematic survey. **IEEE transactions on image processing**, IEEE, v. 14, n. 3, p. 294–307, 2005.

UNIVERSITY, Carnegie Mellon. **The computer vision home page**. 2016. Disponível em: <<https://www.cs.cmu.edu/~cil/vision.html>>.

APÊNDICE A – Título do Apêndice

APÊNDICE B – Exemplo do pacote Algorithm

Algoritmo 1 Estimador ML otimizado.

- 1: Inicializar o contador: $j \leftarrow 1$;
 - 2: Fixar o limiar de variação das estimativas: $e_{\text{out}} \leftarrow 10^{-4}$;
 - 3: Fixar o número máximo de iterações: $N \leftarrow 1000$;
 - 4: Computar o ponto inicial: $\hat{\gamma}(0)$;
 - 5: Determinar o limiar inicial: $e_1 \leftarrow 1000$;
 - 6: Estabelecer o valor inicial de α : $\hat{\alpha}(0) \leftarrow -10^{-6}$;
 - 7: **enquanto** $e_j \geq e_{\text{out}}$ e $j \leq M$ **fazer**
 - 8: Solucionar $\hat{\alpha}_j \leftarrow \arg \max_{\alpha} l_1(\alpha; \gamma_{j-1}, \mathbf{z}, n)$;
 - 9: Solucionar $\hat{\gamma}_j \leftarrow \arg \max_{\gamma} l_2(\gamma; \alpha_j, \mathbf{z}, n)$;
 - 10: $j \leftarrow j + 1$
 - 11: Computar o critério de convergência: e_j ;
 - 12: **fim enquanto**
-