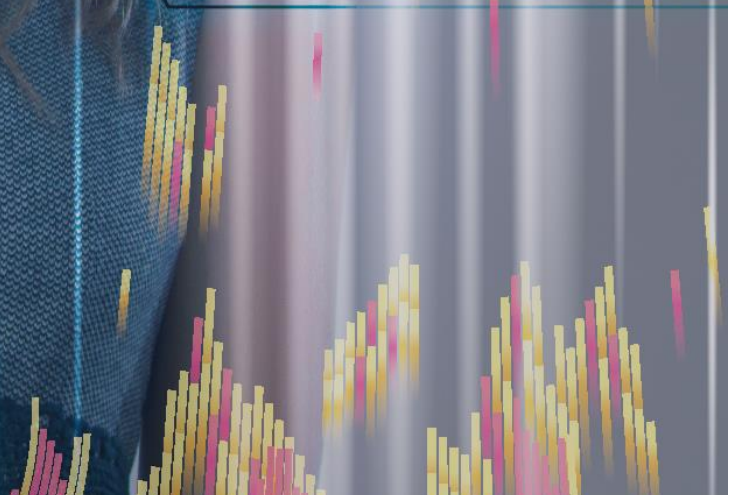
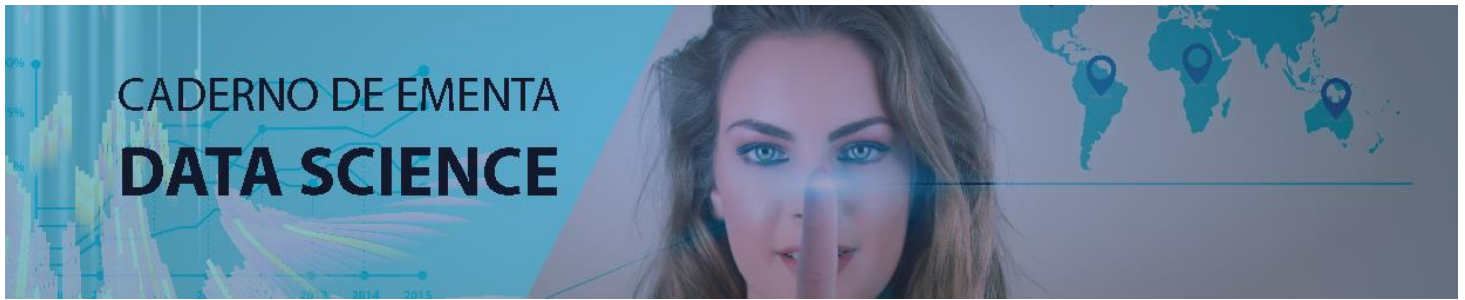




CADERNO DE EMENTA  
**DATA SCIENCE**





## Sumário

ALGORITMOS E PROGRAMAÇÃO .....	3
APRENDIZADO DE MÁQUINA (MACHINE LEARNING) .....	4
ECONOMETRIA: TÓPICOS ESPECIAIS.....	6
PROJETOS I.....	7
OTIMIZAÇÃO NÃO-LINEAR E ALGORITMOS .....	9
APRENDIZADO PROFUNDO (DEEP LEARNING) E POR RECOMPENSAS (REINFORCEMENT LEARNING).....	9
PROCESSAMENTO DE LINGUAGEM NATURAL (NLP) E DE IMAGENS .....	11
PROJETOS II.....	12

# CADERNO DE EMENTA DATA SCIENCE

## PROGRAMAÇÃO E BANCOS DE DADOS

Créditos	Carga Horária	Disciplina
3	45 horas	Obrigatória

### Objetivos

Esta disciplina tem dois objetivos. O primeiro é apresentar a modelagem matemática de problemas computacionais. Nesta primeira parte, o curso apresenta os algoritmos mais comuns e estruturas de dados usadas para resolver este problema. O curso enfatiza a relação entre algoritmos e programação, introduzindo as medidas básicas de performance destes algoritmos. Na segunda parte, o curso apresenta uma introdução a programação em Python

### Ementa

#### 1- Introdução

- 1.1- Estrutura de um algoritmo
- 1.2- Modelos Computacionais
- 1.3- Introdução ao Python

#### 2- Programação

- 2.1- Iteração e ramificação
- 2.2- Manipulação de string, "guess and check", aproximações
- 2.3- Decomposição, abstrações, funções
- 2.4- Listas, mutação, clones
- 2.5- Recursão
- 2.6- Programação Orientada
- 2.7- Exemplos práticos

#### 3- Algoritmos

- 3.1- "Sorting" e Árvores
- 3.2- "Hashing"

# CADERNO DE EMENTA

## DATA SCIENCE

- 3.3- Numéricos
- 3.4- Grafos
- 3.5- “Shortest Path”
- 3.6- Aplicações em Python

### Bibliografia Básica

- Cormen, Thomas, Charles Leiserson, Ronald Rivest, and Clifford Stein. Introduction to Algorithms. 3rd ed. MIT Press, 2009. ISBN: 9780262033848.
- Guttag, John. Introduction to Computation and Programming Using Python: With Application to Understanding Data Second Edition. MIT Press, 2016. ISBN: 9780262529624.
- Miller, Bradley, and David Ranum. Problem Solving with Algorithms and Data Structures Using Python. 2nd ed. Franklin, Beedle & Associates, 2011. ISBN: 9781590282571.

### APRENDIZADO DE MÁQUINA (MACHINE LEARNING)

Créditos	Carga Horária	Disciplina
4	60 horas	Obrigatória

### Objetivos

A disciplina Machine Learning tem como principal objetivo capacitar os alunos a compreender e aplicar os principais algoritmos de aprendizagem estatística (aprendizagem de máquina). Ao final do curso, espera-se que tenham adquirido a habilidade de implementar estes algoritmos em conjuntos de dados reais, utilizando a linguagem de programação Python e seus pacotes especializados.

### Ementa

1. Introdução à aprendizagem de máquina e sua relação com o problema de inferência estatística. Ferramentas computacionais.
2. Aprendizado não-supervisionado
  - 2.1.1.1. Clustering





# CADERNO DE EMENTA DATA SCIENCE

- 2.1.1.2. Redução de Dimensionalidade
- 3. Aprendizado supervisionado
  - 3.1.1.1. Classificação
  - 3.1.1.2. K-Nearest Neighbors
  - 3.1.1.3. Support Vector Machines
  - 3.1.1.4. Árvores de Decisão
  - 3.1.1.5. Naive Bayes
- 3.2. Regressão
  - 3.2.1.1. Árvores de Decisão
  - 3.2.1.2. Support Vector Machines
  - 3.2.1.3. Random Forests
  - 3.2.1.4. Bagging
  - 3.2.1.5. Boosting
- 4. Mecanismos de Regularização e Avaliação de Modelos
- 5. Ferramentas para limpeza e organização de dados
- 6. Computação em Nuvem

## Bibliografia Básica

Existe uma vasta literatura sobre machine learning e inteligência artificial, e o rápido desenvolvimento da disciplina requer a atualização constante através de sites como o StackOverflow. Entre os vários livros e manuais disponíveis, uma bibliografia básica é composta por:

Murphy, K, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

Este manual é bastante abrangente em termos das técnicas disponíveis, e apresenta alguma formalização matemática dos conceitos subjacentes.

Hastie, T. and Tibshirani, R.: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, 2016 (second edition).

Um clássico escrito por dois dos principais nomes da área, com aplicações usando a linguagem R. Também disponibilizado legalmente em versão eletrônica gratuita na internet.

Para uma boa introdução à implementação dos principais métodos usando a linguagem Python:

Raschka, S. and Mirjalili, V.: *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*, Packt Publishing, 2017 (second edition). Existe uma vasta literatura sobre machine learning e inteligência artificial, e o rápido desenvolvimento da disciplina requer a atualização constante através

# CADERNO DE EMENTA

## DATA SCIENCE

de sites como o StackOverflow. Entre os vários livros e manuais disponíveis, uma bibliografia básica é composta por:

Murphy, K, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

Este manual é bastante abrangente em termos das técnicas disponíveis, e apresenta alguma formalização matemática dos conceitos subjacentes.

Hastie, T. and Tibshirani, R.: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, 2016 (second edition).

Um clássico escrito por dois dos principais nomes da área, com aplicações usando a linguagem R. Também disponibilizado legalmente em versão eletrônica gratuita na internet.

Para uma boa introdução à implementação dos principais métodos usando a linguagem Python:

Raschka, S. and Mirjalili, V.: *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*, Packt Publishing, 2017 (second edition).

---

### INFERÊNCIA (ML, GMM E BAYESIANA)

Créditos	Carga Horária	Disciplina
3	45 horas	Obrigatória

#### Objetivos

Esta disciplina tem por objetivo fazer uma revisão do estimador de máxima verossimilhança, apresentar o estimador do método generalizado dos momentos (GMM) e fazer uma introdução a Econometria Bayesiana.

#### Ementa

- 1- Máxima Verossimilhança
  - 1.1- Apresentação do estimador
  - 1.2- Propriedades assintóticas: consistência e distribuição
  - 1.3- Discussão de eficiência
  - 1.4- Inferência: o trio de testes
  - 1.5- Algoritmos numéricos
  - 1.6- Programação em Python

# CADERNO DE EMENTA

## DATA SCIENCE

- 2- Método Generalizado dos Momentos
  - 2.1- Apresentação do estimador
  - 2.2- Exemplos: MQO, MLE
  - 2.3- Propriedades assintóticas: consistência e distribuição
  - 2.4- Inferência: o trio de testes
  - 2.5- Aplicações em Python
  
- 3- Econometria Bayesiana
  - 3.1- Introdução (como pensar em probabilidade)
  - 3.2- Estimação pontual
  - 3.3- Inferência Bayesiana
  - 3.4- Estimação por intervalos
  - 3.5- Predição
  - 3.6- O modelo de regressão linear
  - 3.7- Modelos hierárquicos
  - 3.8- Algoritmos bayesianos em Python

### Bibliografia Básica

#### Livros Principais:

- Dougherty, C. "Introduction to Econometrics", Oxford University Press, 2016.
- Heijl, Cjristiaan., de Bôer, Paul, Franses Philip Hans, Kloek, Teun e van Dijk, Herman K., Econometric Methods with applications in Business and Economics, Oxford University Press, 2004.
- Koop, G., Poirier, D., & Tobias, J. "Bayesian Econometric Methods". Cambridge: Cambridge Univeristy Press.

### PROJETOS I

Créditos	Carga Horária	Disciplina
2	30 horas	Obrigatória

### Objetivos

# CADERNO DE EMENTA

## DATA SCIENCE

Esta disciplina tem por objetivo a elaboração de um projeto de pesquisa a ser executado ao longo do curso. Este projeto deve contemplar a aplicação de uma das técnicas apresentadas ao longo do curso, quer seja do ponto de vista de investigação de suas propriedades teóricas, quer seja em um problema prático envolvendo a aquisição, tratamento e análise de bases de dados do mundo real.

### Ementa

1. Entrega parcial 1: Introdução (motivando a pergunta a ser respondida e deixando claro qual é essa pergunta) e Revisão da Literatura
2. Entrega parcial 2: Introdução, Revisão da Literatura e Montagem do banco de dados
3. Entrega parcial 3: Introdução, Revisão da Literatura, Montagem do banco de dados e Metodologia
4. Entrega final: projeto completo contendo todo o conteúdo desenvolvido acima.

Todas as entregas devem conter a assinatura do orientador do projeto de pesquisa.

### Bibliografia Básica

## DATA ANALYTICS E COMPUTAÇÃO EM NUVEM

Créditos	Carga Horária	Disciplina
3	45 horas	Obrigatória

### Objetivos

A disciplina de programação não-linear e algoritmos busca dar aos alunos uma visão analítica e computacional para a otimização de programas não-lineares. Os tópicos incluem otimização sem restrição e otimização com restrição. Os tópicos de otimização sem restrição incluem gradiente, direção conjugada, métodos de Newton. Os métodos de otimização com restrição incluem direções factíveis, projeções, ponto interior e multiplicador de lagrange. Além disso, incluem também análise convexa, otimização não-diferenciável, condições de otimalidade e teoria dual. Os algoritmos serão aplicados a solução de problemas reais e simulados.

### Ementa



# CADERNO DE EMENTA

## DATA SCIENCE

- 1- Introdução
- 2- Otimização sem restrição
  - 2.1- Condições para otimalidade
  - 2.2- Método do Gradiente
  - 2.3- Métodos de Newton e Gauss
  - 2.4- Outros métodos
- 3- Otimização em um espaço convexo
  - 3.1- Métodos direcionais; projeções
  - 3.2 – Métodos alternativos
- 4- Otimização com restrição: métodos de multiplicador de lagrange
- 5- Restrições de Desigualdade
- 6- Introdução a teoria dual
- 7- Métodos do ponto interior
- 8- Métodos de penalização
- 9- Teoria dual

### Bibliografia Básica

Livros Principais:

- Bertsekas, Dimitri. Nonlinear Programming: 2nd Edition. Belmont, MA: Athena Scientific Press, 1999. ISBN: 1886529000.
- Bazaraa, Mokhtar S., Hanif D. Sherali, and C. M. Shetty. Nonlinear Programming: Theory and Algorithms. New York: John Wiley & Sons, 1993. ISBN: 0471557935.

### APRENDIZADO PROFUNDO (DEEP LEARNING) E POR RECOMPENSAS (REINFORCEMENT LEARNING)

Crédito	Carga Horária	Disciplina
4	60 horas	Obrigatória

### Objetivos



# CADERNO DE EMENTA DATA SCIENCE

Na sequência dos métodos apresentados na disciplina Aprendizado de Máquina, os dois conjuntos de algoritmos dentro das áreas de Deep Learning e Reinforcement Learning vem ganhando grande relevância em função do desempenho em problemas concretos. Este curso tem como objetivo apresentar a teoria por trás destes métodos, bem como sua implementação através de softwares que possibilita a aplicação em problemas concretos.

## Ementa

1. Representação de dados para redes neurais
  - 1.1. Vetores
  - 1.2. Tensores
  - 1.3. Numpy
2. Estruturas básicas de redes neurais
3. Aplicações
  - 3.1. Classificação
  - 3.2. Regressão
4. Alternativas de Softwares: TensorFlow, Keras e PyTorch
5. Modelos com camadas ocultas completamente conectadas
6. Arquiteturas de Redes Neurais
  - 6.1. Convolucionais
  - 6.2. Recorrentes
7. Processos de Decisão Markovianos
8. Programação Dinâmica
  - 8.1. Avaliação de políticas
  - 8.2. Iteração de funções valor
9. Aprendizagem por Recompensas
  - 9.1. Métodos de Solução
  - 9.2. Aplicações

## Bibliografia Básica

- Chollet, F., “Deep Learning with Python”, Manning Publications, 2018.
- Sutton, R. e Barto, G., “Reinforcement Learning: An Introduction (2nd edition)”, MIT Press, 2019. SIMON,

# CADERNO DE EMENTA DATA SCIENCE

## PROCESSAMENTO DE LINGUAGEM NATURAL (NLP) E DE IMAGENS

Créditos	Carga Horária	Disciplina
3	45 horas	Obrigatória

### Objetivos

O conceito de *Big Data* diz respeito não somente à dimensão das bases de dados, mas de forma igualmente importante à natureza das informações. Um conjunto de métodos e algoritmos vem sendo desenvolvido e aplicado com bastante sucesso à interpretação de informações não só na forma de números, mas também de textos e imagens. O objetivo desta disciplina é capacitar o aluno a utilizar de forma eficiente estes tipos de informações dentro do contexto de aprendizagem de máquina.

### Ementa

1. Conceitos básicos de NLP
  - 1.1. Corpus de um texto
  - 1.2. Parágrafos, sentenças, frases e palavras
  - 1.3. N-gramas
  - 1.4. *Bag-of-words*
2. Análise de Sentimentos
3. Classificação de Textos
4. Implementações em softwares
5. Classificação de Imagens utilizando Aprendizado Profundo
6. Transformando informações de imagens em atributos para modelos de aprendizagem de máquina.

### Bibliografia Básica

- Arumugan, R. e Shanmugamani, R. “Hands-On Natural Language Processing with Python”, Packt Publishing, 2018.
- Ganegedara, T. “Natural Language Processing with TensorFlow”, Packt Publishing, 2018.
- Ballard, W., “Hands-On Deep Learning for Images with TensorFlow: Build Intelligent Computer Vision Applications using TensorFlow and Keras, Packt Publishing, 2018.

# CADERNO DE EMENTA DATA SCIENCE

## PROJETOS II

Créditos	Carga Horária	Disciplina
2	30 horas	Obrigatória

### Objetivos

Esta disciplina tem por objetivo o aluno escrever um artigo final para o curso de ciências de dados. Esse artigo deve usar técnicas apresentadas ao longo do curso para responder uma pergunta aplicada. Esse artigo deve ser desenvolvido com base do projeto de pesquisa apresentado na disciplina de Projetos I.

### Ementa

- 1- Entrega parcial 1: Introdução (motivando a pergunta a ser respondida e deixando claro qual é essa pergunta) e Revisão da Literatura
- 2- Entrega parcial 2: Introdução, Revisão da Literatura e Montagem do banco de dados
- 3- Entrega parcial 3: Introdução, Revisão da Literatura, Montagem do banco de dados e Metodologia
- 4- Entrega parcial 4: Introdução, Revisão da Literatura, Montagem do banco de dados, Metodologia e Apresentação dos Resultados e Discussão dos mesmos.
- 5- Entrega final: artigo aplicado completo.

Todas as entregas devem conter a assinatura do orientador do projeto de pesquisa.

### Bibliografia Básica