Especialização em Análise de Dados para o Controle

PLANO INSTRUCIONAL DE DISCIPLINA

|  |
| --- |
| **Disciplina: Modelos de Regressão Modalidade: Presencial Carga horária: 8h/aula** |
| **Professor responsável:** Thiago Paulo Faleiros (Doutor) |
| **Currículo Resumido:** Possui graduação em Ciência da Computação pela Universidade Federal de Goiás (2007), mestrado em Ciência da Computação pela Universidade Estadual de Campinas (2010) e doutorado em Ciência da Computação pela Universidade de São Paulo (2016), com período sanduíche na University of Maryland, EUA. Tem experiência na área de Ciência da Computação, com ênfase em Inteligência Artificial, atuando principalmente nos seguintes temas: algoritmos, modelos probabilísticos de tópicos, extração de informação, teoria dos grafos, aprendizado não supervisionado e otimização combinatória. |
| **Endereço do currículo Lattes:** <http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4131269Z5> |
| **Ementa da disciplina:** Esta disciplina trata da análise por regressão, mínimos quadrados e inferência utilizando modelos de regressão. Casos especiais de modelos de regressão, como o ANOVA, ANCOVA e a regressão logística serão tratados também. A disciplina abordará técnicas recentes de seleção de modelos e novos usos dos modelos de regressão. |
| **Objetivo geral**: *Formular em parceria com o coordenador acadêmico*. |
| **Objetivos específicos:** *Formular em termos de competências a serem adquiridas, em parceria com o coordenador acadêmico*. |
| **Avaliação da disciplina:** *exemplos: trabalho individual, trabalhos em grupo ao longo da disciplina, prova escrita etc*. |
| **Bibliografia Básica:** *Material que deverá ser lido pelos alunos para acompanhamento da disciplina – dar preferência a material que possa ser compartilhado no ambiente virtual do curso*. |
| **Bibliografia Complementar:** *Material para leitura opcional ou que possa servir de subsídio para elaboração de trabalhos da disciplina*. |

**CRONOGRAMA GERAL DA DISCIPLINA**

|  |  |  |  |
| --- | --- | --- | --- |
| **Aula** | **Data** | **Horário** | **Conteúdo** |
| **1** | 19/11/2018 | 8h-18h | *Formular em parceria com o coordenador acadêmico*. |
| **2** | 26/11/2018 | 8h-18h | *Formular em parceria com o coordenador acadêmico*. |

**PLANEJAMENTO DAS AULAS**

| **AULA** | **Objetivos** | **Conteúdo programático** | **DESENVOLVIMENTO (cronograma/estratégias)** | **Recursos**  **(equipamentos/ materiais)** | **Avaliação**  **(metodologia/ critérios)** |
| --- | --- | --- | --- | --- | --- |
| **Aula 1 (8h)**  **19/11/2018**  **8h às 18h** | *Ao final desta aula os alunos serão capazes de:*   * *compreender os conceitos fundamentais para aplicação de técnicas de regressão linear.* * *Identificar potenciais problemas solucionáveis com técnicas de regressão.* * *Implementar um regressor linear simples utilizando a linguagem python.* * *Avaliar resultados provenientes de técnicas de regressão e regressão multivariada.* * *Compreender técnicas para seleção de modelos* * *compreender técnicas de teste de variância e diferença entre médias de grupos de amostras de dados* | * *Conceitos fundamentais de estatística e probabilidade.* * *Regressão via otimização do erro mínimo quadrático* * *Implementando “do zero” um modelo de regressão linear simples em python.* * *Calculando medidas para avaliação do modelo* * *Cálculo da correlação de variáveis do modelo* * *Implementação de um modelo de regressão multivariado* * *Regressão baseado em método de gradiente descendente.* * *Aplicação de regressores em dados* | *30 min – Apresentação inicial:*   * *Motivação e caracterização de problemas de regressão.*   *1h – Fundamentação e revisão de estatística e implementação de medidas de cálculos estatísticos em python.*  *30 mim – Explicação de técnicas de otimização baseada em otimização via gradiente*  *15 min – Intervalo*  *1h min – Formulação e desenvolvimento das equações para obtenção de um algoritmo de regressão linear simples baseado na minimização do erro quadrático.*  *1h – Aplicação em ambiente Jupyter do algoritmo implementado.*  *- conceitos básicos de python e pandas*  *- plotar dados para visualização*  *30 min – Discussão dos resultados do exercício em conjunto.*  *30 min – Regressão Multivariada* | *- Projetor*  *- Computador para apresentação*  *- Computadores para os alunos desenvolverem os trabalhos* | *A participação do grupo será contabilizado para fins de avaliação da disciplina.* |
| **Aula 2 (8h)**  **26/11/2018**  **8h às 12h** | * *compreender técnicas de análise de resíduos.* | * *Revisão de Teste de hipóteses*   1. *t-student test*   2. *distribuição F* * *Análise Residual de modelos Lineares* * *Seleção de Variáveis em modelos de Regressão Linear* * *Regressão Logística* * *Ridge Regression* * *Lasso Regression* * *Elastic-net Regression* * *Regressores Não-Lineares*   1. *Support Vector Machine*   2. *Random Forest*   3. *XGBoost* * *Técncias de Comites (Essembles) para regressão* * *Apresentação de um notebook Jupyter da análise do dataset* | *1h – Revisão de Teste de Hipóteses*  *1h – Análise Residual de modelos*  *2h – Seleção de variáveis em modelos de Regressão Linear.*  *Atividades prática aplicando técnica de seleção de váriáveis*  *Intervalo Almoço*  *1h – Ridge Regression, Lasso Regressioon e Elastic-net Regression. (Apresentação teórica das técnicas e demonstração da aplicação na biblioteca sklearn)*  *30min – Regressores Não-Lineares. Apresentação teórica das técnicas SVM, Random Forest e XGBoost.*  *30min – Técnicas baseadas em comitês para regressão*  *2h – Apresentação de um notebook jupyter para análise do dataset. Será realizada a análise do conjunto de dados “House Prices”. Conjunto de dados para predição de preços de casas. Serão aplicados os Regressores não lineares neste conjunto de dados.* | *- Projetor*  *- Computador para apresentação*  *- Computadores para os alunos desenvolverem os trabalhos* | A avaliação final será em trabalho extra-classe. O trabalho é referente ao processo de pré-processamento, engenharia de características, e seleção de modelos em um conjunto de dados a definir. |