



## PROGRAMA DA DISCIPLINA

### DISCIPLINA:

APRENDIZADO PROFUNDO I

CODCRED	CARGA HORÁRIA	MÓDULO
98H09-04	60	60

### EMENTA:

Conceito de Aprendizado Profundo. Redes neurais profundas como estratégia para aprendizado automático de representações (features). Apresentação detalhada de algoritmos de treinamento de redes neurais profundas. Discussão sobre os principais desafios no projeto, na otimização de hiperparâmetros e na avaliação de redes neurais profundas. Apresentação e prototipação de diversas arquiteturas de redes neurais profundas para tarefas que envolvem dados não estruturados.

### OBJETIVOS:

O cumprimento da disciplina busca dar ao aluno, ao final do semestre, condições de:

1. Identificar as diferenças de aprendizado profundo e outras técnicas de aprendizado de máquina;
2. Explicar os elementos fundamentais do projeto de redes neurais profundas;
3. Comparar as principais arquiteturas de redes neurais profundas, identificando vantagens e desvantagens de cada arquitetura estudada;
4. Prototipar modelos de redes neurais profundas para diversas tarefas.

### CONTEÚDO:

Unidade de Aprendizagem 1: Introdução ao Aprendizado Profundo (30%)

- 1.1. Introdução ao Aprendizado Profundo
  - 1.1.1. Aprendizado automático de representações
- 1.2. Perceptron
  - 1.2.1. Algoritmo de Aprendizado do Perceptron
  - 1.2.2. O problema do ou-exclusivo
  - 1.2.3. Problemas de otimização da função sinal
- 1.3. Perceptron Multicamadas
  - 1.3.1. Algoritmo de retropropagação de erros
  - 1.3.2. Funções de Custo
  - 1.3.3. Grafo Computacional
  - 1.3.4. Descida de Gradiente





- 1.3.4.1. ☐ Batch, Minibatch e Estocástica
- 1.3.4.2. ☐ Com e sem momento
- 1.3.5. ☐ Funções de Ativação
- 1.3.5.1. ☐ Propriedades
- 1.3.5.2. ☐ Derivadas
- 1.3.6. ☐ Algoritmos para inicialização de pesos
- 1.3.6.1. ☐ Problemas da Inicialização Aleatória de Pesos
- 1.3.6.2. ☐ Inicialização He
- 1.3.6.3. ☐ Inicialização Xavier
- 1.3.7. ☐ Algoritmos para regularização
- 1.3.7.1. ☐ Regularização L1 e L2
- 1.3.7.2. ☐ Batch Normalization
- 1.3.7.3. ☐ Drop Out
- 1.3.7.4. ☐ Early Stopping
- 1.3.8. ☐ Otimização de hiper-parâmetros
- 1.3.9. ☐ Problemas de explosão e de dissipação do gradiente (Exploding and Vanishing Gradient)
- 1.3.10. ☐ MLP Aplicado a dados não estruturados
- 1.3.10.1. ☐ Classificação de Imagens
- 1.3.10.2. ☐ Classificação de Texto

Unidade de Aprendizagem 2: Redes Convolucionais (30%)

- 2.1. ☐ Limitações do Perceptron Multicamadas para tarefas de Visão Computacional
- 2.2. ☐ Motivação da convolução como forma de introduzir viés indutivo
- 2.3. ☐ Redes Convolucionais
- 2.3.1. ☐ Camadas Convolucionais
- 2.3.2. ☐ Camadas de Agrupamento (Pooling)
- 2.3.3. ☐ Conexões Residuais
- 2.3.4. ☐ Visualização e explicabilidade
- 2.3.4.1. ☐ Oclusão de Imagens
- 2.3.4.2. ☐ Mapas de Saliência
- 2.3.4.3. ☐ Gradient Based Class Activation Maps (GradCAM)
- 2.3.5. ☐ Introdução a Pré-Treino e Transferência de Aprendizado
- 2.3.6. ☐ Introdução a Aprendizado Auto-supervisionado
- 2.4. ☐ Aplicações
- 2.4.1. ☐ Classificação de Imagens
- 2.4.2. ☐ Detecção de Objetos
- 2.4.3. ☐ Segmentação Semântica de Imagens

Unidade de Aprendizagem 3: Redes Recorrentes (15%)

- 3.1. ☐ Motivação para modelos sequenciais
- 3.2. ☐ Grafo computacional ao longo do tempo
- 3.3. ☐ Redes Recorrentes
- 3.3.1. ☐ Retropropagação de Erros através do tempo
- 3.3.2. ☐ Encoder-Decoder





- 3.3.3. ☐ Problema da dissipação do gradiente (vanishing gradient) para redes recorrentes
- 3.3.4. ☐ LSTM (Long Short Term Memory)
- 3.3.5. ☐ GRU (Gated Recurrent Unit)
- 3.3.6. ☐ Redes Recorrentes Bidirecionais
- 3.3.7. ☐ Problemas do treinamento sequencial
- 3.4. ☐ Aplicações de Redes Recorrentes

#### Unidade de Aprendizagem 4: Atenção e Transformers (25%)

- 4.1. ☐ Mecanismo de Atenção
  - 4.1.1. ☐ Atenção aplicada a Redes Neurais Recorrentes
  - 4.1.2. ☐ Atenção aplicada a CNN
- 4.2. ☐ Transformers
  - 4.2.1. ☐ Camadas de Atenção
    - 4.2.1.1. ☐ Self-Attention
    - 4.2.1.2. ☐ Multi-Head Attention
  - 4.2.2. ☐ Encoding Posicional
  - 4.2.3. ☐ Encoder-Decoder
  - 4.2.4. ☐ Viés indutivo
  - 4.2.5. ☐ Comparação de Redes Recorrentes e Transformers
  - 4.2.6. ☐ Arquiteturas
    - 4.2.6.1. ☐ Somente com Encoder
    - 4.2.6.2. ☐ Somente com Decoder
    - 4.2.6.3. ☐ Com Encoder e Decoder
  - 4.2.7. ☐ Vision Transformers

#### PROCEDIMENTOS E CRITÉRIOS DE AVALIAÇÃO:

A nota de G1 será calculada da seguinte maneira:

$$G1 = (P1 + P2 + 2*MT) / 4$$

onde:

P1: Prova 1, abrange as unidades 1 e 2,

P2: Prova 2, abrange as unidades 3 e 4,

MT: Média dos trabalhos práticos que abrangem todas as unidades

#### BIBLIOGRAFIA BÁSICA:

1. ☐ NIELSEN, Michael A. Neural networks and deep learning. San Francisco, CA, USA: Determination press, 2015. (<http://neuralnetworksanddeeplearning.com/index.html>)
2. ☐ GOODFELLOW, I., BENGIO, Y., COURVILLE, A. Deep Learning. MIT Press, 775p., 2016. (<https://www.deeplearningbook.org/>)





3. □ZHANG, Aston et al. Dive into deep learning. arXiv preprint arXiv:2106.11342, 2021. Disponível online em <https://d2l.ai/index.html#>

#### **BIBLIOGRAFIA COMPLEMENTAR:**

1. □HAYKIN, Simon. Neural networks and learning machines, 3/E. Pearson Education India, 2009.

2. □WEIDMAN, Seth. Deep Learning from Scratch: Building with Python from First Principles. O`Reilly Media, 2019.

3. □REDMON, J., DIVVALA, S., GIRSHICK, R., FARHADI, A. You only look once: Unified, real-time object detection. In CVPR, 2016.

4. □VASWANI, Ashish et al. Attention is all you need. Advances in neural information processing systems, v. 30, 2017.

5. □YAN LECUN, Yann et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, v. 86, n. 11, p. 2278-2324, 1998.

