

# Inteligência Computacional

## Especificação do Projeto III

### Regressão com Redes Neurais

#### Objetivo

---

Desenvolver uma solução completa para prever o preço de venda de uma casa com base em 79 variáveis explicativas que descrevem (quase) todos os aspectos das residências. Neste projeto, os alunos também terão que preparar a base de dados antes de construir a solução. Ao contrário do projeto anterior, esse é um projeto de regressão, ou seja, a variável alvo é contínua. A base de dados já está na granularidade do projeto.

#### Descrição da base de dados

---

A base de dados para o projeto é `ic_house_pred_train.csv` e `ic_house_pred_test.csv` que estão disponíveis no Github da disciplina. A descrição da base de dados pode ser encontrada no site abaixo:

<http://web.stanford.edu/class/stats191/data/amesdoc.txt>

OBS1.: A base de dados usada na disciplina é um **subconjunto** dos dados originais selecionado pelo prof. da disciplina.

OBS2.: A base de dados da disciplina possui **81 variáveis** (já eliminada a primeira variável: *Order (Discrete): Observation number*).

#### Etapas do Projeto

---

- 1- Importar as bases de dados;
- 2- Eliminar a variável irrelevante *PID (Nominal): Parcel identification number*;
- 3- Tratamento dos *Missing Values*; (Apenas colocar para a média)
- 4- Transformação da variável alvo com a função logaritmo natural `np.log(x)`;
- 5- Exportar duas bases de dados:

- *A: Missing* para a média e variável alvo original;
- *B: Missing* para a média e variável alvo transformada pela função logaritmo natural;

5- Execução dos experimentos. (Rede Neural MLP no Keras);

- Variar número de parâmetros livres do modelo (influenciado pelo número de neurônios na camada escondida). Escolher três opções (justifique sua escolha).
- Variar a taxa de aprendizado. Escolher três opções (justifique sua escolha).

6- Apresentação dos resultados: com e sem a transformação logarítmica na variável alvo (incluir análise dos resultados do item 5). Usar as seguintes métricas de avaliação de desempenho:

- $r^2$ ;
- Gráfico de Dispersão;
- Erro médio absoluto (MAE, *Mean Absolute Error*).

7- Fazer uma conclusão contendo entre 200 e 250 palavras.