



Inteligência Artificial Computacional

Prática 1 AV1 - Modelo Preditivo 1 (Regressão Linear Múltipla)

Professor: Prof. Msc. Paulo Cirillo Souza Barbosa

Você é um projetista de modelos de IA, e seu cliente te pediu para desenvolver um sistema que lhe auxilia em suas pesquisas. O cliente em questão, é um profissional da área da química que faz diversos experimentos de níveis de solubilidade a partir de relações estruturais de componentes químicos. Para poder desenvolver o modelo preditivo, você solicitou ao cliente os dados coletados em suas pesquisas. Contudo, em um momento inicial, o cliente resolveu lhe fornecer parte dos experimentos, e lhe enviou dados referentes a 951 amostras com os preditores (variáveis regressoras, ou variáveis independentes) **quantidade de carbono** (x_1) e **peso molecular** (x_2). Em conjunto dessas informações, ele também lhe enviou as 951 medições da de solubilidade (variável dependente) realizadas. **Observação:** Esses dados podem ser encontrados no AVA, bloco IV, no arquivo chamado *Solubilidade2.csv*. O arquivo disponibilizado contém 951 linhas e 3 colunas, em que as duas primeiras colunas são as variáveis regressoras e a terceira coluna trata-se da variável dependente (solubilidade). Com essas informações faça o que se pede:

- Dada as informações fornecidas, quais são os valores de N e p ?
- A tarefa a ser resolvida pelo modelo, é de regressão ou classificação? Justifique.
- Carregue os dados com numpy utilizando o método **np.loadtxt()**, e em seguida plote o gráfico de espalhamento. Neste caso, tente fornecer ao gráfico, o máximo de informações (rótulos dos eixos x,y e z, título da imagem, cor diferente ao espalhamento...).
- Você deve projetar um sistema inteligente que aprende os padrões fornecidos, ou seja, que captura a relação entre ($x_{numeroCarbono}$, $x_{pesoMolecular}$ e $y_{solubilidade}$). Assim, pede-se que utilize **hipóteses** os seguintes modelos:
 1. A equação linear em que β_1 até β_p tem o valor nulo, e que tenha β_0 um valor definido. A estimativa de $\hat{\beta}_0$ (algoritmo de aprendizagem) deve ser realizada via média dos valores da variável dependente (ou média da variável observável).
 2. A equação linear de regressão múltipla (que tenha todos seus parâmetros definidos, β_0 até β_p) e se faça a estimação de seus parâmetros (**algoritmo de aprendizagem**) via Método dos Mínimos Quadrados Ordinários (MMQO).
- Após os modelos serem estimados, o cliente te forneceu uma única amostra sem nível de solubilidade. Essa entrada aos modelos desenvolvidos, trata-se do valor $x_1 = 11$ e $x_2 = 190,27$. Para esses casos, qual é a predição de solubilidade realizada para cada modelo?