

ÁLGEBRA LINEAR NUMÉRICA
AULA PRÁTICA 4
MÉTODO DOS MÍNIMOS QUADRADOS

- 1) Resolva o exercício 32, da seção 7.3, página 588, do livro Álgebra Linear, tradução da 4ª.edição americana, David Poole. Relate a modelagem utilizada e use o SciLab para os cálculos.
- 2) Resolva o exercício 33, da seção 7.3, página 588, do livro Álgebra Linear, tradução da 4ª.edição americana, David Poole. No item b), pesquise a população real dos EEUU em 2010 e compare com a população que você estimou. Relate a modelagem utilizada e use o SciLab para os cálculos.
- 3) Resolva o exercício 34, da seção 7.3, página 588, do livro Álgebra Linear, tradução da 4ª.edição americana, David Poole. Relate a modelagem utilizada e use o SciLab para os cálculos.

- 4) Agora vamos usar o método dos mínimos quadrados para implementar um método rudimentar de “machine learning” para diagnosticar câncer de mama a partir de um conjunto de características fornecidas para cada paciente. São dados dois arquivos: um arquivo para “treinamento” (cancer_train.csv) do modelo e um arquivo para “teste” (cancer_test.csv). O primeiro arquivo contém 300 registros e o segundo 260 registros, partes do “Wiscosin Diagnostic Breast Cancer dataset”. Cada registro de cada arquivo contém 11 valores: os 10 primeiros correspondem a valores reais de 10 características dos núcleos celulares observados em imagens digitalizadas de uma fina camada de massa mamária coletada de cada paciente. O décimo primeiro valor é +1 se a paciente tem câncer de mama e -1, caso contrário.

Sendo \mathbf{x} o vetor das 10 características de cada paciente (variáveis independentes) e y o valor (+1 ou -1) que indica o diagnóstico (variável dependente), a ideia é, usando o arquivo de treinamento, obter o hiperplano

$$y = h(\mathbf{x}) \quad (y = \alpha_0 + \sum_{i=1}^{10} \alpha_i x_i) \text{ que “melhor se ajuste aos dados fornecidos”}$$

usando o método dos mínimos quadrados.

Uma vez obtido o hiperplano, o mesmo será usado para classificar cada paciente da seguinte forma: se $h(\mathbf{x}) \geq 0$, então o diagnóstico é +1 (tem câncer), caso contrário, o diagnóstico é -1 (não tem câncer).

Use o seu classificador (hiperplano) e calcule a porcentagem de acertos sobre o arquivo de treinamento (de certa forma é uma medida do ajuste do seu modelo aos dados de treinamento) e sobre o arquivo de teste (de certa forma é uma medida da capacidade de generalização do seu modelo).

Construa uma Matriz de Confusão (Confusion Matrix) (pesquise a respeito) com o conjunto de teste e calcule as diversas medidas daí decorrentes, tais como: acurácia, precisão, recall, probabilidade de falso alarme, probabilidade de falsa omissão de alarme. Interprete essas medidas e comente os resultados obtidos.