

# Relatório - Implementação do Model Oculto de Markov (HMM)

Guilherme Albarrans e Rodrigo Caland

July 2023

Neste trabalho, foi realizada a implementação do modelo oculto de Markov univariado assumindo independência condicional entre as variáveis observáveis dado o estado oculto, conforme apresentado em aula. O pseudocódigo da nossa implementação pode ser encontrado no notebook *relatorio.ipynb*.

As implementações desenvolvidas foram comparadas com os outputs dados pelo HMMTeacher, uma plataforma online construída por acadêmicos da área de biologia computacional. O HMMTeacher recebe como inputs uma sequência de observações, uma distribuição inicial, uma matriz de transição e as probabilidades de emissão. A partir disso, ele gera a função de verossimilhança calculada pelo método Forward-Backward e os estados ocultos obtidos através do algoritmo de Viterbi. Os testes comparativos com o HMMTeacher podem ser encontrados no notebook *validacao.ipynb*.

Para realizar a comparação foi considerado um HMM com dois estados ocultos ( $k=2$ ) e dois possíveis outputs ( $Y=0$  ou  $Y=1$ ). Os parâmetros fixados estão discriminados a seguir:

- Distribuição inicial:

$$P(Y = 0) = 0.122077 ; P(Y = 1) = 0.87793$$

- Matriz de transição:

$$\begin{bmatrix} 0.52106 & 0.47894 \\ 0.63783 & 0.36217 \end{bmatrix}$$

- Probabilidades de emissão:

- No regime 1:

$$P_1(Y = 0) = 0.68121 ; P_1(Y = 1) = 0.31879$$

- No regime 2:

$$P_2(Y = 0) = 0.083737 ; P_2(Y = 1) = 0.91627$$

Além disso, uma amostra de 0s e 1s de tamanho 47 gerada aleatoriamente foi fixada.

O valor da função de verossimilhança calculado pelo HMMTeacher para a configuração descrita acima foi de  $4,475442e-15$ , enquanto a verossimilhança obtida usando a implementação do nosso grupo foi de  $4,475442e-15$ . Com relação aos estados ocultos obtidos, verificou-se que houve concordância de 100% entre o output do HMMTeacher e o output da implementação do grupo, ou seja, os estados ocultos obtidos pelo Algoritmo de Viterbi nas duas implementações distintas foram iguais em 100% do número total de observações. Ademais, a probabilidade conjunta da sequência de estados ocultos na nossa implementação foi de  $7,433071e-20$ , enquanto a mesma quantidade na implementação do HMMTeacher foi de  $7,433071e-20$ .