

Modelos Markovianos

Trabalho No.3

Entregar até o dia 3 de novembro de 2025.

O Hidden Markov Model (HMM) sendo o termo em português Modelos Ocultos de Markov, é uma técnica de modelagem geral adequada para representar uma sequência de recursos ocultos no tempo ou no espaço, em que cada recurso oculto causa ou emite uma observação. Na prática, dado um modelo, uma sequência observada é a entrada, para a qual o HMM mais básico calcula sua probabilidade ou produz uma previsão da sequência mais provável de recursos ocultos nela.

Um dos primeiros exemplos de aplicação conhecidos foi encontrar a sequência mais provável de letras ou palavras em uma sequência de sons ou reconhecimento de fala, foi o primeiro problema que inspirou o desenvolvimento do HMM nas décadas de 1960 e 1970.

O foco na modelagem de variáveis de um HMM é responder, através dos algoritmos HMM mais básicos Forward/Backward (BaumWelch) e Viterbi 2 questões:

- (1) Qual é a probabilidade de uma sequência observada dado um modelo? e
- (2) Qual é a sequência mais provável de recursos ocultos nele?

Por fim, interpretar os resultados em uma análise dos parâmetros à luz do problema modelado.

Modelo oculto de Markov sobre dados do mercado de ações:

A previsão do mercado de ações tem sido uma das áreas de pesquisa mais ativas no passado, dado o interesse óbvio de muitas grandes empresas. Historicamente, vários algoritmos de aprendizado de máquina foram aplicados com vários graus de sucesso. No entanto, a previsão de estoque ainda é severamente limitada devido à sua natureza não estacionária, sazonal e imprevisível. Fazer previsões apenas com base nos dados de ações anteriores é uma tarefa ainda mais desafiadora, pois ignora vários fatores externos.

Os HMMs são capazes de modelar transições de estado oculto a partir dos dados observados sequenciais. O problema de previsão de estoque também pode ser pensado seguindo o mesmo padrão. O preço da ação depende de uma infinidade de fatores, que geralmente permanecem invisíveis para o investidor (variáveis ocultas). A transição entre os fatores subjacentes muda com base nas políticas e decisões da empresa, suas condições financeiras e decisões de gestão, e estas afetam o preço da ação (dados observados).

Portanto, os HMMs são um ajuste natural para o problema de previsão de preços. Como aplicação, considere os preços históricos do ouro em:

```
> base_dir = "https://github.com/natsunoyuki/Data_Science/  
            blob/master/gold/gold/gold_price_usd.csv?raw=True"  
> data = read.csv(base_dir)  
> head(data)
```

	datetime	gold_price_usd
1	1978-12-29	137.06
2	1979-01-01	137.06
3	1979-01-02	137.29
4	1979-01-03	134.01
5	1979-01-04	136.79
6	1979-01-05	138.35

A descoberta de preços é crucial para qualquer mercado. O ouro não só tem um preço à vista, mas também tem o preço do ouro LBMA, bem como vários preços regionais. O preço do ouro LBMA é usado como uma referência importante em todo o mercado de ouro, enquanto os outros preços regionais do ouro são importantes para os mercados locais.

Este conjunto de dados fornece o preço do ouro em vários períodos (diário, semanal, mensal, anual) desde 1978 e nas principais moedas comerciais, produtoras e de consumo. O endereço fornecido acima acessa diretamente os dados diários do ouro em USD, ou seja, em dólares americanos até 2021.

```
> gold.price = data$gold_price_usd
> datas = as.Date(data$datetime)
> par(mar=c(3,4,2,1))
> plot(datas,gold.price, ylab = "USD", xlab = "", type = "l",
      main = "Preços históricos do ouro em USD")
> grid()
```

Em vez de modelar o preço do ouro diretamente, modelamos a variação diária do preço do ouro, isso nos permite capturar melhor o estado do mercado. Para isso fazemos:

```
> retornos = diff(log(gold.price))
> par(mar=c(3,4,2,1))
> plot(datas[-1], retornos, ylab = "", xlab = "", type = "l",
      main = "Variação diária do preço do ouro")
> grid()
```

Obtém-se assim os dados que serão utilizados na análise, os retornos ou log-retornos.

- Sugere-se ajustar a variação diária do preço do ouro a um modelo de emissões Gaussiano com 2 e 3 estados ocultos. A possível razão para usar até 3 estados ocultos é que esperamos pelo menos 3 regimes diferentes nas mudanças diárias: baixa, média e alta de volatilidade.
- Uma vez identificado o modelo mais adequado encontre a matriz de probabilidades de transição para os estados ocultos e observe os elementos diagonais: estes são grandes em comparação com os elementos fora da diagonal? Caso isto seja identificado significaria que o modelo tende a querer permanecer naquele estado específico em que se encontra, a probabilidade de transição para cima ou para baixa.
- Encontre a distribuição estacionária da Cadeia de Markov oculta, quais estados são transientes e quais recorrentes? Caso decida-se utilizar o preço do ouro como opção em uma carteira de investimentos, qual seria a probabilidade de obter ganho investindo em ouro?