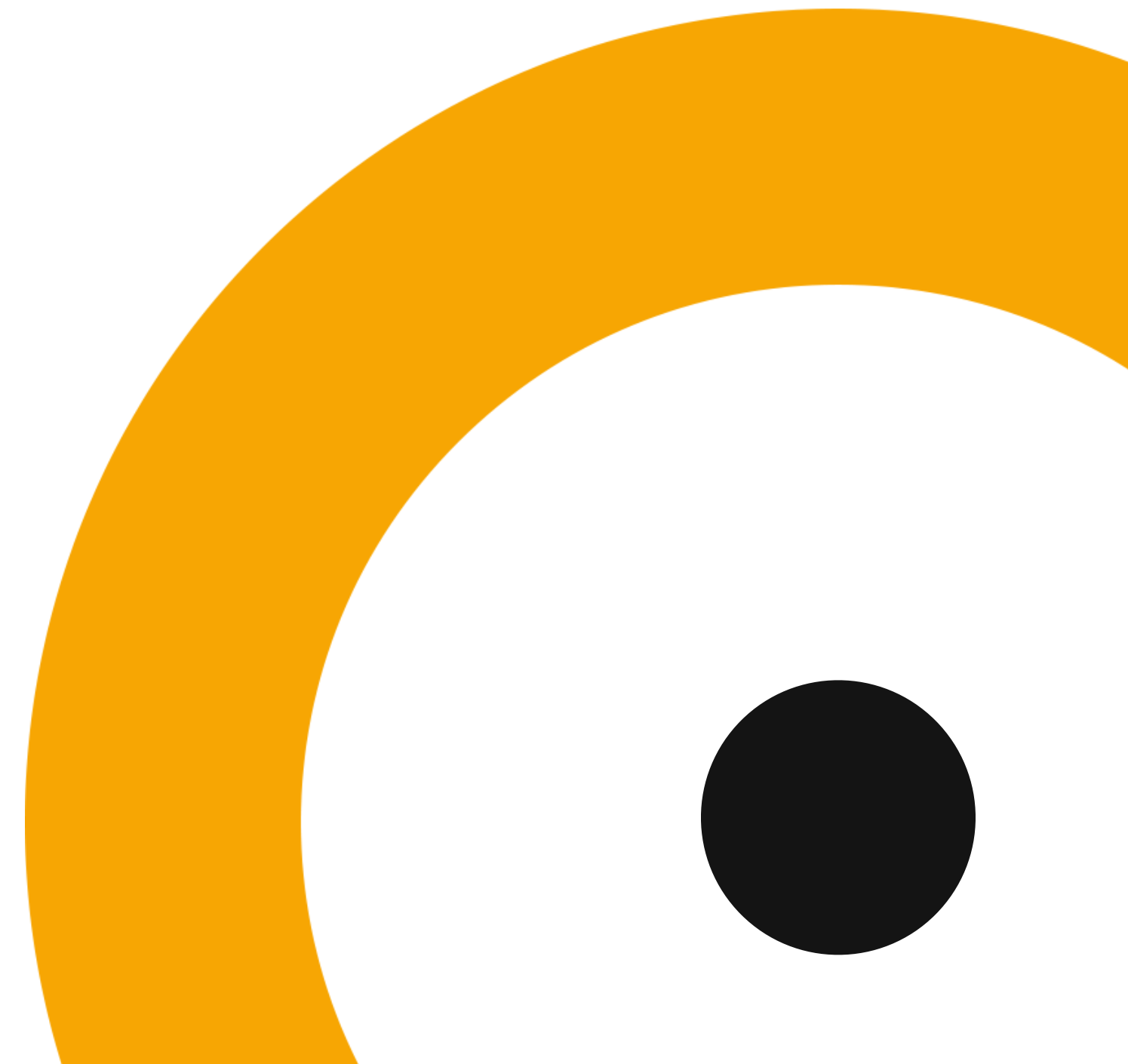


Mestrado em Data Science

Séries Temporais

Desafios Relativos a Tarefas de Previsão





Conteúdo

10.1 Desafios de Previsão na Prática

10.2 Estudo de Caso - Previsão



10.1

Desafios de Previsão

na Prática





Por que a previsão é difícil?

Auto-correlação

- Dependência temporal entre as observações
- O que vemos hoje depende do que vimos anteriormente

Suscetibilidade à Mudança - Não Estacionariedade

- Previsão é um processo de extrapolação
 - Assume-se que os padrões se mantêm para o futuro
- O fenómeno muda ao longo do tempo e a distribuição de dados também



Problemas Específicos Acerca de Previsão

1. Interesse em vários horizontes temporais
2. Interesse em casos raros
3. Múltiplas dependências
4. Pontos de mudança
5. Ruído
6. Amostras pequenas





1. Previsão a Longo-Prazo

Interesse em Múltiplos Horizontes

- Prever o próximo passo muitas vezes não é suficiente
- Precisamos prever vários pontos com antecedência
- Mais difícil à medida que tentamos prever um futuro mais longínquo
 - Prever a temperatura máxima de amanhã é mais fácil do que a temperatura máxima daqui a 2 meses

Dependência do Caminho (*Path Dependency*)

Suponha que queremos prever H passos com antecedência.

- O que acontece entre agora e H pode ser importante na tomada de decisão



1. Previsão a Longo-Prazo

Horizontes de Previsão na Competição M4 por Frequência

- Horária: 48
- Diária: 14
- Semanal: 13
- Mensal: 18
- Trimestral: 8
- Anual: 6



1. Previsão a Longo-Prazo

Métodos para previsão *multi-step*

- Tipicamente, métodos de aprendizagem lidam com um output
- Então são necessários métodos específicos para previsão *multi-step*

Métodos Principais

1. Recursivo

- a. Treina um único modelo para prever o ponto seguinte
- b. As previsões em pontos anteriores são usadas como input para prever o próximo ponto

2. Direto

- a. Treina um modelo para cada ponto de previsão

3. DirRec (Direto-Recursivo)

- a. Treina um modelo para cada ponto de previsão
- b. Cada modelo sucessivo adiciona as previsões dos modelos anteriores

4. Multi-output

- a. Um modelo para todo o horizonte



2. Interesse em Casos Raros

Valores Extremos

- Exemplos:
 - Picos de consumo de energia
 - Grandes subidas/descidas no preço de produtos financeiros
- Casos raros mas com grande impacto no domínio
- Poderão haver consequências a longo-prazo
- O problema é que, por definição, existe pouca informação sobre esses casos raros

Abordagens

- Usar distribuições desenvolvidas para valores extremos
 - Teoria de Valores Extremos (*extreme value theory*)
- Re-amostragem do conjunto de dados
 - Melhora a representatividade de casos raros
 - Mas existe um trade-off
- Introduzir custos mais elevados em casos raros



3. Múltiplas Dependências

Múltiplas Dependências

- Uma série temporal pode ter dependências adicionais para além do tempo
 - Espaço
 - Outras séries relacionadas

Dados Espaço-Temporais

- Dependências no Tempo+Espaço
 - As observações também dependem do que está acontecendo nas proximidades
- Localizações podem ou não ser fixas
 - Sensores (fixo)
 - Taxi GPS (não fixo)
- Modelar essas dependências é fundamental para construir modelos de previsão com bom desempenho
 - Por exemplo, com abordagens Auto-Regressive Distributed Lag



4. Detecção de Mudanças

Mudança (*Concept Drift*)

A distribuição dos dados muda à medida que a série evolui ao longo do tempo

Diferentes Tipos de Mudança

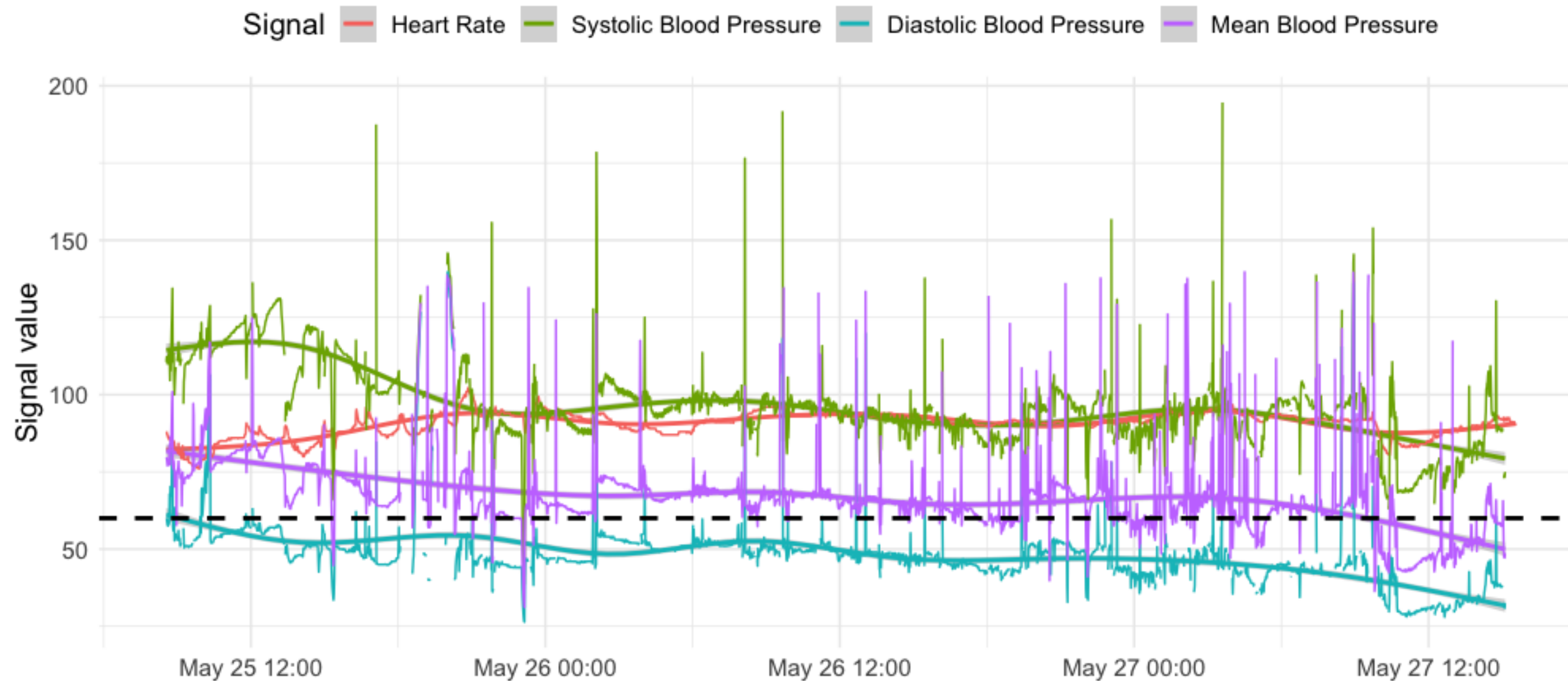
- Recorrente
 - Se a série alternar entre diferentes conceitos ou regimes
- Abrupta (Quebra estrutural)
 - Mudança na distribuição é dramática
 - Por exemplo, por causa de um evento externo
- Gradual
 - Mudança na distribuição é lenta
 - Exemplo: Interesses de utilizadores ao longo do tempo

Não confundir com *outliers*

5. Ruído

Séries temporais são muitas vezes caracterizadas por ruído e dados omissos

- sensores podem falhar temporariamente devido a avarias
- mudanças no equipamento de medição
 - e.g. substituição do sensor com parâmetros diferentes
- interferência nos sensores





6. Amostras Pequenas

- As séries temporais podem ser pequenas (baixa frequência)
- Mudança recente
- Problema *Cold start*

Previsão com Modelos Globais

Uma possível solução:

- Treinar modelos com várias séries temporais
 - *Deep Learning* é um tipo de modelos que segue esta abordagem



10.2

Estudo de Caso





Previsão para Energia Renovável

Duas questões impedem a disseminação de fontes de energia renováveis:

- **Alta variabilidade**
- **Dificuldade de previsão**





Altura significativa das ondas

Energia Renovável

- As ondas dos oceanos são uma fonte de energia renovável
- Prever o tamanho das ondas é fundamental para estimar a produção de energia através desta fonte.

Prevenção de Ondas Grandes

- Gestão de processos marítimos e a sua segurança:
 - Desastres na costa
 - Passagem de navios



Dados

Smart Buoy

- Junto da costa da Irlanda
- https://data.gov.ie/dataset/smartbay-ireland-cork-buoy-wave/resource/e58ebc3e-edcc-4bb8-9cc2-112da4465477?inner_span=True

Amostra

- https://github.com/vcerqueira/wave_height_data

Séries Temporais

Vitor Cerqueira

[https://www.linkedin.com/in/vcerq/
cerqueira.vitormanuel@gmail.com](https://www.linkedin.com/in/vcerq/cerqueira.vitormanuel@gmail.com)

