

Méliuz

Concept Drift

O que é, como acontece e como resolver?

Luísa Reis

Quem sou eu?

Luísa Reis

- Sistemas de Informação - UFAM
- Mestre em Informática - UFAM
- Engenheira de Dados - Méliuz



Méliuz



Índice

1. O que é?
2. Como Acontece?
3. Como Resolver?

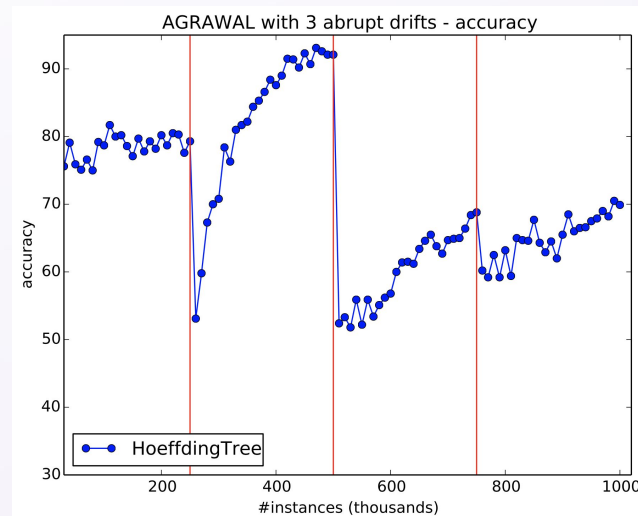


O que é Concept Drift?

O que é Concept Drift?

Mudanças implícitas que tornam o modelo defasado

- Modelos preditivos são criados com base em **dados históricos**
- Porém os **dados** podem sofrer **alterações ao longo do tempo**
- Algumas dessas mudanças são influenciadas por **fatores externos não previstos**
- Isso pode resultar na **degradação de modelos** preditivos que assumem uma relação estática entre o input e o output



Modelo para prever qual filme será mais assistido no próximo mês



Modelo para prever qual filme será mais assistido no próximo mês



Modelo para prever qual filme será mais assistido no próximo mês



Como Acontece?



Como Acontece?

Mudança Gradual

- Mudanças nos dados ocorrem de forma sutil
- Pode levar meses ou anos para ser detectada

Ex.: Detecção de Faces



Como Acontece?

Mudança Cíclica

- Mudança em determinados períodos de tempo
- Mudanças sazonais

Ex: Black Friday

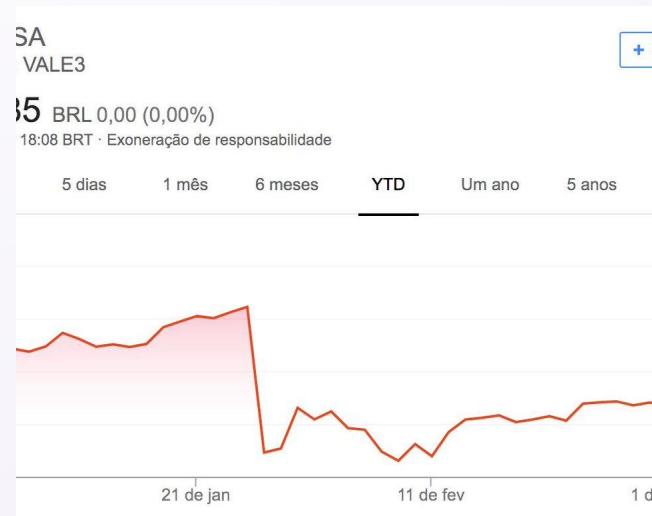


Como Acontece?

Mudança Abrupta

- Mudança brusca no padrão dos dados

Ex.: Ações da vale em queda por causa do desastre de Brumadinho



Como Resolver?

Como Resolver?

Não Fazer Nada!

- Podemos assumir que os dados não vão mudar com o tempo.
- Permite o desenvolvimento de um único modelo “ótimo”
- Importante monitorar para verificar se alguma anomalia irá acontecer
- Ex.: Modelo para diagnóstico de tuberculose.
 - Por mais que mudanças ocorram no corpo do paciente, dificilmente isso impactará o padrão de detecção da tuberculose.



Como Resolver?

Retreinar Periodicamente

- Atualizar o seu modelo estático com dados históricos mais recentes
- Importante identificar qual o período adequado a ser adicionado no modelo
- Podemos atualizar, por exemplo, um modelo anualmente com dados referentes ao ano anterior



Como Resolver?

Refinar Periodicamente

- Muito parecido com o retreinamento periódico
 - Mas não descarta o modelo anterior
 - Utiliza o modelo anterior como ponto de início para o processo de atualização.
- Adiciona uma amostra de dados históricos mais recentes
- Usado em algoritmos que usam pesos e coeficientes, como modelos de regressão.



Como Resolver?

Usar Pesos

- Adicionar pesos em variáveis que sejam inversamente proporcionais ao período do dado.
- Dessa forma, o modelo prestará mais atenção nos dados mais recentes (com peso maior).



Méliuz

Muito obrigado!
luisareis94@gmail.com

Referências

- <https://machinelearningmastery.com/gentle-introduction-concept-drift-machine-learning/>
- Alexey Tsymbal, The problem of concept drift: definitions and related work. Trinity College Dublin, Ireland — 2004.
- <https://medium.com/@ericabertan/o-que-%C3%A9-concept-drift-em-machine-learning-40ae3c4f0b67>