

Disciplina:

Processamento de linguagem natural

Professor: Gabriel Assunção



Apresentação do curso

Módulos

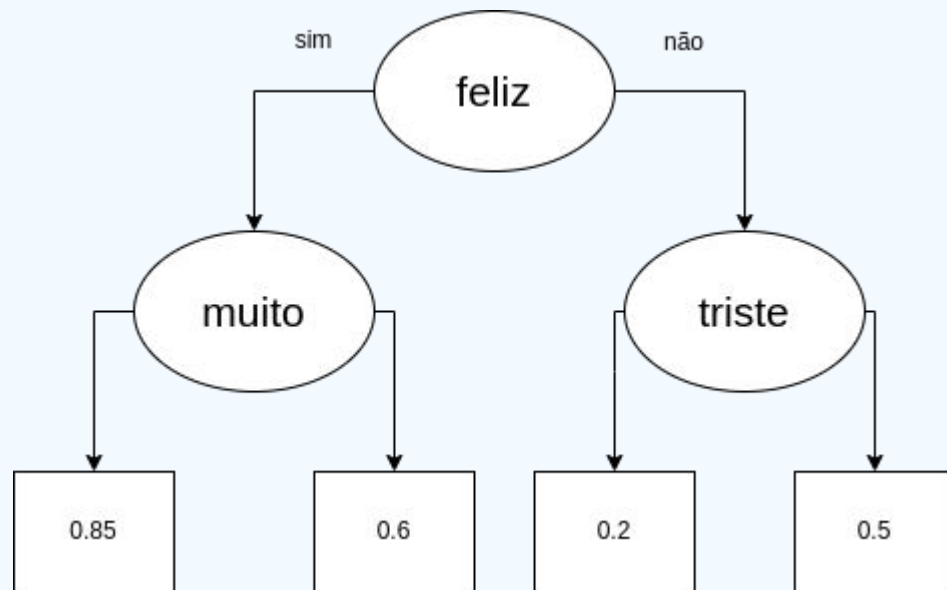
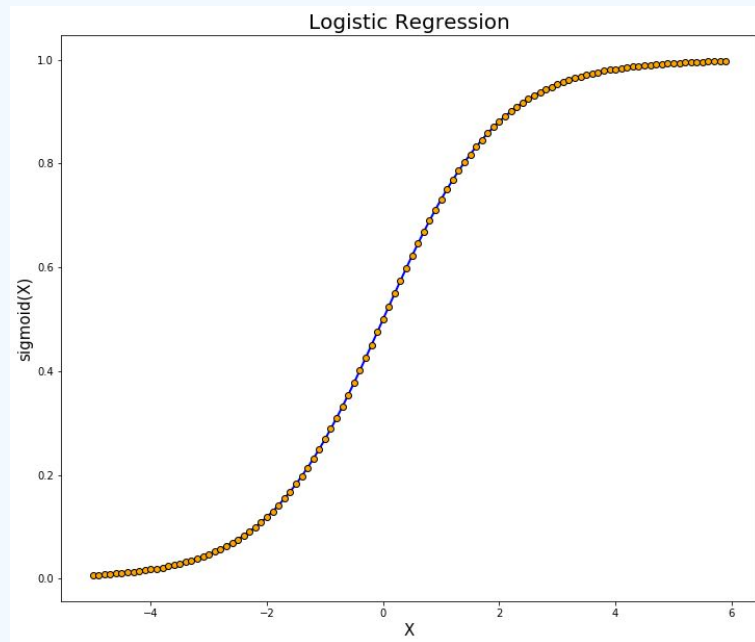
1. Introdução
2. Expressão regular
3. Processamento de texto
4. Representação textual
5. **Classificação de texto**
6. Modelos de NLP

Modelos

Modelos de ML

- Técnicas conhecidas em diversas aplicações de classificações.
- Aceitam diferente tipos de entrada de dados.
- No geral não consideram contextos das frases e a semântica.
- Mais fáceis de ser treinados e não é necessário volume muito grande de dados.
- Possuem formas de interpretação.

Modelos de ML



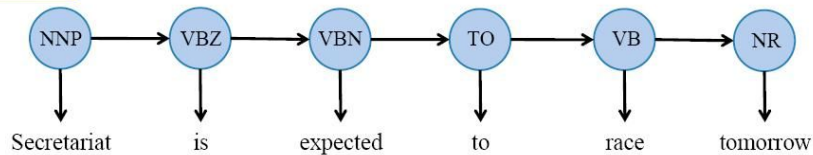
Modelos HMM, MEMM, CRF

- Modelos probabilísticos.
- Modelam a $P(c|x)$
- Modelam a transição de um rótulo para o outro
- Considera o contexto porém em uma ordem similar a leitura humana.
- Pode ser encaixado com outros modelos para ficar mais robusto.

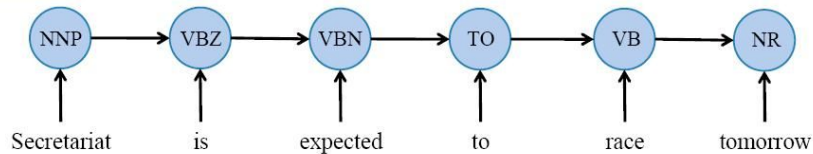
Modelos HMM, MEMM, CRF

HMM v.s. MEMM

HMM



MEMM



Modelo de LSTM

- Modelo de rede neural recorrente (RNN)
- Considera contexto (leitura em ambos os lados)
- Modelo black-box
- Melhor performance que os modelos anteriores
- Necessário maior volume de dados e maior custo para treinar

[Recomendação de leitura](#)

Classificação de textos

Classificação

- Objetivo: classificar o texto baseado em alguma classe.

Exemplos:

- Texto do e-mail é ou não um spam?
- Segmento do texto, é uma notícia esportiva, ou de tecnologia?
- Comentário em rede social é ou não discurso de ódio?
- Sentimento expressado na frase: raiva, alegria, neutro.
- Polaridade da frase: frase positiva ou negativa?

Classificação

- Entrada: representação textual e rótulos dos textos.
- Output: modelo/regras capazes de ou gerar a probabilidade ou mapeamento de um rótulo.
- Em alguns casos podemos ver qual é a associação de uma palavra ao rótulo. Por exemplo: o texto contém “feliz” é indicativo de um sentimento positivo.
- Ponto negativo de utilização de regras: complexidade da informação, nível de detalhamento necessário ser grande.

Como classificar

- Define a representação
- Separa o dado entre treino e teste
- Aplica um algoritmo
- Avalia métricas de performance (precisão, acurácia)

Análise de Sentimentos

Análise de sentimentos

- É um dos problemas de classificação mais conhecido.
- Aparece também como análise de polaridade, extração de opinião.
- Muito usado para avaliar o sentimento das pessoas respeito de um evento, de um produto, um filme, uma pessoa, política.
- Exemplos:
 - O que as pessoas pensam do candidato X?
 - Como as pessoas receberam o filme Y?
 - O que as pessoas estão comentando sobre o meu produto?
 - O que as pessoas estão achando da utilização do meu chatbot?
 -

Análise de sentimentos

- Para realizar a análise de sentimentos podemos optar por algoritmos com regras utilizando dicionário com polaridade das palavras e conjunto de expressão regulares, ou utilizar modelos de ML e DeepLearning.
- Quando há poucos dados rotulados uma opção é utilizar os dicionários
- Os sentimentos podem ser separados em:
 - Positivo ou negativo
 - Escala de muito positivo até muito negativo
 - Emoções: raiva, triste, alegre.
 - Intenção: quero comprar ou não quero comprar.

Desafios na análise de sentimentos

- Construção da frase. Exemplo: não estou feliz. O não traz um sentimento oposto para a palavra feliz.
- Contexto da frase, o mesmo trecho de frase pode ter significados diferentes a depender do contexto.
 - O que mais gostei foi tudo que eles fizeram.
 - O que eu menos gostei foi tudo que eles fizeram.
- Ironia e sarcasmo
- Comparações