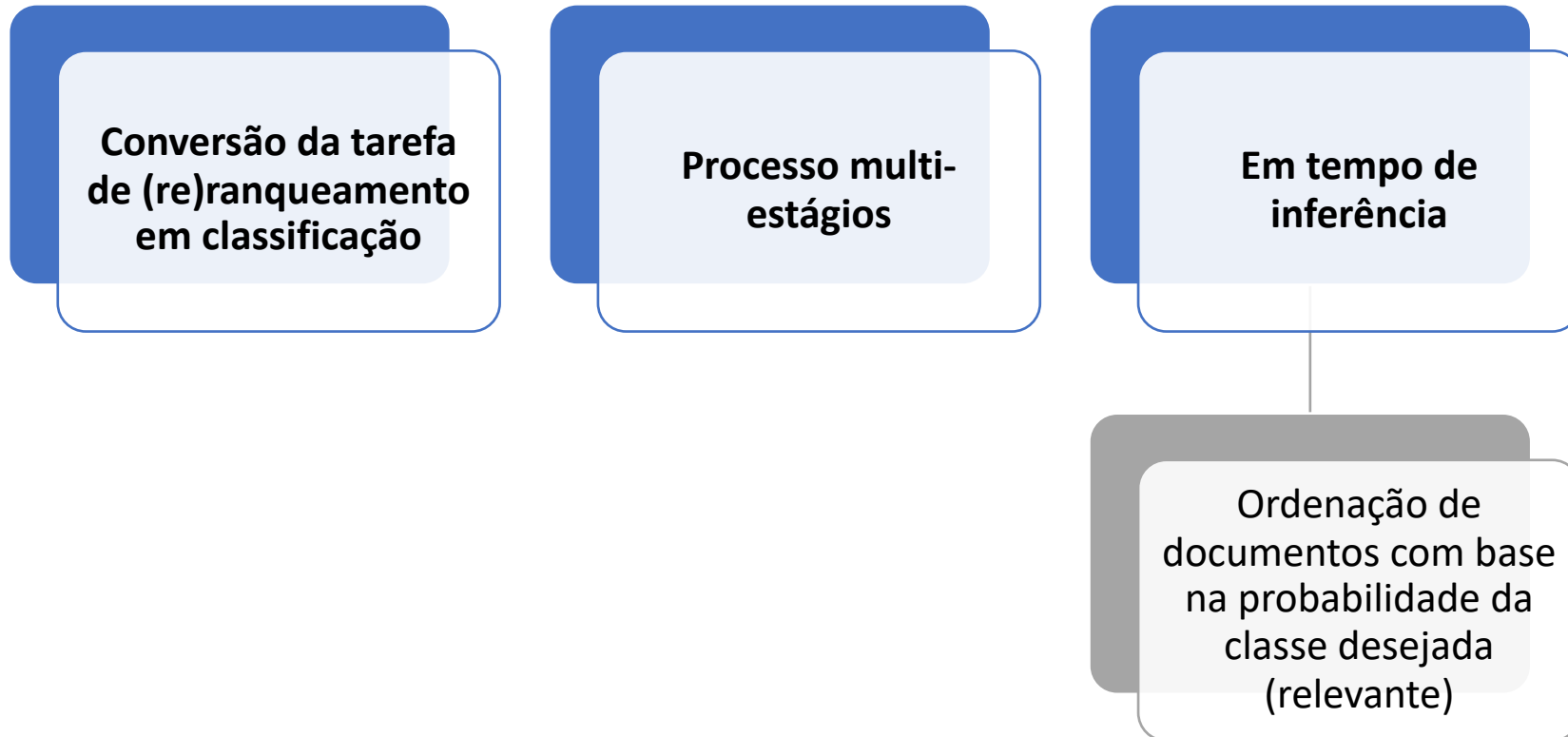
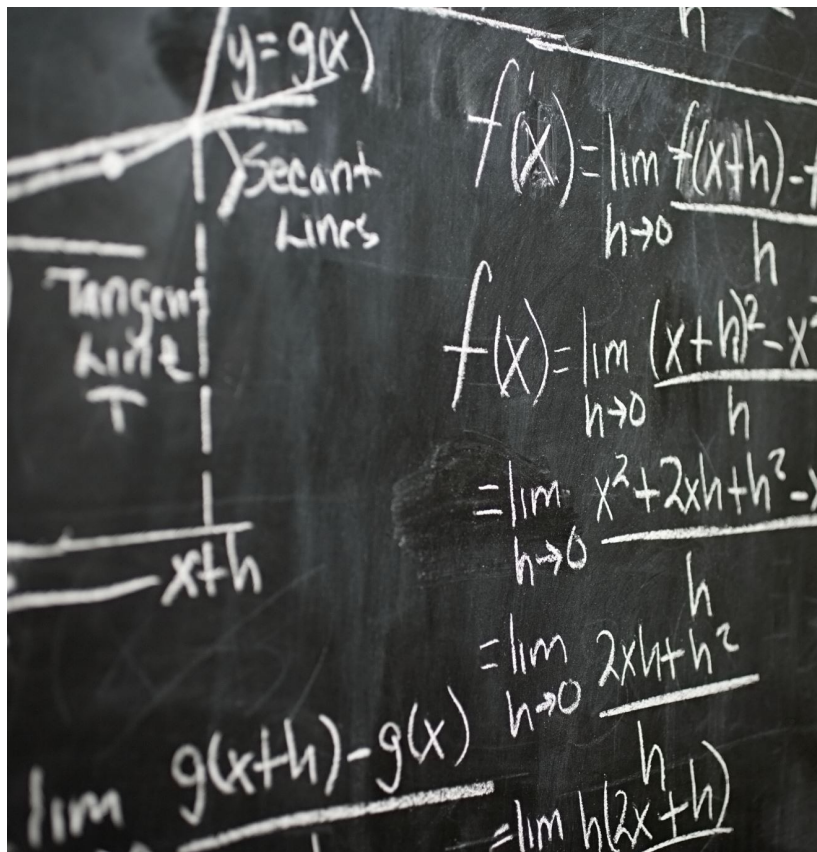


Classificação de Texto e Rerankeadores

Monique Monteiro – moniquelouise@gmail.com

Conceitos importantes





Experimentos / Resultados

- Melhor (último!) resultado obtido:
 - microsoft/MiniLM-L12-H384-uncased
 - nDCG@10 = 0,6316

Experimentos / Resultados (cont.)

- Histórico dos experimentos
 1. *microsoft/MiniLM-L12-H384-uncased*
 - Uso da função de ativação softmax: $nDCG@10 = 0,621$
 - Uso direto dos *logits*: $nDCG@10 = 0,6223$
 - Principais hiperparâmetros:
 - Taxa de aprendizagem (AdamW): 5×10^{-5}
 - 5 épocas
 2. *nreimers/MiniLM-L6-H384-uncased*
 - Com 20 épocas de treinamento, $nDCG@10 \cong 0,42$ ($< BM25$)
 3. Repetição dos experimentos com *microsoft/MiniLM-L12-H384-uncased*
 - Função de ativação softmax:
 $\overline{nDCG@10} = 0,6043, \sigma = 0,0172$
 - Uso direto dos logits:
 $\overline{nDCG@10} = 0,6041, \sigma = 0,0182$



Experimentos / Resultados (cont.)

- Histórico dos experimentos

- 4. Otimização de hiperparâmetros

- 10 épocas de treinamento: não gerou melhores resultados
 - *Grid search* para taxas de aprendizagem: $[2 \times 10^{-5}, 3 \times 10^{-5}, 5 \times 10^{-5}]$
 - 2×10^{-5} : **nDCG@10 = 0,6316**
 - 3×10^{-5} : nDCG@10 = 0,6315

- Principais conclusões:

- Não é possível afirmar que o uso dos *logits* leva a um melhor resultado em geral, dada a pouca diferença no valor médio para a média e ligeira variação para maior do desvio-padrão.
 - A busca em um espaço de 3 taxas de aprendizagem, apesar de ter levado ao modelo "vencedor" dentre aqueles aqui testados, não levou a variações muito significativas para a métrica NDCG@10.



Tópico avanzado

- Em situações “do mundo real”:
 - Bases de documentos longos (notícias, páginas da Web, documentos jurídicos, etc.)
 - Opções:
 - Truncagem (baixo custo)
 - Quebra em subdocumentos e uso do escore máximo dentre os subdocumentos

