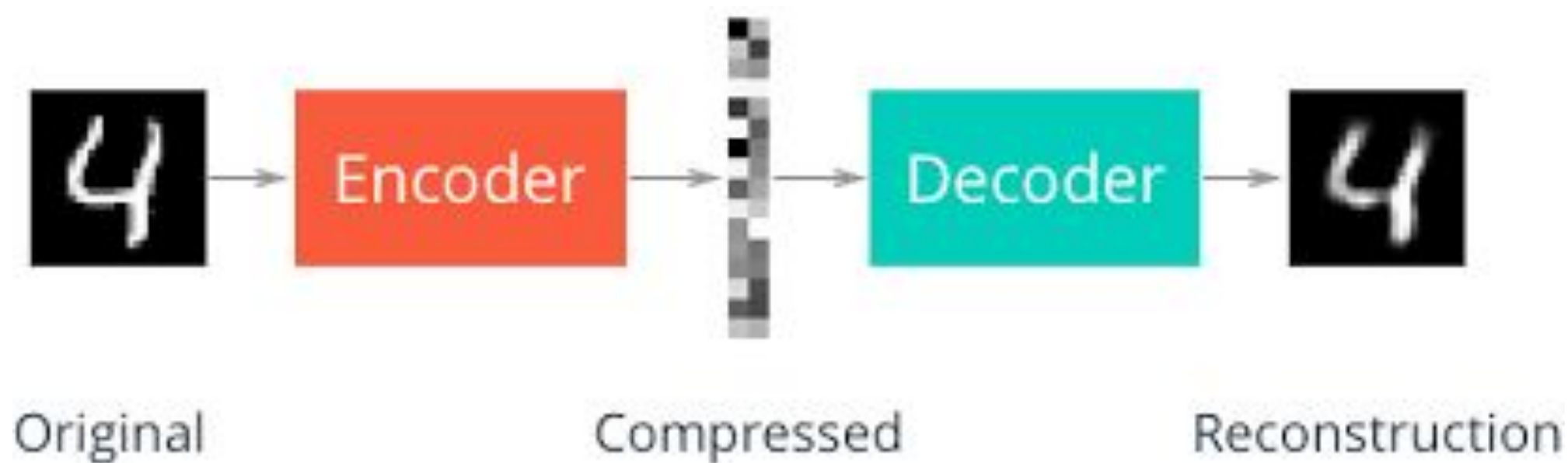
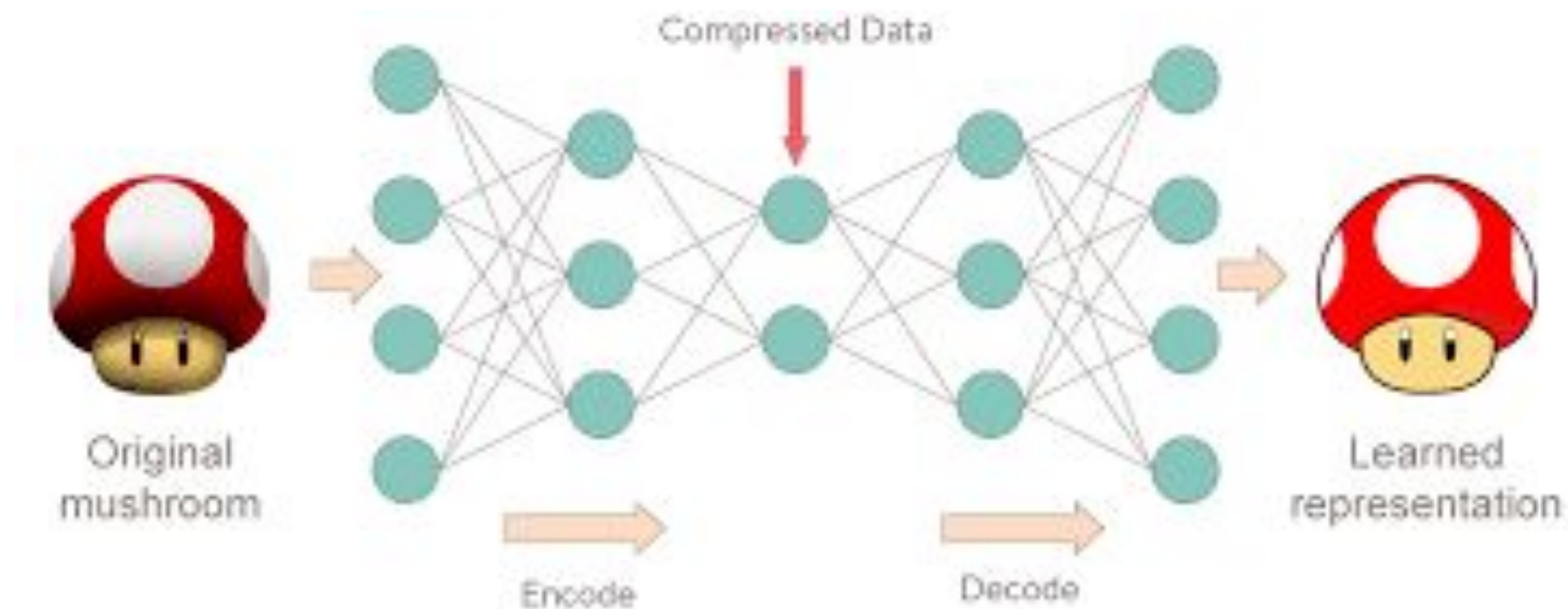


Análise de Mídias Sociais e Mineração de Texto

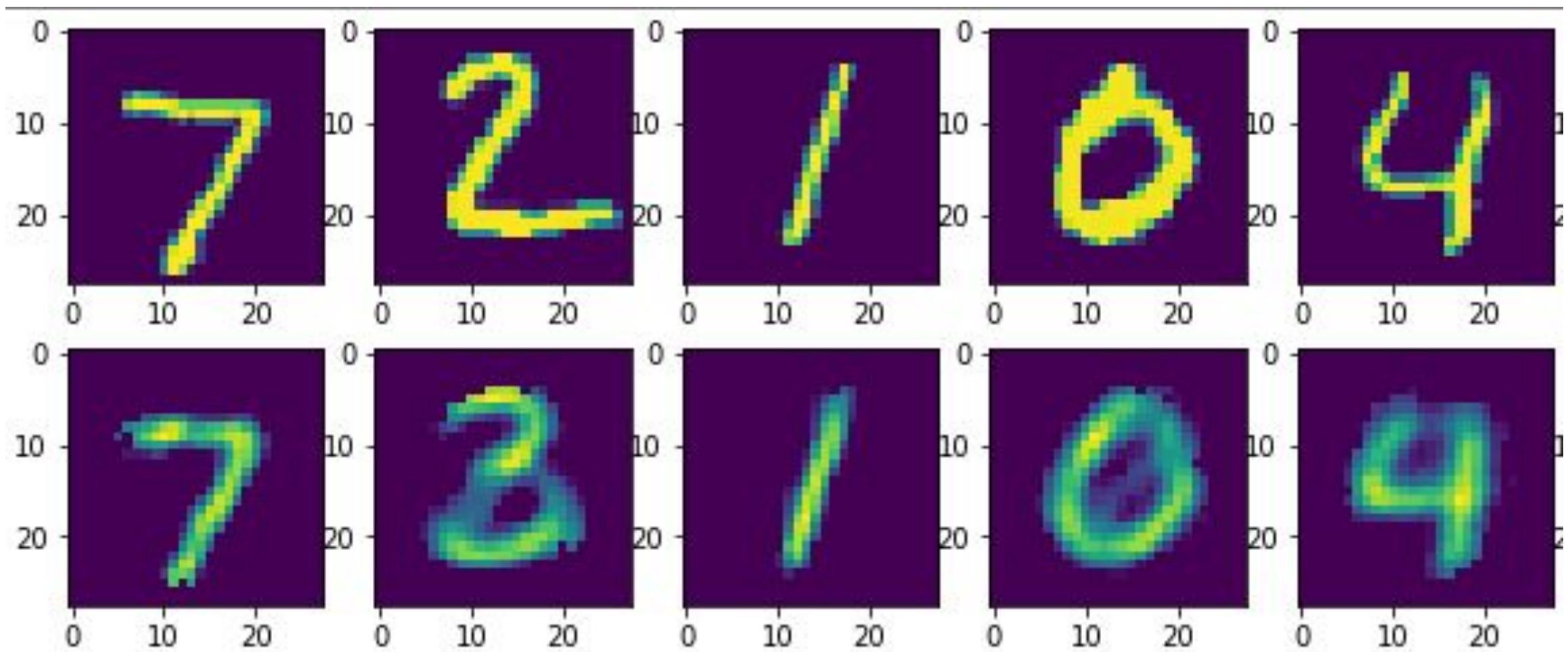
Word2Vec

Laura de Oliveira F. Moraes

Autoencoder

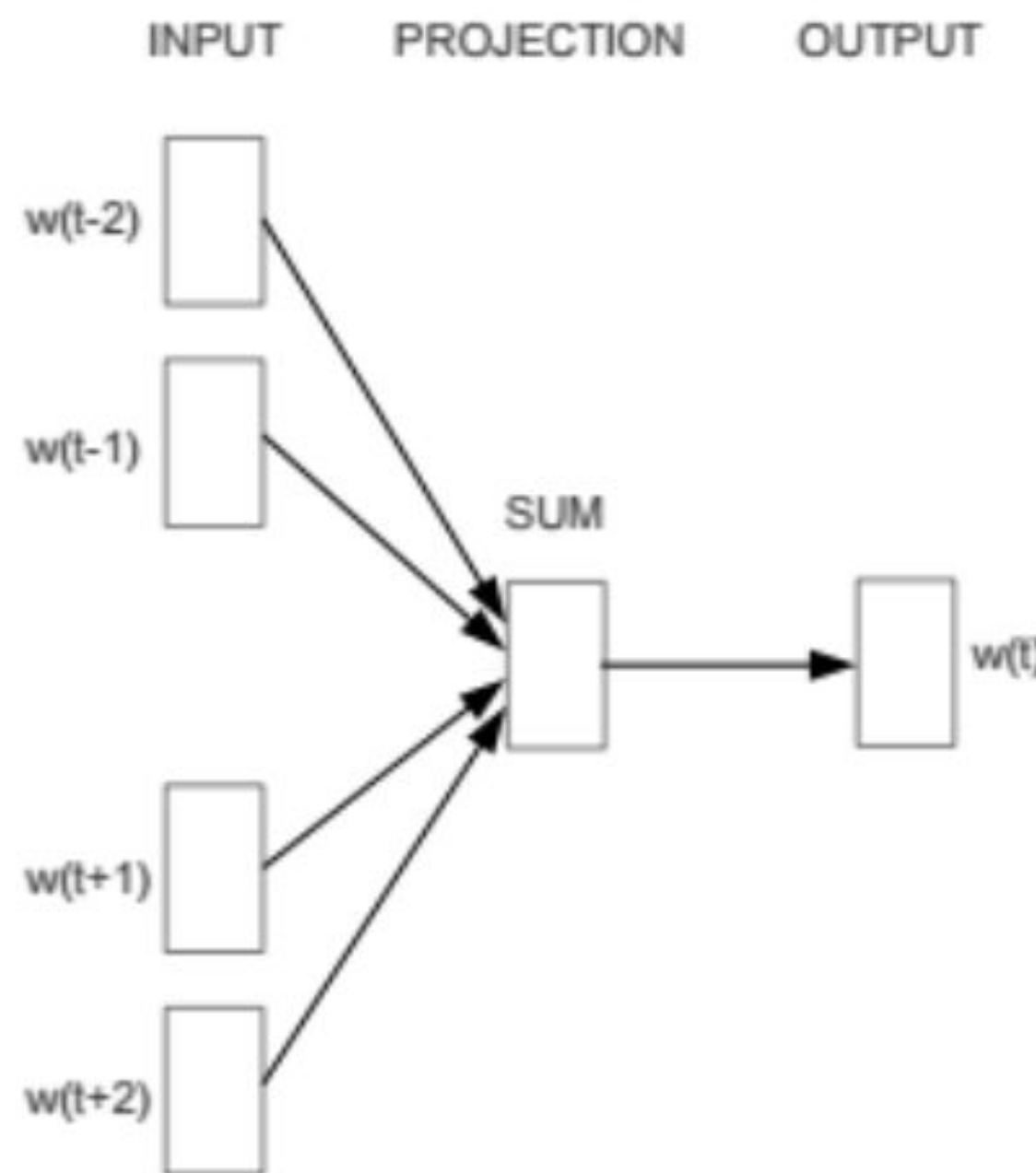


Resultado do autoencoder



- Treina mais rápido
- Acurácia boa para palavras mais frequentes

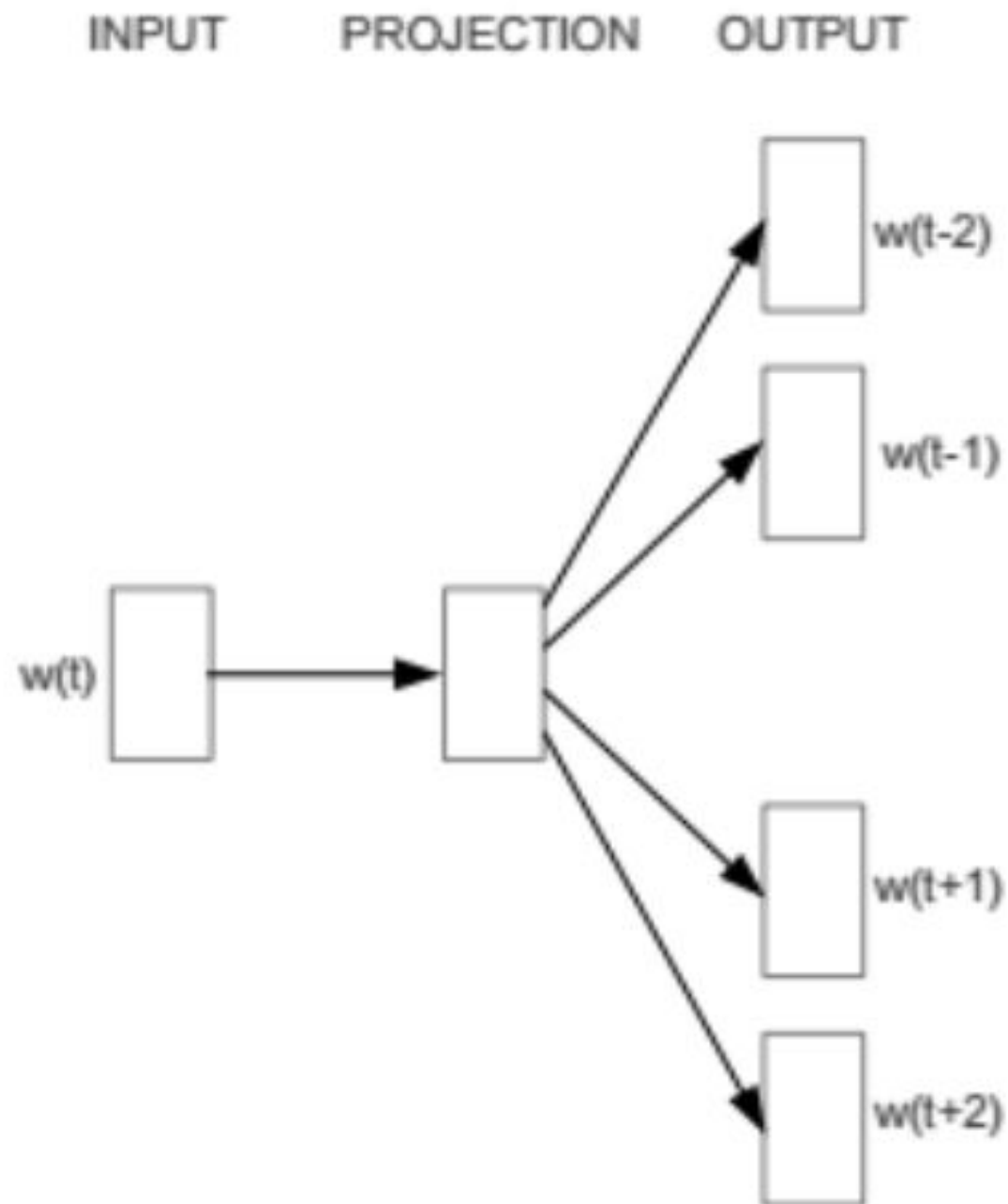
Given a set of (neighboring) words, **guess single words** that potentially occur along with this set of words.



CBOW

□ (Continuous Bag Of Words) □

Arquitetura da Rede - opção 2



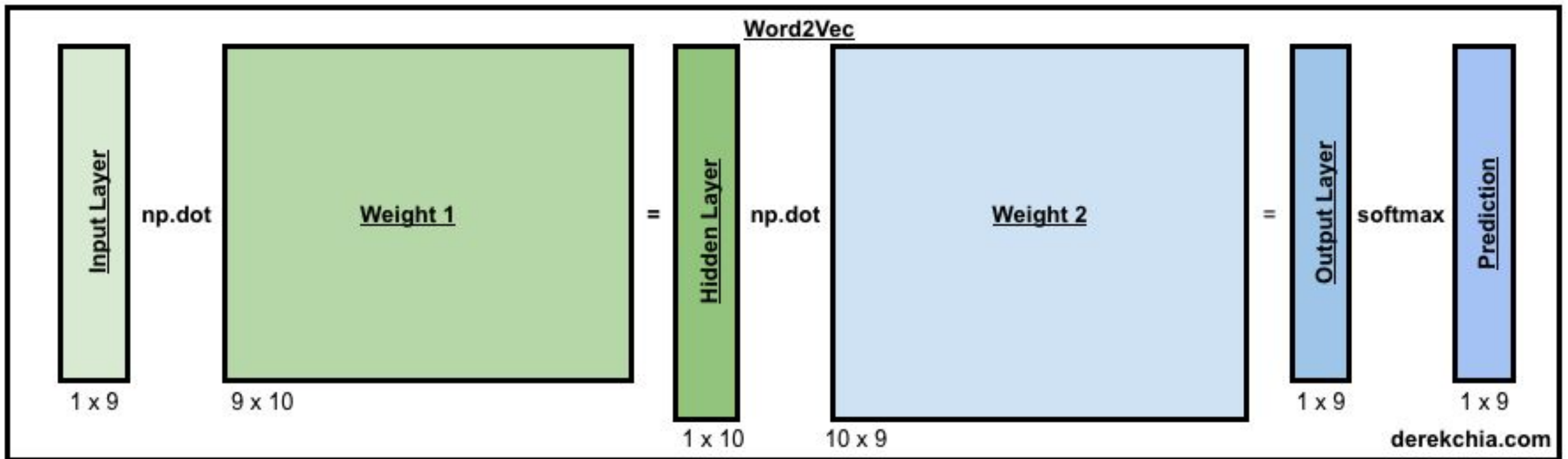
Skip-gram

- Trabalha bem com um número menor de dados
- Melhor na identificação de palavras mais raras

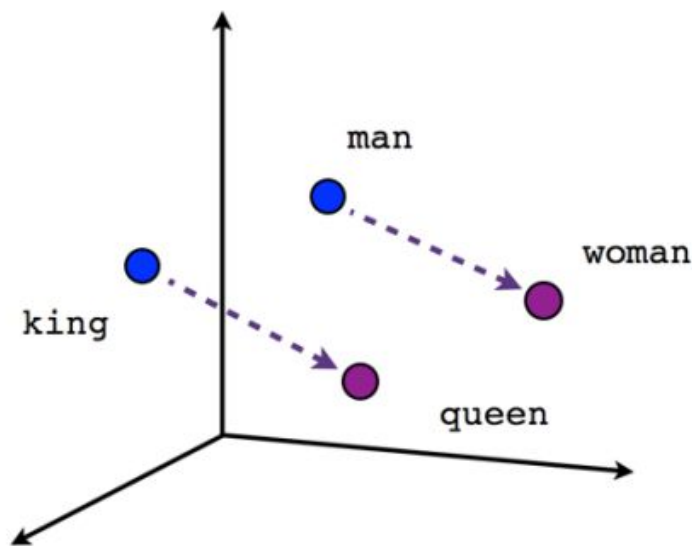
Guess potential neighboring words based on the single word being analyzed.

https://docs.google.com/spreadsheets/d/1mgf82Ue7MmQixMm2ZqnT1oWUucj6pEcd2wDs_JgHmco/edit#gid=0

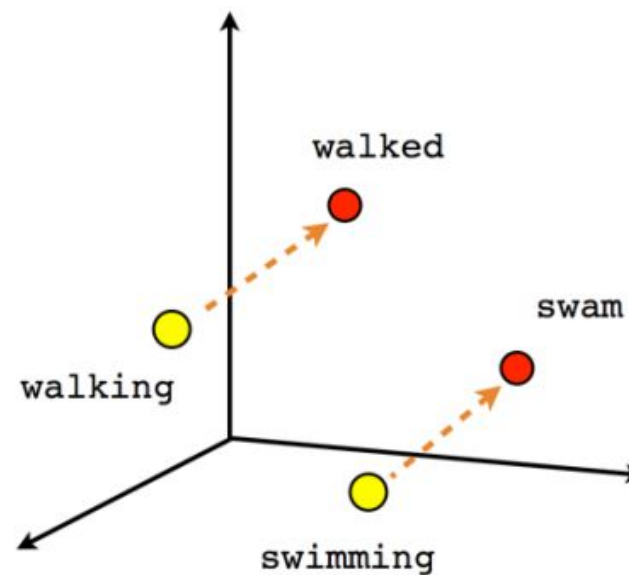
Estrutura da Rede Neural



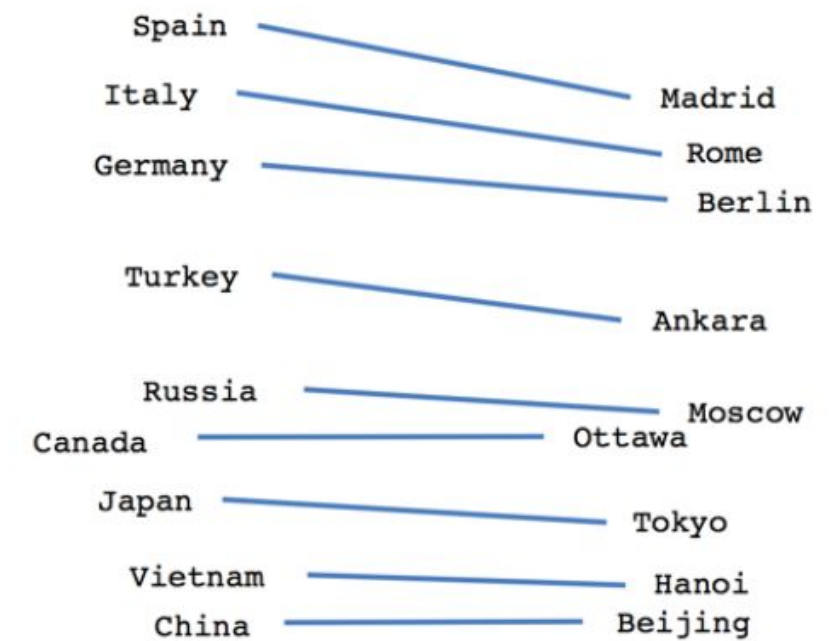
Espaço comprimido/codificado



Male-Female



Verb tense



Country-Capital

Czech + currency	Vietnam + capital	German + airlines	Russian + river	French + actress
koruna	Hanoi	airline Lufthansa	Moscow	Juliette Binoche
Check crown	Ho Chi Minh City	carrier Lufthansa	Volga River	Vanessa Paradis
Polish zolty	Viet Nam	flag carrier Lufthansa	upriver	Charlotte Gainsbourg
CTK	Vietnamese	Lufthansa	Russia	Cecile De