

# 深度学习 训练营

一步一个脚印，掌握深度学习

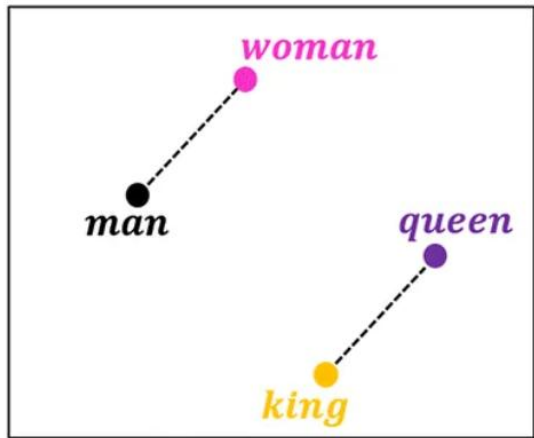


# 第8周, word2vec词嵌入算法

## 第8周, 课程核心内容:

- 8.1-文本向量化和OneHot编码
- 8.2-实验, OneHot编码实现文本向量化
- 8.3-词嵌入, Word Embedding算法
- 8.4-相似度计算和类比推理
- 8.5-实验, 基于gensim使用词向量模型
- 8.6-实验, 基于gensim训练中文词向量
- 8.7-连续词袋模型CBOW
- 8.8-实验, pytorch实现CBOW
- 8.9-跳字模型skip-gram
- 8.10-实验, pytorch实现skip-gram

<i>man</i> →	0.6	-0.2	0.8	0.9	-0.1	-0.9	-0.7
<i>woman</i> →	0.7	0.3	0.9	-0.7	0.1	-0.5	-0.4
<i>king</i> →	0.5	-0.4	0.7	0.8	0.9	-0.7	-0.6
<i>queen</i> →	0.8	-0.1	0.8	-0.9	0.8	-0.5	-0.9



# 预习指导问题

同学们在视频学习前，请尝试回答如下问题，并在视频学习后，将答案进行归纳与整理：

- 1.什么是文本向量化?在处理NLP问题时，为什么需要文本向量化?
- 2.OneHot编码的原理是什么?它有什么局限性?
- 3.什么是词嵌入(Word Embedding)，相比传统的文本向量化方法，它有什么优势?
- 4.简述word2vec的工作原理，它有哪些具体的实现方法?
- 5.如何使用词向量计算词语的相似度?
- 6.什么是词语的类比推理，为什么可以基于king, man, queen三个词推出woman?
- 7.CBOW模型和skip-gram模型，有哪些相同和不同的地方?
- 8.什么是CBOW模型?它是如何工作的，在模型中有哪些结构?
- 9.什么是skip-gram模型?它是如何工作的，在模型中有哪些结构?
- 10.除了word2vec算法，还有哪些算法可以训练出词向量?

# 课后编程练习

同学们在视频学习后，请尝试如下编程练习，并在直播课中，跟着老师一起完成全部编程作业。

- 1.请基于Gensim工具，使用词向量，计算词语的相似度，并进行类比推理。
- 2.请基于Gensim工具，训练中文词向量。
- 3.请基于Gensim工具，调研不同算法(word2vec、fasttext、glove)、不同数据训练出的词向量的效果。
- 4.如何使用pytorch，实现一个CBOW模型，并训练CBOW模型？
- 5.如何使用pytorch，实现一个skip-gram模型，并训练skip-gram模型？

最后将所有问题，都直接与小黑黑老师讨论清楚吧！

微信号:xhh890921

