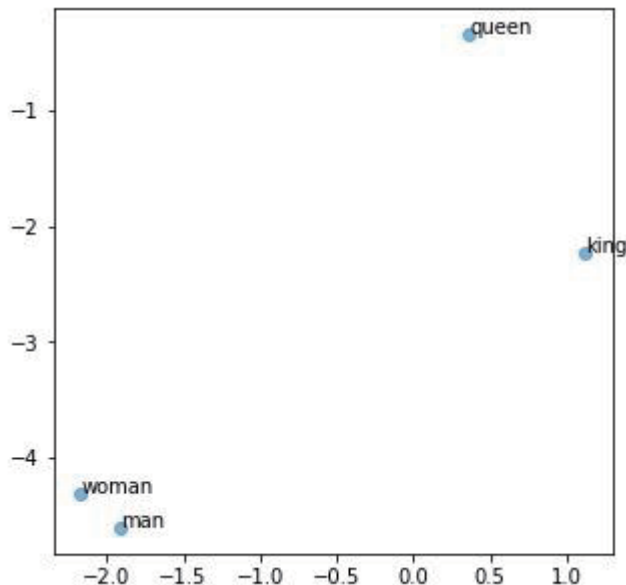


6장 연습문제

- 유추문제를 통해 카운트 기반 모델과 예측 기반 모델을 비교하려 한다.
 - 공정한 비교를 위해 2장 연습문제 3-(i) 코드를 수정하여 윈도우 크기 5로 PPMI행렬 구한 후 pickle파일에 저장하시오.
 - eval.py를 수정하여 카운트 기반 모델로 유추문제를 푸시오. 공정한 비교를 위해 SVD를 적용해 벡터표현 차원을 100으로 줄이고 시작하시오. (count_method_big.py 참고)
- train.py를 수정하여 PTB 데이터셋으로 Skip-Gram 모델을 훈련시킨 후 파라미터들을 pickle파일에 저장하시오.
 - eval.py를 수정하여 위에서 학습시킨 Skip-Gram 모델로 유추문제를 푸시오. CBOW 모델에 못미치는 이유를 설명하시오.
- 통계학 시간에 배운 PCA가 얼마나 효율적인 차원축소 방법인지 유추문제로 확인하려 한다.
 - 인터넷에서 사이킷런의 PCA 클래스 사용법을 검색하시오.
 - eval.py를 수정하여 PCA로 단어의 벡터 표현 차원을 20만큼 줄 후 유추문제를 푸시오.
 - eval.py를 수정하여 단어의 벡터 표현을 단순히 앞에서 20차원 만큼 자른 후 유추문제를 푸시오.
- PCA를 이용해 CBOW 모델의 유추문제에 대한 이해를 기하학적으로 시각화하려 한다.
 - 네 단어 king, man, queen, woman의 벡터 표현에 PCA를 적용해서 차원을 2만큼 줄인 후 산포도로 시각화 하시오. (eval.py와 count_method_small.py 수정)



- 기하학적 의미를 해석하시오.
- 나머지 조합 take, took, go, went와 car, cars, child, children와 good, better, bad, worse에 대해서도 같은 작업을 하시오.