

# Word2Vec & t-SNE

# Word2Vec

*# 초기화 및 모델 학습*

```
from gensim.models import word2vec
```

*# 모델 학습*

```
model = word2vec.Word2Vec(sentences,  
                           workers=num_workers,  
                           size=num_features,  
                           min_count=min_word_count,  
                           window=context,  
                           sample=downsampling)
```

# Word2Vec의 parameter

*# 초기화 및 모델 학습*

```
from gensim.models import word2vec
```

*# 모델 학습*

```
model = word2vec.Word2Vec(sentences,  
                           workers=num_workers,  
                           size=num_features,  
                           min_count=min_word_count,  
                           window=context,  
                           sample=downsampling)
```

# t-SNE

```
from sklearn.manifold import TSNE
```

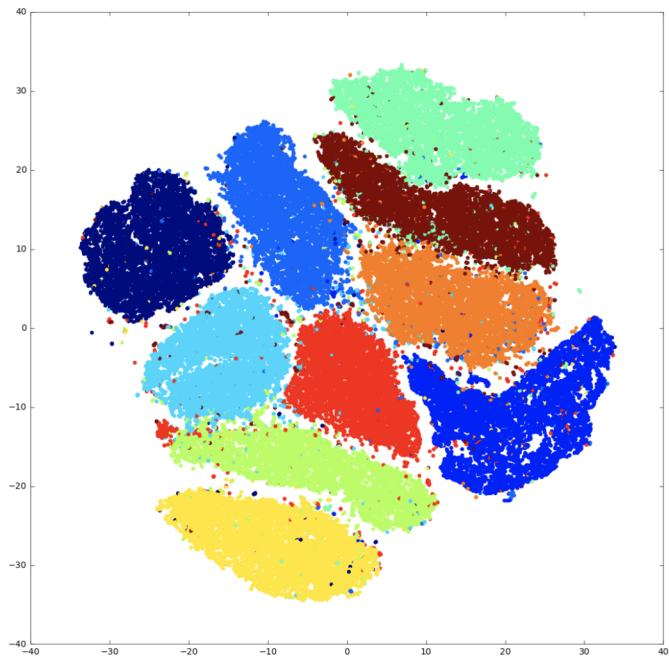
```
TSNE(n_components=2, perplexity=30.0, early_exaggeration=12.0,  
     learning_rate=200.0, n_iter=1000, n_iter_without_progress=300,  
     min_grad_norm=1e-07, metric='euclidean', init='random', verbose=0,  
     random_state=None, method='barnes_hut', angle=0.5)
```

parameter	note
n_components	임베딩 공간의 차원
perplexity	$\sigma_i$ 의 기준, 학습에 영향을 주는 점들의 개수를 조절
metric	원 공간에서의 두 점간의 거리 척도
method	원 공간 데이터의 인덱싱 방식

# t-SNE

## Multicore t-SNE build failing

This is a multicore modification of [Barnes-Hut t-SNE](#) by L. Van der Maaten with python and Torch CFFI-based wrappers. This code also works **faster than sklearn.TSNE** on 1 core.



# t-SNE의 parameter

```
from sklearn.manifold import TSNE
```

```
TSNE(n_components=2, perplexity=30.0, early_exaggeration=12.0,  
     learning_rate=200.0, n_iter=1000, n_iter_without_progress=300,  
     min_grad_norm=1e-07, metric='euclidean', init='random', verbose=0,  
     random_state=None, method='barnes_hut', angle=0.5)
```

parameter	note
n_components	임베딩 공간의 차원
perplexity	$\sigma_i$ 의 기준, 학습에 영향을 주는 점들의 개수를 조절
metric	원 공간에서의 두 점간의 거리 척도
method	원 공간 데이터의 인덱싱 방식

# Word2Vec 관련 paper review

- Word2vec 모델로 학습된 단어 벡터의 의미 관계 분석
- Word2Vec을 활용한 제품군별 시장규모 추정 방법에 관한 연구
- Word2vec 임베딩을 이용한 책 추천 시스템
- Word2Vec과 LDA기법을 이용한 게임 리뷰 분석
- Word2Vec 기반의 의미적 유사도를 고려한 웹사이트 키워드 선택 기법
- Word2Vec과 심층 신경망을 활용한 영화추천 시스템의 정확도 개선
- Word2vec과 앙상블 분류기를 사용한 효율적 한국어 감성 분류 방안
- Word2Vec 과 Convolution 신경망 기반 문서 주제 분류에 관한 연구
- Word2vec을 활용한 RNN기반의 문서 분류에 관한 연구
- Textrank와 word2vec을 활용한 국민청원 게시판 중복주제 탐지

Thank you