Lecture 1- Intro & Word Vectors

01. Human Language and Word meaning

- what is **meaning** ? (Webster dictionary)
  - the Idea that is represented by a word, phrase, etc.

- Commonest lingustic way of thinking meaning:
  - signifier (symbol) ↔ signified (idea or thing)

    = denotational semantics

WordNet

유의어(synonym)와 상의어(hypernyms)를 가지고 있는 백과사전

- denotational semantic (표시 의미론)
  - 단어가 표시하고 있는 것이 곧 의미

- Problems of WordNet
  - 단어의 뉘앙스를 놓쳐버림
  - 단어의 새로운 의미를 놓쳐버림
  - 주관적임
  - 단어의 유사도를 파악하기 어려움

One-Hot Vectors

문장에서 단어가 나타나는 위치에 따라 벡터 값 1, 이외에는 벡터 값 0을 부여

- Problems of One-Hot Vector
  - 무한한 단어를 표현하기에 한계가 존재
  - 단어의 유사도를 표현하기 어려움

02. Word2Vec algorithm introduction

Distributional semantics

- 분산 의미론 ; 문맥을 통해 단어의 의미를 파악. (A word’s meaning is given by the words that **frequently appear close-by.)**
- word vector = word embeddings = word representations

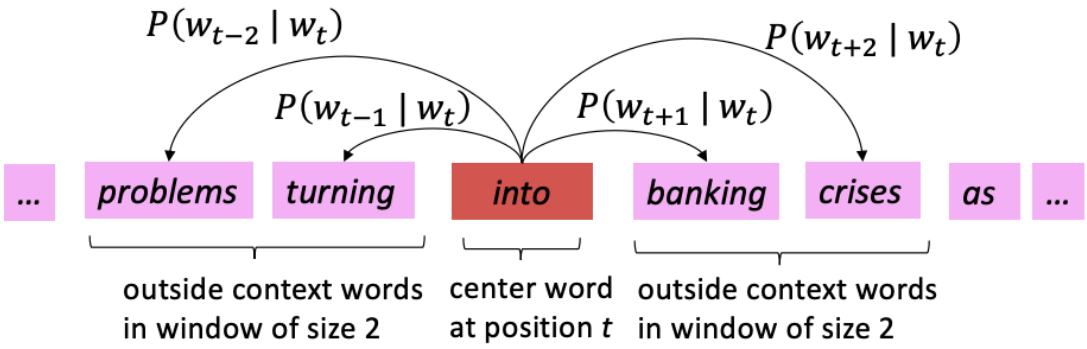
Word2Vec

- large corpus (말뭉치)
- corpus의 모든 단어에 random vector 부여
- 중심 단어 c, 주변 단어 o
- c가 나올 때 o가 나올 확률 계산

- 위 확률이 maximize되도록 단어 vector 조정

<<Overview>>

Example windows and process for computing $P(w_{t+j} \mid w_t)$



03. Word2Vec objective function and gradients

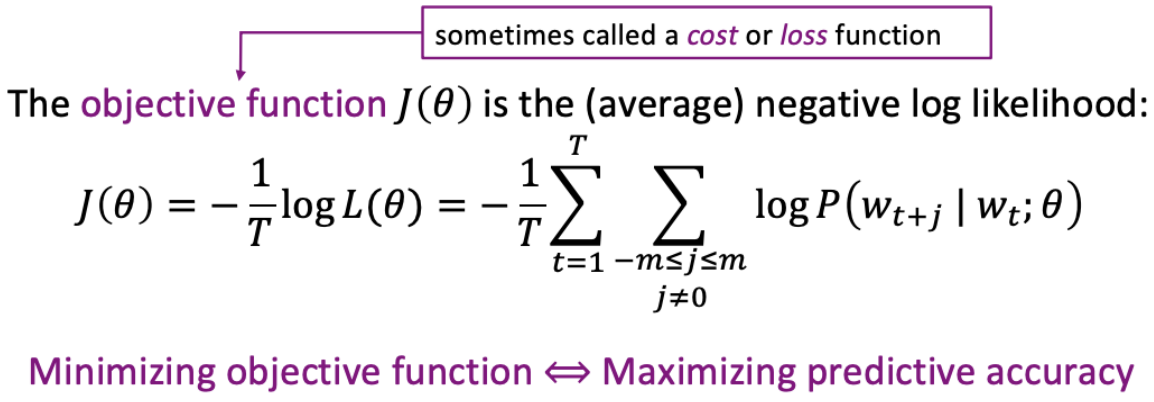
- Likelihood (우도)
  - 모든 단어(T)에 대하여 fixed size window 만큼 주변 단어 확률 곱함
  - 주변 단어 발생 확률 예측

$$\text{Likelihood} = L(\theta) = \prod_{t=1}^{T} \prod_{\substack{-m \le j \le m \\ j \neq 0}} P(w_{t+j} \mid w_t; \theta)$$

θ is all variables to be optimized

- Objective function
  - Objective function 최소화 하는 방향

    == 다른 단어들의 문맥에서 중심 단어를 잘 예측

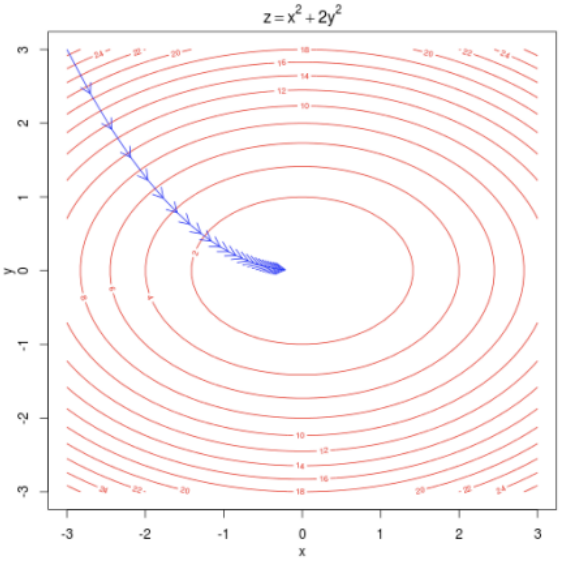


04. Optimization basics

Optimization

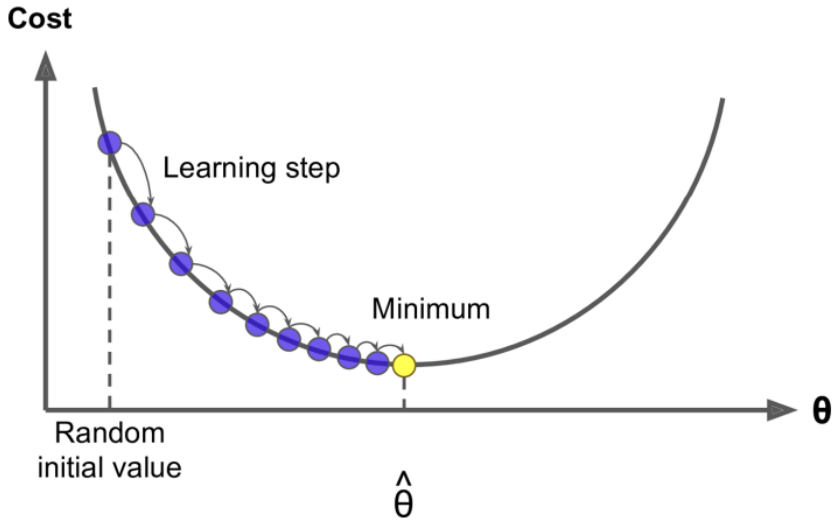
optimize를 통해 objective function을 최소화하는 파라미터 θ 찾는 과정

$$\theta = \begin{bmatrix} v_{aardvark} \\ v_a \\ \vdots \\ v_{zebra} \\ u_{aardvark} \\ u_a \\ \vdots \\ u_{zebra} \end{bmatrix} \in \mathbb{R}^{2dV}$$



- θ : represents all the model parameters, in one long vector
- we gradually adjust parameters to minimize a loss
- optimize these parameters by walking down the gradient
- compute all vector gradients!

- Gradient Descent



Note: Our objectives may not be convex like this ☹

But life turns out to be okay ☺

  - cost function j(θ) ; we want to minimize this

- for current value of θ, calculate gradient of j(θ) , then take small step in direction of negative gradient. Repeat.

05. Looking at word vectors

Gensim

이미 학습되어 배포된 word vector package이며, 벡터로 단어를 표현

입력한 단어와 비슷한 의미를 가지는 단어를 출력