

# Assignment 2 Report

Sai Sriram Yannamani - 2019101029

## NEGATIVE SAMPLING

Negative sampling is a technique that is used to simplify the computational aspect of Word2Vec training. In the vanilla model, the training samples are context vectors which are used to predict the center word as the output. This means that the output is represented across the vocabulary space, with the desired result such that the largest probability is given to the correct center word.

However, if the vocabulary size is large, as is the case for a huge corpus, then the output layer can become expensive to calculate as softmax involves exponentiation of a number for each of the vocabulary element. In Negative sampling, the task is modified into a binary classification, with the new training input being a tuple of context vector and a center word, and the output label being 1 indicating that the center word belongs to that context, and vice versa.

When creating a dataset from the corpus, only positive samples can be created. Thus, negative ones are artificially generated by sampling the unigram probabilities raised to the power  $3/4$ . Once this is done, all the training samples are fed into the network like a usual classification task, with the output layer now simply using a sigmoid function which is much better than softmax computationally.

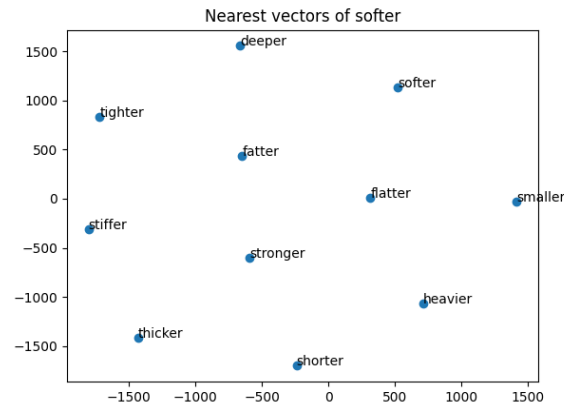
## ANALYSIS

*Nearest words for 'camera' from pre-trained model (glove-wikipedia-gigaword-100):*

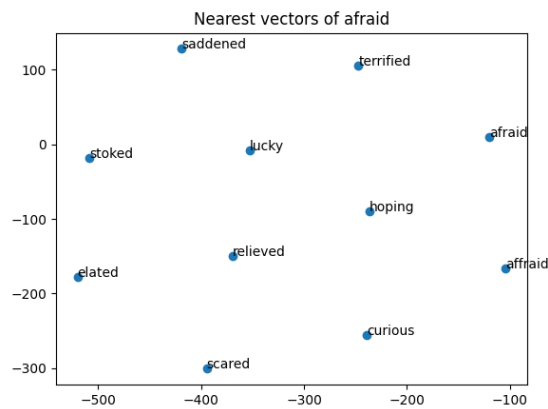
cameras, screen, video, screens, microphone, digital, images, device, window, sensor

*Frequency Based Model from Co-occurrence Matrix and SVD:*

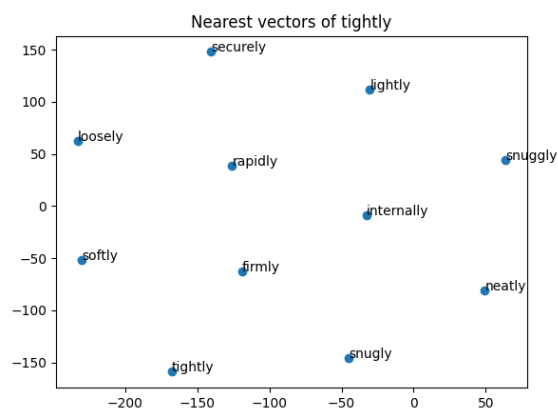
1. **softer** - tighter, smaller, flatter, deeper, stiffer, stronger, thicker, fatter, heavier, shorter



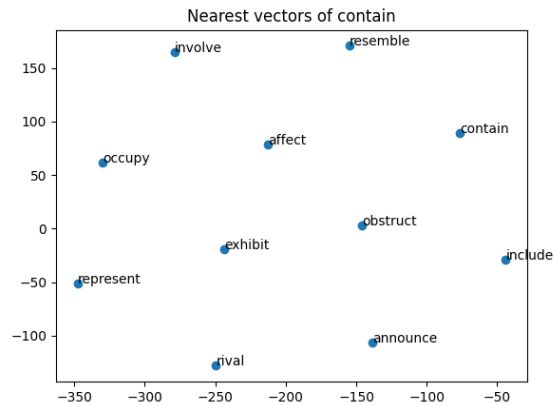
2. **afraid** – affraid, scared, relieved, saddened, hoping, terrified, stoked, lucky, curious, elated



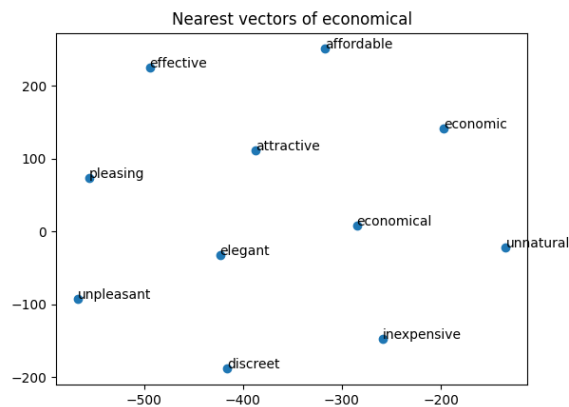
3. **tightly** – loosely, securely, snugly, firmly, snugly, lightly, neatly, rapidly, softly, internally



4. **contain** – include, involve, exhibit, occupy, resemble, announce, affect, represent, obstruct, rival



5. **economical** – affordable, effective, inexpensive, attractive, elegant, unpleasant, discreet, pleasing, unnatural, economic



Vectors output for **camera**: lense, cam, machine, headset, scope, radio, psu, router, detector, mouse

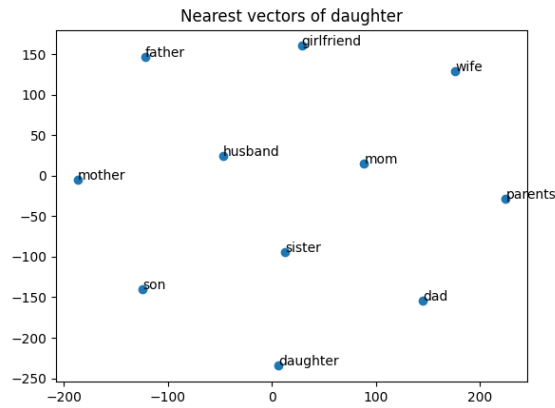
Other notable results:

1. **gripping** – complaining, raves, facts, ravings, raving, whining, myths, nevermind, remarks, comments
2. **capacidades** – almacenamiento, efectuar, espacios, dividendos, descuento, muchisima, convinaciones, copias, filas, compartimientos
3. **lossless** – avi, jpg, 1080i, mp4, wmv, pal, flac, ntsc, wav, dance
4. **surprisingly** – suprisingly, shockingly, remarkably, pretty, very, vey, amazingly, unbelievably, impressively, stunningly
5. **occasions** – occassion, demand, occasion, weekends, friday, vacation, spilled, relied, trips, flights
6. **disliked** – dislike, misunderstood, liked, underestimated, questioned, detest, hated, appreciate, blason, scoured

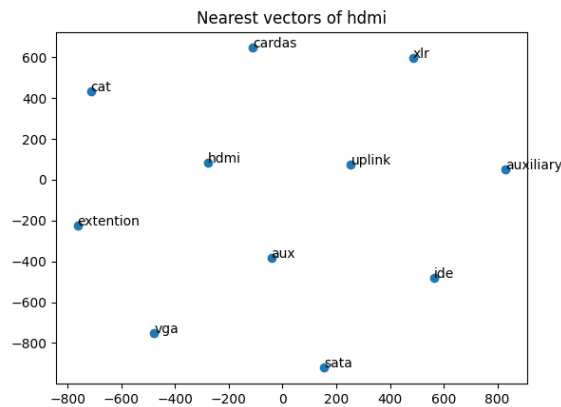
7. **tolerably** – exceptionally, exceedingly, amazingly, phenomenally, fantastically, works, cleverly, worked, reasonably, decently
8. **brilliant** – beautiful, gorgeous, pure, magnificent, weak, silent, sloppy, horrid, lousy, lovely
9. **elegance** – richness, crispness, smoothness, simplicity, openness, versatility, sturdiness, subtlety, usability, sleekness
10. **challenging** – troublesome, tricky, painful, dangerous, awkward, complicated, disconcerting, tough, risky, confusing
11. **downgraded** – upgraded, moved, transfered, copied, uploaded, switched, transferred, read, migrated, donated
12. **kentucky** – vancouver, peru, orlando, london, yosemite, greece, spain, asia, diego, hawaii
13. **mexico** – england, france, paris, antonio, pennsylvania, korea, rome, jose, philadelphia, florida
14. **nylon** – gray, canvas, metal, metallic, rubberized, steel, motorized, stainless, leather, rubbery
15. **customers** – consumers, buyers, employees, shoppers, cassettes, calls, solutions, converters, destinations, answers
16. **daughter** – son, dad, girlfriend, grandson, sister, nephew, niece, granddaughter, fiance, mom
17. **hdmi** – ethernet, hmdi, toslink, coaxial, cat6, optic, coax, svga, extension, cat5
18. **manfrotto** – cobra, celestron, plantronics, linksys, razer, wacom, fiio, d, rode, netgear
19. **toyota** – ford, subaru, honda, chevy, dodge, prius, corolla, pontiac, lexus, vw
20. **impedance** – transmission, resolution, aluminium, suppression, wattage, frequency, poi, adjustment, speed, amperage
21. **disassembled** – unplugged, returned, trashed, uninstalled, deleted, tracked, finally, hated, ignored, unhooked
22. **advanced** – intricate, intense, energy, obscure, arcane, flexibility, aesthetic, ergonomic, esoteric, options
23. **enables** – allows, forces, encourages, appeals, enable, allow, allows, ensure, enabling
24. **flimsy** – stiff, bulky, tacky, sloppy, clunky, rigid, plasticky, heavy, fragile, jerky
25. **rf970r** – rmvl1000, v40xi, mapsever, headsethowever, de925, a816, ip55, 464, pc120, ns92v

*Frequency Based Model from Continuous Bag of Words and Negative Sampling:*

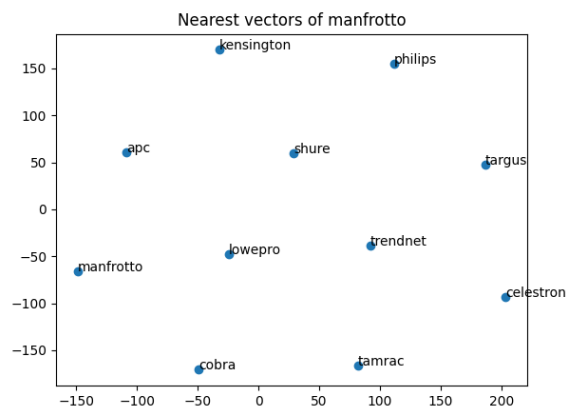
1. **daughter** – son, girlfriend, mother, husband, father, sister, mom, wife, dad, parents



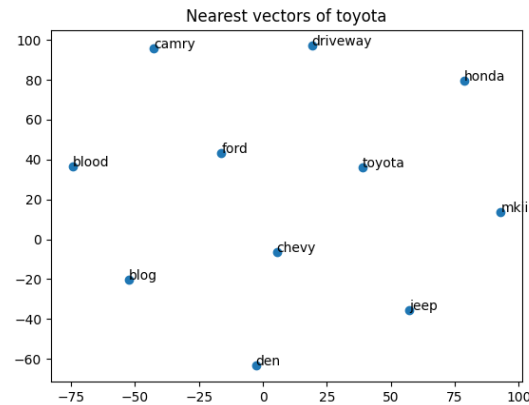
2. **hdmi** – auxiliary, aux, xlr, cat, cardas, uplink, vga, sata, ide, extention



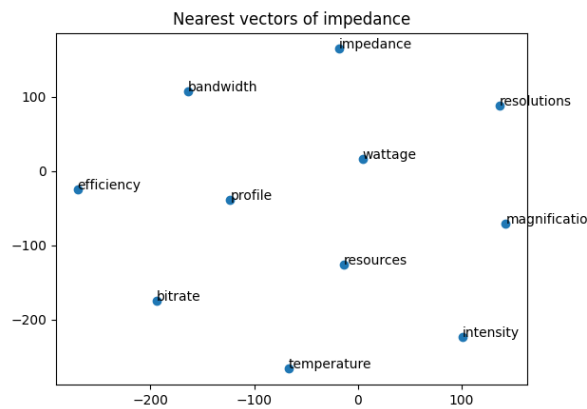
3. **manfrotto** – targus, philips, celestron, trendnet, tamrac, kensington, lowepro, apc, shure, cobra



4. **toyota** – mkii, chevy, ford, jeep, camry, honda, den, blood, blog, driveway



5. **impedance** – bitrate, wattage, intensity, resolutions, efficiency, bandwidth, magnifications, profile, temperature, resources



Vectors output for **camera**: cameras, laptop, gadget, cam, domke, machine, notto, camcorder, lightspheres, messenger

Other notable results:

1. **softer** – brighter, darker, sturdier, thinner, warmer, quieter, bulkier, louder, taller, sharper
2. **tightly** – snugly, snug, loosely, firmly, neatly, snuggly, rapidly, securely, comfortable, badly
3. **contain** – render, rival, reveal, generate, indicate, introduce, represent, establish, occupy, retain
4. **afraid** – advised, aware, suspicious, shocked, forced, unsure, wary, lucky, dissapointed, ashamed
5. **disassembled** – inspected, trashed, uninstalled, ignored, rented, grabbed, hated, bough, outgrew, studied
6. **augment** – supplement, maximize, adapt, refers, relocate, fasten, bypass, compliment, affix, according

7. **advanced** – enjoyable, appropriate, sophisticated, specialized, inopportune, obscure, complex, complicated, costly, forgiving
8. **wasting** – waste, saving, wasted, refunded, serving, spending, consuming, delivered, dying, save
9. **enables** – forces, tells, forcing, permits, owe, appeals, allows, enabling, encourages, bothers
10. **flimsy** – fragile, fuzzy, stiff, delicate, confusing, substantial, complicated, intuitive, cumbersome, sturdy
11. **nylon** – pec, velvet, bean, neoprene, canvas, cotton, velco, domke, nest, vinyl
12. **iraq** – chicago, fairness, arizona, florida, defense, retrospect, hawaii, nyc, dvhs, miami
13. **challenging** – troublesome, painful, satisfactory, tempting, inconvenient, confusing, problematic, uncomfortable, frustrating, irritating
14. **brilliant** – remarkable, lovely, magnificent, fabulous, gorgeous, stunning, superb, terrific, terrible, phenomenal
15. **surprisingly** – exceptionally, incredibly, extremely, very, fairly, stunningly, pretty, remarkably, amazingly, equally