Text Classification

CS 685, Fall 2025

Advanced Natural Language Processing

https://people.cs.umass.edu/~brenocon/cs685_f25/

Brendan O'Connor

College of Information and Computer Sciences

University of Massachusetts Amherst

- Text classification
- Supervised learning for text classif.
  - BOW features
  - Word embeddings features

text classification

- input: some text **x** (e.g., sentence, document)
- output: a label **y** (from a finite label set)
- goal: learn a mapping function *f* from **x** to **y**

fyi: basically every NLP problem reduces to learning a mapping function with various definitions of **x** and **y**!

problem	**x**	**y**
sentiment analysis	text from reviews (e.g., IMDB)	{positive, negative}
topic identification	documents	{sports, news, health, …}
author identification	books	{Tolkien, Shakespeare, …}
spam identification	emails	{spam, not spam}

… many more!

input **x**:

From European Union <info@eu.org>
Subject
Reply to [redacted]

Please confirm to us that you are the owner of this very email address with your copy of identity card as proof.

YOU EMAIL ID HAS WON $10,000,000.00 ON THE ONGOING EUROPEAN UNION COMPENSATION FOR SCAM VICTIMS. CONTACT OUR EMAIL:
CONTACT US NOW VIA EMAIL: [redacted] NOW TO CLAIM YOUR COMPENSATION

label **y**: **spam** or **not spam**

we'd like to learn a mapping *f* such that
$f(\mathbf{x})$ = **spam**

Demo: Keyword count classifier

- Let's consider this task:
  sentiment classification of movie reviews
- Can *manually defined* keyword lists be a useful indicator of text sentiment?
  - For each category, define set of words
  - Predict a category if many of its words are used
- Let's try manually defined keywords!

https://docs.google.com/forms/d/e/1FAIpQLScpufac69IBvXOeZUoUsNB63EIXKN6BcwPZwoq6kkTBcFnNIg/viewform?usp=sharing&ouid=104321982622251425263

bag-of-words representation

i hate the actor i love the movie

word	count
i	2
hate	1
love	1
the	2
movie	1
actor	1

bag-of-words representation

i hate the actor i love the movie

word	count
i	2
hate	1
love	1
the	2
movie	1
actor	1

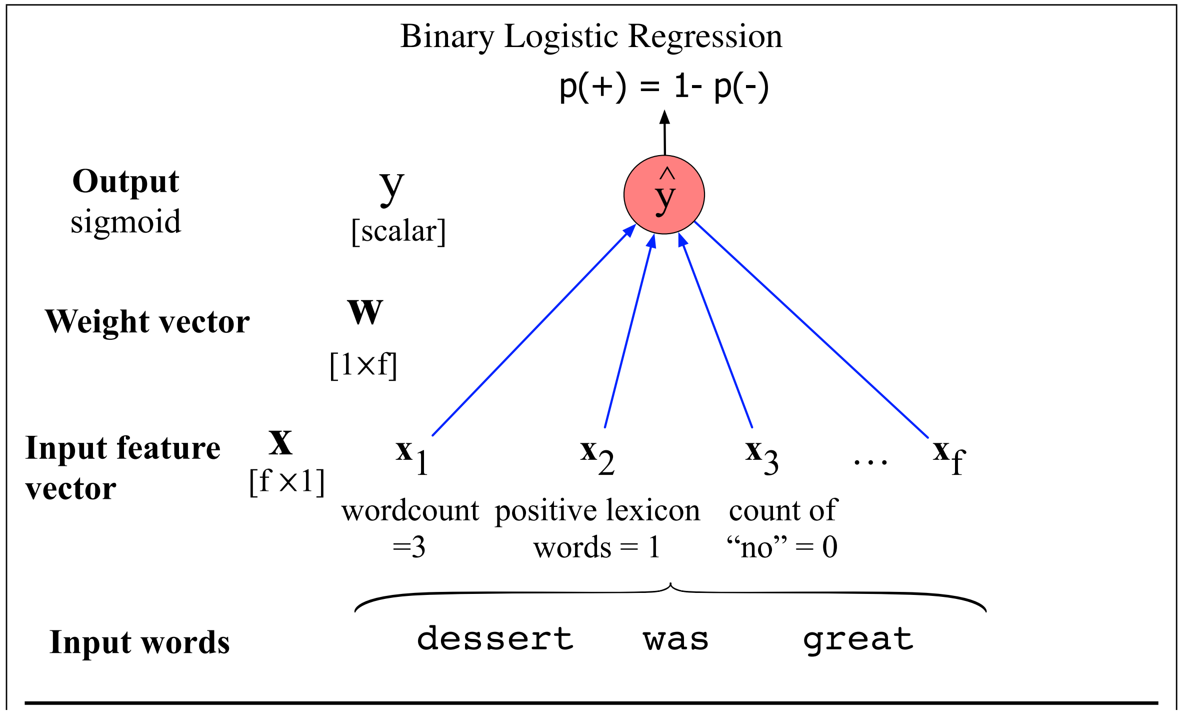
equivalent representation to:

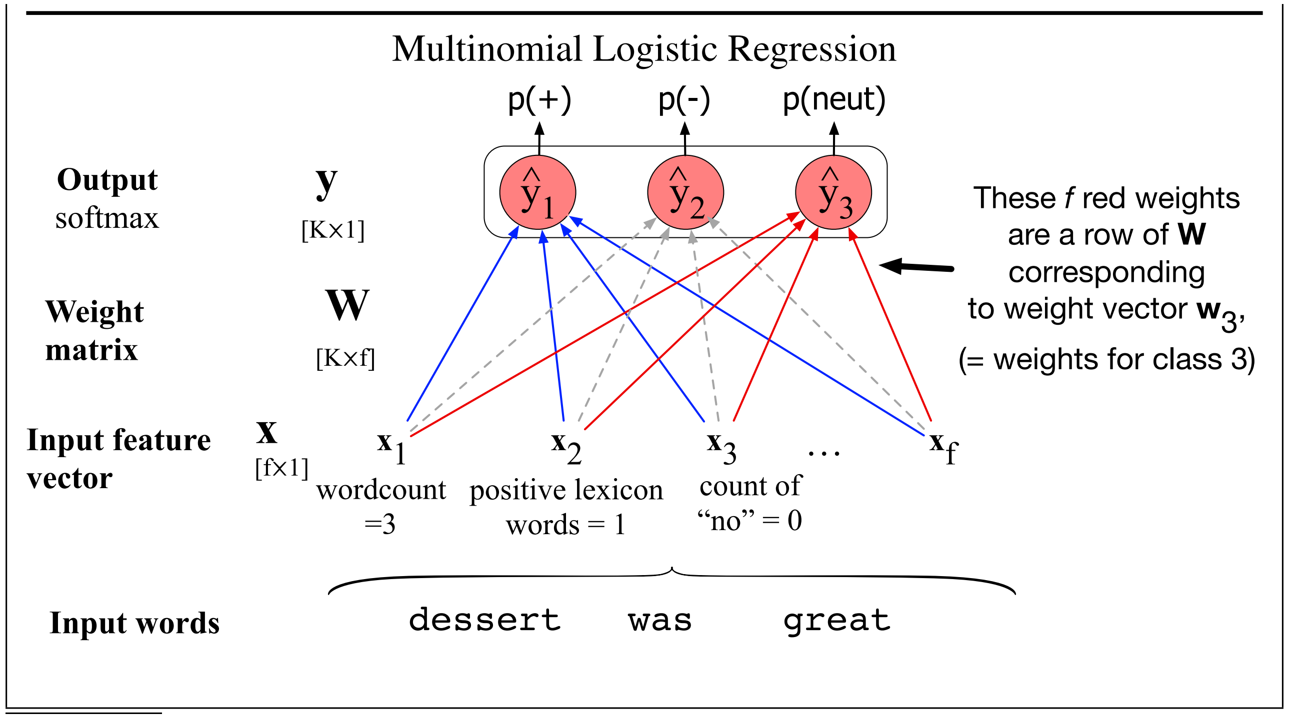
actor i i the the love movie hate

- What's weird about BOW: for many classification tasks it can actually perform well!
  - genre
  - author (e.g. indicator of style)
  - even.. sentiment?!

Multiclass Logistic Regression

- Each class has its own weight vector across features
  - BOW word-count features
  - Specialized or custom phrase features





Rank (r)	Word
1	the
2	and
3	to
4	a
5	she
6	it
7	of
8	said
9	i
10	alice
20	all
30	little
40	about
50	again
60	queen
70	don't
80	quite
90	just
100	voice
200	hand
300	turning
400	hall
500	kind

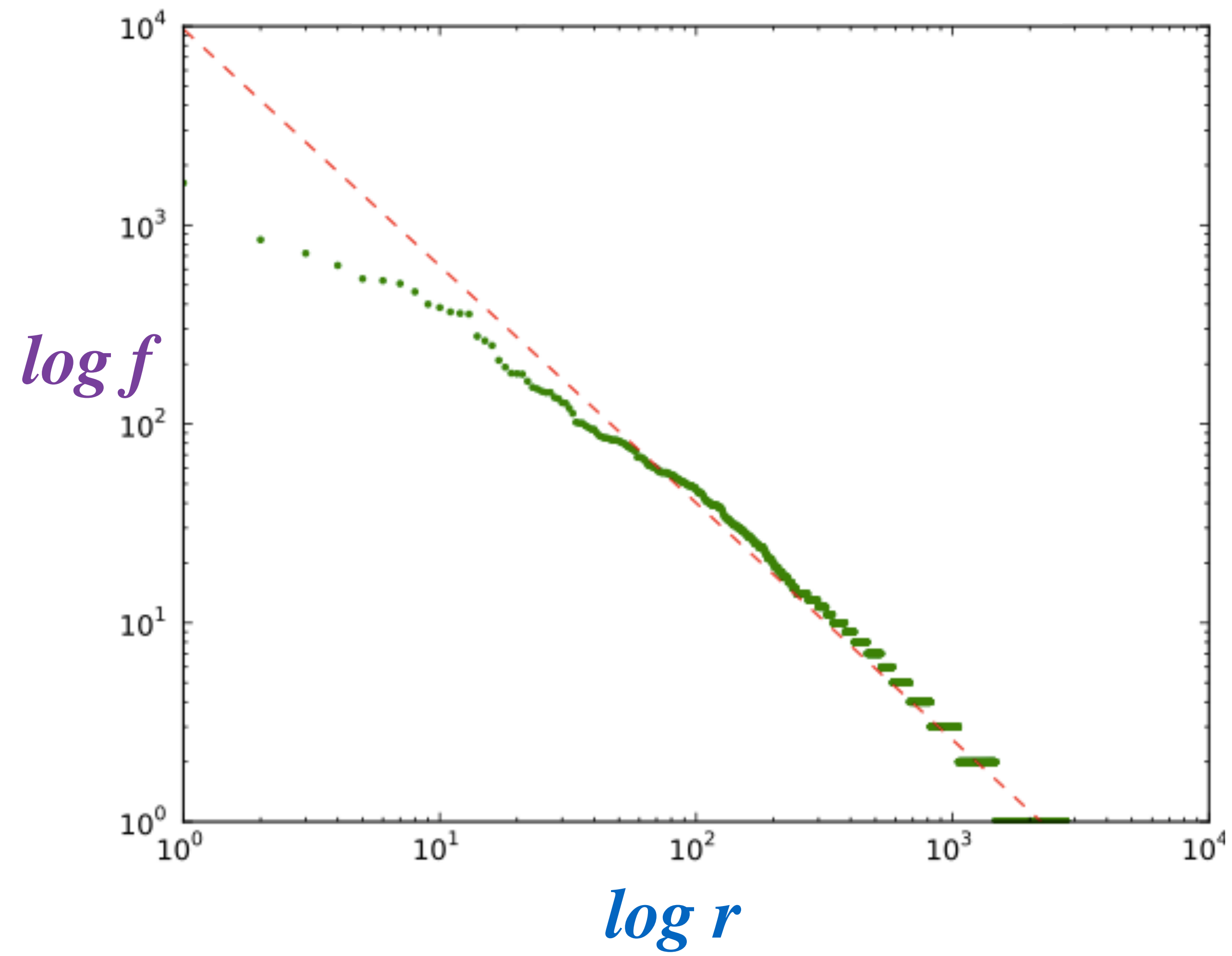
General problem: most words are rare

Zipf's Law

- When word types are ranked by frequency, then frequency (f) * rank (r) is roughly equal to some constant (k)

$$f \times r = k$$

Rank (r)	Word	Frequency (f)
1	the	1629
2	and	844
3	to	721
4	a	627
5	she	537
6	it	526
7	of	508
8	said	462
9	i	400
10	alice	385
20	all	179
30	little	128
40	about	94
50	again	82
60	queen	68
70	don't	60
80	quite	55
90	just	51
100	voice	47
200	hand	20
300	turning	12
400	hall	9
500	kind	7



$$f * r = k$$
$$\log f + \log r = \log k$$

Sup. learning with document embedding

- Instead of bag-of-words, can we derive a latent embedding of a document/sentence?
  - "Bag of embeddings" or "averaged word embeddings" representation
  - You can use it just like a BOW logistic regression - it's just a different type of feature vector
  - Pros/cons?
- Especially for shorter texts, BoE LR typically outperforms BOW LR.

See: Arora et al. 2017

- stopped here 9/11/25

Overfitting and generalization

- Overfitting: your model performs overly optimistically on training set, but generalizes poorly to other data (even from same distribution)
- To diagnose: separate training set vs. test set.
- How did we regularize Naive Bayes and language modeling?

- For logistic regression: L2 regularization for training

Regularization tradeoffs

- No regularization <--------------> Very strong regularization

Visualizing a classifier in feature space

"Bias term"

Feature vector $x = (1,\ \text{count “happy”},\ \text{count “hello”}, \ldots)$

Weights/parameters $\beta =$

50% prob where

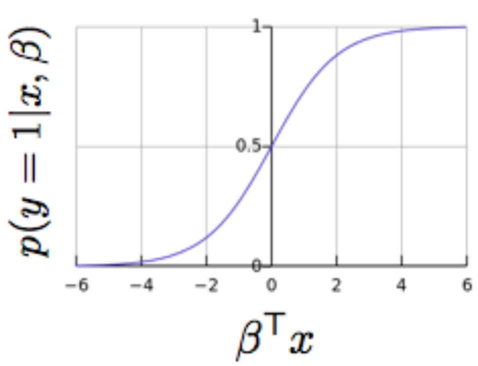
$\beta^\mathsf{T} x = 0$

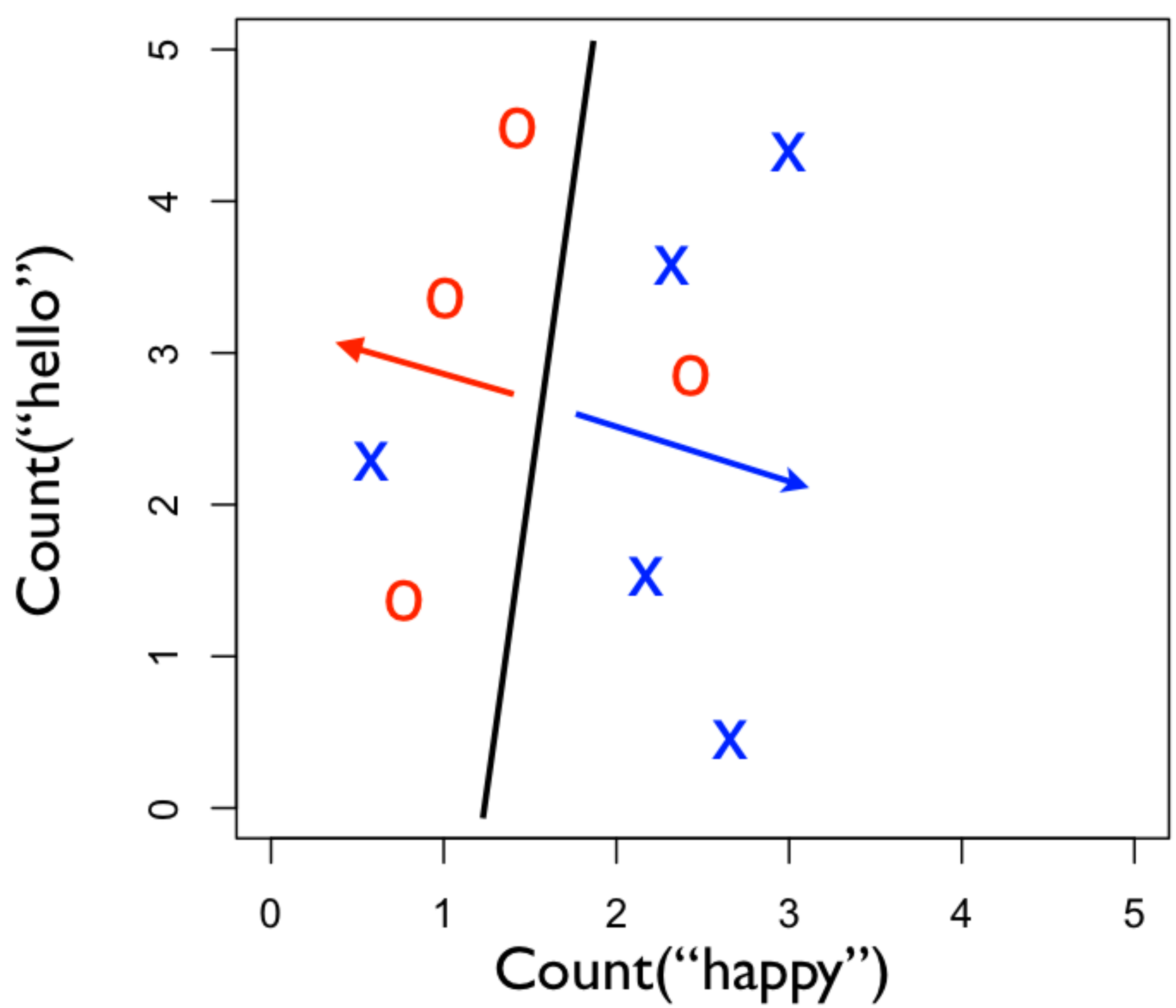
Predict y=1 when

$\beta^\mathsf{T} x > 0$

Predict y=0 when

$\beta^\mathsf{T} x \leq 0$





Application: doc sim to words

- Given a word list to represent a concept, can we score a document for how much it expresses that concept?
- DDR is a very simple embedding approach:
  - Average the word lists embeddings to create a concept vector
  - Average a doc's words to create a document vector
  - Apply cosine similarity!
- Supplying a set of keywords is *low-supervision*, or low-expertise, approach compared to labeling docs
  - Though you don't get a nice logreg probability (until you label some...)

[Garten et al. 2018]

LIWC "posemo" list

accept, accepta*, accepted, accepting, accepts, active*, admir*, ador*, advantag*, adventur*, affection*, agree, agreeab*, agreed, agreeing, agreement*, agrees, alright*, amaz*, amor*, amus*, aok, appreciat*, assur*, attachment*, attract*, award*, awesome, beaut*, beloved, benefic*, benefit, benefits, benefitt*, benevolen*, benign*, best, better, bless*, bold*, bonus*, brave*, bright*, brillian*, calm*, care, cared, carefree, careful*, cares, caring, casual, casually, certain*, challeng*, champ*, charit*, charm*, cheer*, cherish*, chuckl*, clever*, comed*, comfort*, commitment*, compassion*, compliment*, confidence, confident, confidently, considerate, contented*, contentment, convinc*, cool, courag*, create*, creati*, credit*, cute*, cutie*, daring, darlin*, dear*, definite, definitely, delectabl*, delicate*, delicious*, deligh*, determina*, determined, devot*, digni*, divin*, dynam*, eager*, ease*, easie*, easily, easiness, easing, easy*, ecsta*, efficien*, elegan*, encourag*, energ*, engag*, enjoy*, entertain*, enthus*, excel*, excit*, fab, fabulous*, faith*, fantastic*, favor*, favour*, fearless*, festiv*, fiesta*, fine, flatter*, flawless*, flexib*, flirt*, fond, fondly, fondness, forgave, forgiv*, free, freeb*, freed*, freeing, freely, freeness, freer, frees*, friend*, fun, funn*, genero*, gentle, gentler, gentlest, gently, giggl*, giver*, giving, glad, gladly, glamor*, glamour*, glori*, glory, good, goodness, gorgeous*, grace, graced, graceful*, graces, graci*, grand, grande*, gratef*, grati*, great, grin, grinn*, grins, ha, haha*, handsom*, happi*, happy, harmless*, harmon*, heartfelt, heartwarm*, heaven*, heh*, helper*, helpful*, helping, helps, hero*, hilarious, hoho*, honest*, honor*, honour*, hope, hoped, hopeful, hopefully, hopefulness, hopes, hoping, hug, hugg*, hugs, humor*, humour*, hurra*, ideal*, importan*, impress*, improve*, improving, incentive*, innocen*, inspir*, intell*, interest*, invigor*, joke*, joking, joll*, joy*, keen*, kidding, kind, kindly, kindn*, kiss*, laidback, laugh*, libert*, like, likeab*, liked, likes, liking, livel*, lmao, lol, love, loved, lovely, lover*, loves, loving*, loyal*, luck, lucked, lucki*, lucks, lucky, madly, magnific*, merit*, merr*, neat*, nice*, nurtur*, ok, okay, okays, oks, openminded*, openness, opport*, optimal*, optimi*, original, outgoing, painl*, palatabl*, paradise, partie*, party*, passion*, peace*, perfect*, play, played, playful*, playing, plays, pleasant*, please*, pleasing, pleasur*, popular*, positiv*, prais*, precious*, prettie*, pretty, pride, privileg*, prize*, profit*, promis*, proud*, radian*, readiness, ready, reassur*, relax*, relief, reliev*, resolv*, respect, revigor*, reward*, rich*, rofl, romanc*, romantic*, safe*, satisf*, save, scrumptious*, secur*, sentimental*, share, shared, shares, sharing, silli*, silly, sincer*, smart*, smil*, sociab*, soulmate*, special, splend*, strength*, strong*, succeed*, success*, sunnier, sunniest, sunny, sunshin*, super, superior*, support, supported, supporter*, supporting, supportive*, supports, suprem*, sure*, surpris*, sweet, sweetheart*, sweetie*, sweetly, sweetness*, sweets, talent*, tehe, tender*, terrific*, thank, thanked, thankf*, thanks, thoughtful*, thrill*, toleran*, tranquil*, treasur*, treat, triumph*, true, trueness, truer, truest, truly, trust*, truth*, useful*, valuabl*, value, valued, values, valuing, vigor*, vigour*, virtue*, virtuo*, vital*, warm*, wealth*, welcom*, well, win, winn*, wins, wisdom, wise*, won, wonderf*, worship*, worthwhile, wow*, yay, yays

Fig. 4 Nearest neighbors of the LIWC positive emotions dictionary

[Garten et al. 2018]