Maximum Entropy Classifier

Email Classification

Imagine

- You are not a human but a computer
- How would you answer the following question

Fried Chicken or Dog?

Maximum Entropy Classification

- For each feature
- e.g., “has a leash”,
- “has two dots that resemble eyes”, etc.
- Found for the data
- A weight is applied
- Weighted sum is normalized to give a fraction between 0 or 1
- Use this fraction to tell us the “score” of how confident
- Is this a dog?

Email #1:

From: LeChuck@yahoo.com

Har har!

Allow mes to introduce myself. Myname is Lechuck ,and I got your email from the interwebs mail directory. I threw a dart at the directory and hit your name. You seam like a good lad based on your name, so I here I am writing this email.

I live out here in this faraway island, and I have some moneys ($122K USD, to be exact) that I need to send overseas. Could you do mes a favor and help me with my moneys transfer?

1) Provide mes a bank account where this money would be transferred to.

2) Send me a picture of yourselfs so I know who to look for and thank when I sail to the US. Click heres to my Facebook [www.lechuck.myfacebook.com/fake.link/give_me_money] and post me your pic.

Email #2:

From: GuybrushThreepwood@gmail.com

Hi Nade,

This is Guybrush from James Logan High School. It's been a while! How have you beens?

I heard from Elaine that you wuld be visiting Melee Island soon, so I thought I should check in to see if u wanted to meet up. Are you avalable for dinner on Friday, September 8? We could catch up and hang out, like old times. I know a bar that serves some horrible grog.

Let me know! If we can't meet up, then no problem; I hope you enjoy your visit!

Cheers,

Guybrush

1. $f_1(email)$: Email contains spelling/grammatical errors
2. $f_2(email)$: Email asks for money to be transferred
3. $f_3(email)$: Email mentions account holder's name

Let's say that some of these features matter more than the others. Let's give the following weights for each of these features, for when they indicate spam:

1. $w_1(spam)$ (Email contains spelling/grammatical errors): 0.5. Seriously, proofread your emails.
2. $w_2(spam)$ (Email asks for money to be transferred): 0.2. Many email scams somehow involve some kind of bank transfer.
3. $w_3(spam)$ (Email mentions account holder's name): -0.5. If the sender mentions me by name, then maybe it isn't spam. By the way, notice that weights can be negative for a decision.

1. $w_1(spam)$ (Email contains spelling/grammatical errors): 0.5. Seriously, proofread your emails.
2. $w_2(spam)$ (Email asks for money to be transferred): 0.2. Many email scams somehow involve some kind of bank transfer.
3. $w_3(spam)$ (Email mentions account holder's name): -0.5. If the sender mentions me by name, then maybe it isn't spam. By the way, notice that weights can be negative for a decision.

Here are the weights for each of the features indicating NOT spam:

1. $w_1(notspam)$ (Email contains spelling/grammatical errors): -0.2
2. $w_2(notspam)$ (Email asks for money to be transferred): 0
3. $w_3(notspam)$ (Email mentions account holder's name): 1

Let a feature function, $f_i(x)$, take in an input, x, and return either 0 or 1, depending if the feature is present in x:

$$f(x) = \begin{cases} 1, & \text{if the feature is present in } x \\ 0, & \text{otherwise} \end{cases}$$

Furthermore, for N features, associate each feature function $f_i(x)$ with a weight $w_i(d)$, which is a number that denotes how "important" $f_i(x)$ is compared to other features for a decision, d (In this case, spam or not spam).

Entropy Classification

We can "model" (in my opinion, this word could be understood as "estimate") the score of a decision d on input x using the following procedure:

1. For each $f_i(x)$ in a set of N features, determine if $f_i(x)$ should be 1 or 0
2. Multiply each $f_i(x)$ with the associated weight $w_i(d)$, which depends on the decision d being evaluated.
3. Add up all of the weight*feature pairs: $sum_d = \sum_{i=1}^{N} w_i(d) * f_i(x)$
4. Throw the sum up into an exponent: $numerator_d = \exp(sum_d)$
5. Divide the sum by a number that will range the score between 0 and 1, and such that the sum of scores across all decisions is 1. It turns out that this is the sum of the numerators for every possible decision d: $denominator = \sum_d \exp(\sum_{i=1}^{N} w_i(d) * f_i(x))$

Entropy Classification

$$Score_d(x) = \frac{\exp(\sum_{i=1}^{N} w_i(d) * f_i(x))}{\sum_d \exp(\sum_{i=1}^{N} w_i(d) * f_i(x))}$$

Email #1: LeChuck

Here's the email again, for reference:

From: LeChuck@yahoo.com

Har har!

Allow mes to introduce myself. Myname is Lechuck ,and I got your email from the interrwebs mail directory. I threw a dart at the directory and hit your name. You seam like a good lad based on your name, so I here I am writing this email.

I live out here in this faraway island, and I have some moneys ($122K USD, to be exact) that I need to send overseas. Could you do mes a favor and help me with my moneys transfer?

1) Provide mes a bank account where this money would be transferred to.

2) Send me a picture of yourselfs so I know who to look for and thank when I sail to the US. Click heres to my Facebook

Entropy Classification

Let's see what features match up.

1. f_1: Email contains spelling/grammatical errors. YES.
2. f_2: Email asks for money to be transferred. YES.
3. f_3: Email mentions account holder's name. NO.

Spam Score

$$Score_{spam}(email_{LeChuck}) = \frac{\exp(\sum_{i=1}^{N} w_i(spam) * f_i(email_{LeChuck}))}{\sum_d \exp(\sum_{i=1}^{N} w_i(d) * f_i(email_{LeChuck}))} =$$

$$\frac{\exp(0.5*1+0.2*1-0.5*0)}{\exp(0.5*1+0.2*1-0.5*0)+\exp(-0.2*1+0*1+1*0)} = 0.71$$

Not Spam Score

$$Score_{notspam}(email_{LeChuck}) = \frac{\exp(\sum_{i=1}^{N} w_i(notspam) * f_i(email_{LeChuck}))}{\sum_d \exp(\sum_{i=1}^{N} w_i(d) * f_i(email_{LeChuck}))} =$$

$$\frac{\exp(-0.2*1+0*1+1*0)}{\exp(0.5*1+0.2*1-0.5*0)+\exp(-0.2*1+0*1+1*0)} = 0.29$$

LeChuck's spam score (0.71) is higher than the not spam score (0.29), so this is spam.

Email #2: Guybrush

Here's Guybrush's email below, for reference:

From: GuybrushThreepwood@gmail.com

Hi Nade,

This is Guybrush from James Logan High School. It's been a while! How have you beens?

I heard from Elaine that you wuld be visiting Melee Island soon, so I thought I should check in to see if u wanted to meet up. Are you avalable for dinner on Friday, September 8? We could catch up and hang out, like old times. I know a bar that serves some horrible grog.

Let me know! If we can't meet up, then no problem; I hope you enjoy your visit!

Entropy Classification

Once again, let's see what features match up.

1. f_1: Email contains spelling/grammatical errors. Yes… "beens"? "u"? "avalable"? C'mon Guybrush.
2. f_2: Email asks for money to be transferred. No.
3. f_3: Email mentions account holder's name. Yes.

Spam Score

$$Score_{spam}(email_{Guybrush}) = \frac{\exp(\sum_{i=1}^{N} w_i(spam) * f_i(email_{Guybrush}))}{\sum_d \exp(\sum_{i=1}^{N} w_i(d) * f_i(email_{Guybrush}))} =$$

$$\frac{\exp(0.5*1+0.2*0-0.5*1)}{\exp(0.5*1+0.2*0-0.5*1)+\exp(-0.2*1+0*1+1*1)} = 0.31$$

Not Spam Score

$$Score_{notspam}(email_{Guybrush}) = \frac{\exp(\sum_{i=1}^{N} w_i(notspam) * f_i(email_{Guybrush}))}{\sum_d \exp(\sum_{i=1}^{N} w_i(d) * f_i(email_{Guybrush}))} =$$

$$\frac{\exp(-0.2*1+0*1+1*1)}{\exp(0.5*1+0.2*0-0.5*1)+\exp(-0.2*1+0*0+1*1)} = 0.69$$

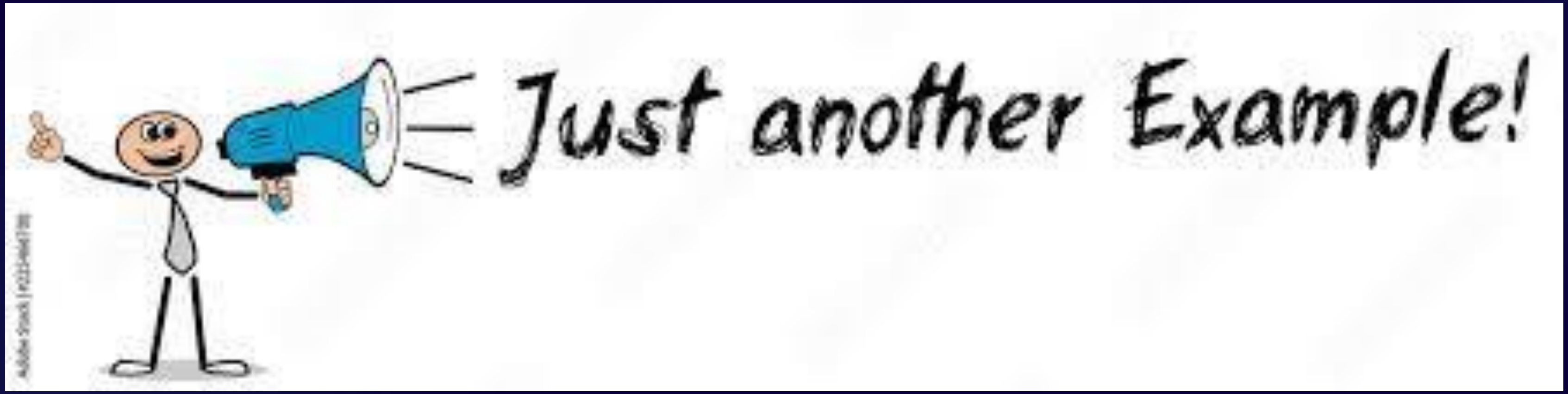
Guybrush's spam score (0.31) is lower than his not spam score (0.69)

Maximum Entropy Classification

Applications

MaxEnt classification is one of the more classical machine learning tasks, and solves problems beyond natural language processing. Here are a few:

- Sentiment analysis (e.g., given a product review, what does the reviewer like and dislike about the product?)
- Preferences (e.g., Given a person's demographics, who will a person vote for? Would they prefer Superman, Batman, or the Teenage Mutant Ninja Turtles? etc.)
- Diagnosis (e.g., Given characteristics of several medical images and patient history, what medical condition is a person at risk of having?)


Just another Example!

Maximum Entropy Classification

$$p(c|x) = \frac{\exp(\sum_{i=1}^{N} f_i(x) \times w_i(d))}{\sum_d \exp(\sum_{i=1}^{N} f_i(x) \times w_i(d))}$$

$$p(c|x) = \frac{1}{Z} exp(\sum_i w_i f_i)$$

For NLP,c is POS tag and x is observed word

Demonimator (Z) is normalizing constant

Stentence is: “The complex record”

The	complex	record
Det.	Verb	Verb
Noun	Adj.	Noun

Context/Features:

POS Tags	T_{i-1}	T_i	
Word	W_{i-1}	W_i	W_{i+1}

Assumption: All features having same weight which is equal to 1.

Beam Size=2

Maximum Entropy Classification

Features are taken as below. Observed word(W_i). Its previous word (W_{i-1}) and POS Tag (T_i and T_{i-1}).

Stentence is: "The complex record"

Feature	**The (Det.)**	**The (Noun)**	**Complex (Verb)**	**Complex (Adj.)**	**Record (Verb)**	**Record (Noun)**
F1:T_{i-1}=Det and T_i=Adj				**1**		
F2:T_{i-1}=Noun and T_i=Verb			**1**			
F3:T_{i-1}=Adj. And T_i=Noun						**1**
F4: W_{i-1}=the and T_i=Adj				**1**		
F5:W_{i-1}=the & W_{i+1}=record and T_i=Adj				**1**		
F6=:W_{i-1}=complex and T_i=Noun						**1**
F7:W_{i+1}=complex and T_i=Det	**1**					
F8:W_{i-1}=NULL and T_i=Noun		**1**				

Maximum Entropy Classification

The Word

The (Det)   F7 feature is present

The(Noun) F8 feature is present

$$p(\text{Det}|\text{The}) = \frac{\exp(1 * 1)}{\exp(1 * 1) + \exp(1 * 1)} = 0.5$$

$$p(\text{Noun}|\text{The}) = \frac{\exp(1 * 1)}{\exp(1 * 1) + \exp(1 * 1)} = 0.5$$

Maximum Entropy Classification

Prvious word The(Det), The (Noun) 0.5 probability

Observed Word is **complex** Word

For The(Det)

complex(Verb)   No feature is present

complex (Adj.) F1,F4,F5 feature is present

For The(Noun)

complex(Verb)   F2 feature is present

complex (Adj.) F4,F5 feature is present

Maximum Entropy Classification

Prvious word The(Det), The (Noun) 0.5 probability.   Observed Word is **complex** Word

For The(Det) :   complex(Verb)   No feature is present

complex (Adj.) F1, F4, F5 feature is present

$$p(\text{Verb}|\text{complex}) = \frac{\exp(0)}{[\exp(1*1)+\exp(1*1)+\exp(1*1)]+\exp(0)} = 0.11$$

$$= 0.11 \times 0.5 = 0.055$$

$$p(\text{Adj.}|\text{complex}) = \frac{\exp(1*1)+\exp(1*1)+\exp(1*1)}{[\exp(1*1)+\exp(1*1)+\exp(1*1)]+\exp(0)} = 0.89$$

$$= 0.89 \times 0.5 = 0.455$$

Prvious word The(Det), The (Noun) 0.5 probability. Observed Word is **complex** Word

For The(Noun)

complex(Verb) F2 feature is present

complex (Adj.) F4,F5 feature is present

$$p(\text{Verb}|\text{complex}) = \frac{\exp(1)}{[\exp(1*1) + \exp(1*1)] + \exp(1)} = 0.33$$

$$= 0.33 \times 0.5 = 0.165$$

$$p(\text{Adj.}|\text{complex}) = \frac{\exp(1*1) + \exp(1*1)}{[\exp(1*1) + \exp(1*1)] + \exp(1)} = 0.66$$

$$= 0.66 \times 0.5 = 0.33$$

Maximum Entropy Classification

First Word: The	Second Word:Complex	Answer
Det	Verb	0.055
Det	Adj.	**0.455**
Noun	Verb	0.165
Noun	Adj.	**0.330**

Beam Size=2

First Word: The	Second Word:Complex	Answer
Det	Adj.	**0.455**
Noun	Adj.	**0.330**

First Word: The	Second Word: Complex	Answer
Det	Adj.	**0.455**

For record : record(Verb) No feature is present

record (Noun) F3, F6 feature is present

$$p(\text{Verb}|\text{record}) = \frac{\exp(0)}{[\exp(1*1)+\exp(1*1)]+\exp(0)} = 0.16$$

$$= 0.16 \times 0.455 = 0.0728$$

$$p(\text{Noun}|\text{record}) = \frac{\exp(1*1)+\exp(1*1)}{[\exp(1*1)+\exp(1*1)]+\exp(0)} = 0.85$$

$$= 0.85 \times 0.455 = 0.387$$

First Word: The	Second Word:Complex	Answer
Noun	Adj.	**0.330**

For record : record(Verb) No feature is present

record (Noun) F3, F6 feature is present

$$p(\text{Verb}|\text{record}) = \frac{\exp(0)}{[\exp(1*1)+\exp(1*1)]+\exp(0)} = 0.16$$

$$= 0.16 \times 0.330 = 0.0528$$

$$p(\text{Noun}|\text{record}) = \frac{\exp(1*1)+\exp(1*1)}{[\exp(1*1)+\exp(1*1)]+\exp(0)} = 0.85$$

$$= 0.85 \times 0.330 = 0.281$$

Maximum Entropy Classification

First Word The	Second Word Complex	Third Word Record	Answer
Det	Adj. (0.455)	Verb	0.0728
Det	Adj. (0.455)	Noun	**0.387**
Noun	Adj. (0.330)	Verb	0.0528
Noun	Adj. (0.330)	Noun	0.281

The	complex	record
Det.	Adj	Noun


Take a
Break!

Study steps in Porter Stemmer

Porter stemmer algorithm https://vijinimallawaarachchi.com/2017/05/09/porter-stemming-algorithm/