Classification

Anand Paul

Regression as a Classifier

- Regression model can predict a continue values such as height, weight.

- They can also predict probabilities, such as the probability that an sound signal contain human voice. A probability-predicting regression model can be used as part of a classifier by imposing a decision rule - for example, if the probability is 50% or more, decide it's a human voice.

Difference Between Regression and Classification

Regression: the output variable takes continuous values.
Classification: the output variable takes class labels.

Regression means to predict the output value using training data.

Classification means to group the output into a class.

Pointers on Classification

- Classical regression as a basis for classification (say for example logistic regression, as it has regression values 0 and 1.
- That is similar to a binary classification (we can use a simple threshold value to classify
- We have to **model the threshold** properly as it would affect the overall performance. For example we have a threshold value and if the values are above this threshold value then we mark it as a viral disease else normal.

“spam” or “non-spam” for emails.

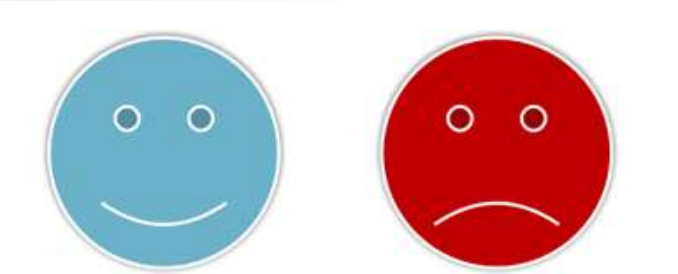
Types of Classification

- Sentiment Classifier : This classifier determines if a text is positive or negative. It is well suited for both short and long texts (tweets, Facebook statuses, blog posts, product reviews etc)

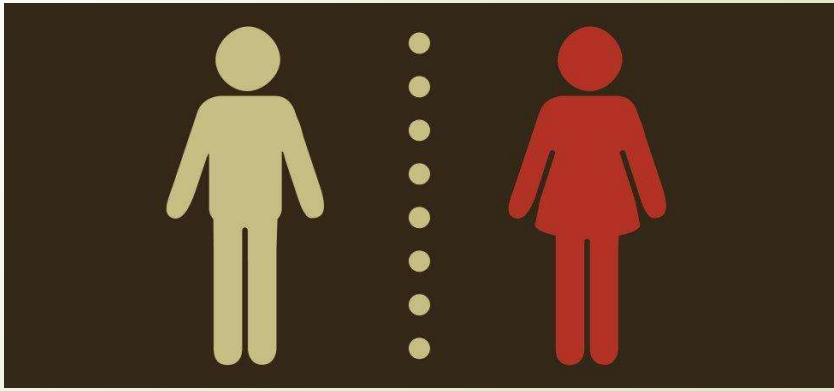
- Topic Classifier : While visiting a webpage we can Categories English text in the page into a topic (Arts, Business, Computers, Games, Health, Home, Recreation, Science, Society and Sports).

- Image Classification : Input an image and classify it as a boy or girl, dog or cat

- Mood Classifier : The state of mind of the writer - upset or happy.



- Gender Classifier : This classifier tries to figure out if a text is written by a male or female



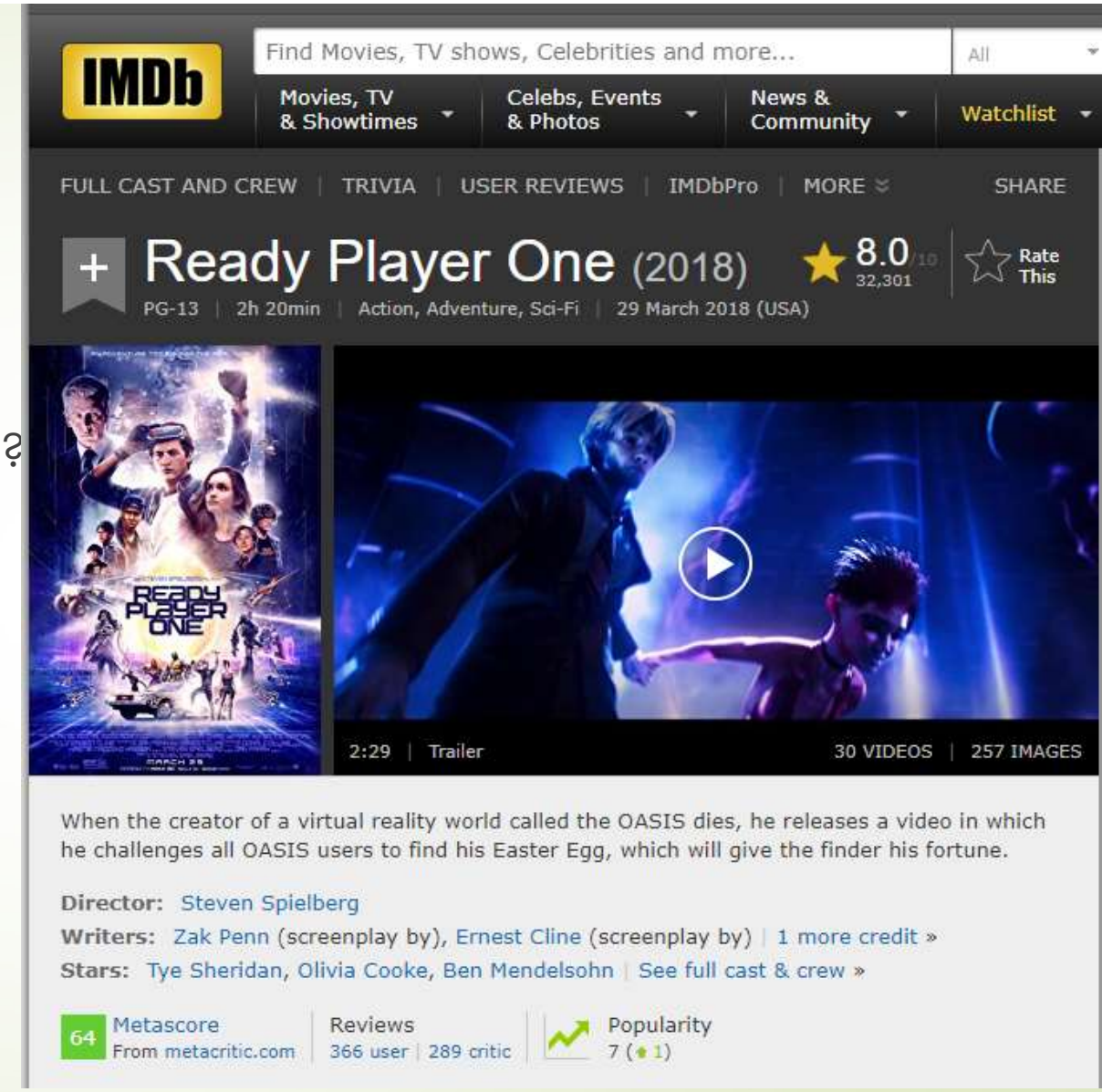
Movie Reviews

- Kind of sentiment analysis (classifier)
- Similar principal can be use for product review, Restaurant review.

Movie Review

- You want to see its rating
- What do people say about the movie?
- how do we classify a positive and Negative reviews

This classifier determines if a text is positive or negative.



WU1

Sample reviews

A Very Cool and Very Exciting Movie, but suitable for teens and up.

This Super Duper Cool and Exciting Movie is Coming Up on SO Track and Field Day at Ron Poe Stadium. But a Little More Violent Than "Star Wars: The Last Jedi"(which is suitable for ages 10+ or ages 12+) For This Movie: Ready Player One, Is for ages 13+ (Forecast)

Stellar Sci-fi Adventure with incredible effects, characters and story

Ready Player One is an incredibly impressive film overall. The majority of the film takes place in a virtual world called the Oasis, and the effects are fantastic. The characters are good and played well, and I loved the story. I thought everything about the movie was great. A lot of the movie has video game style violence and extremely over the top action and battle scenes. There are horror movie references and an entire scene based off of The Shining. Few characters die in the real world, but things get pretty crazy in the virtual one. There is a scene with a nude woman, but the angles don't reveal much. There is also a lot of swearing. This is definitely not for younger kids, but teens should be okay if they like lots of crazy action. I highly recommend this film.

Not for fans of the book

My husband and I love the book, unortunately the screen writer didn't. So much of the story was changed that I was confused with the story unfolding on the screen. If you loved the book, don't waste your money or time on this movie. If you haven't read the book, you'll probably think it's great. We were very disappointed.

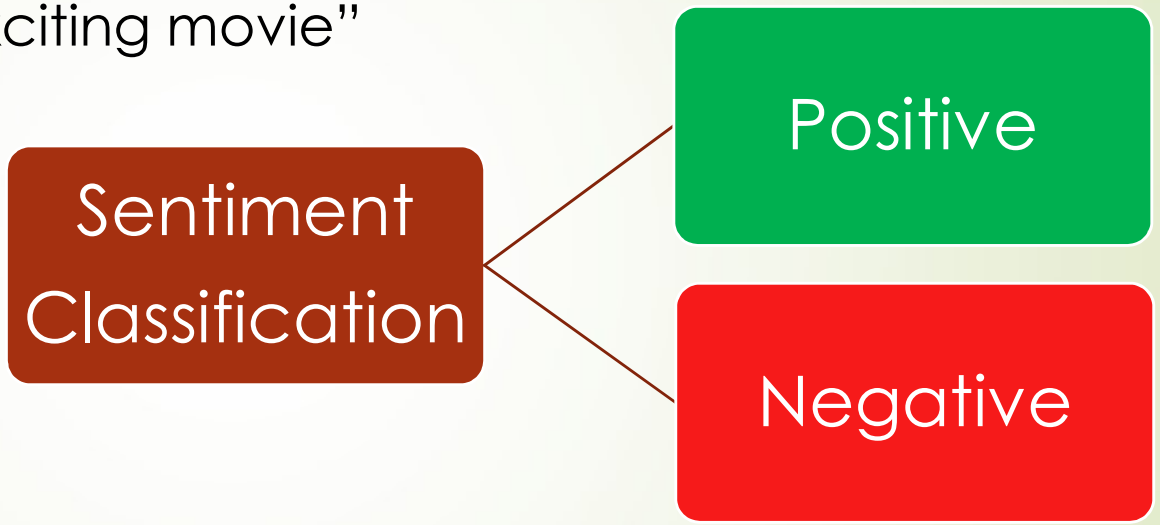
https://www.commonsensemedia.org/movie-reviews/ready-player-one/user-reviews/adult

Slide 9

WU1 Windows User, 02/04/2018

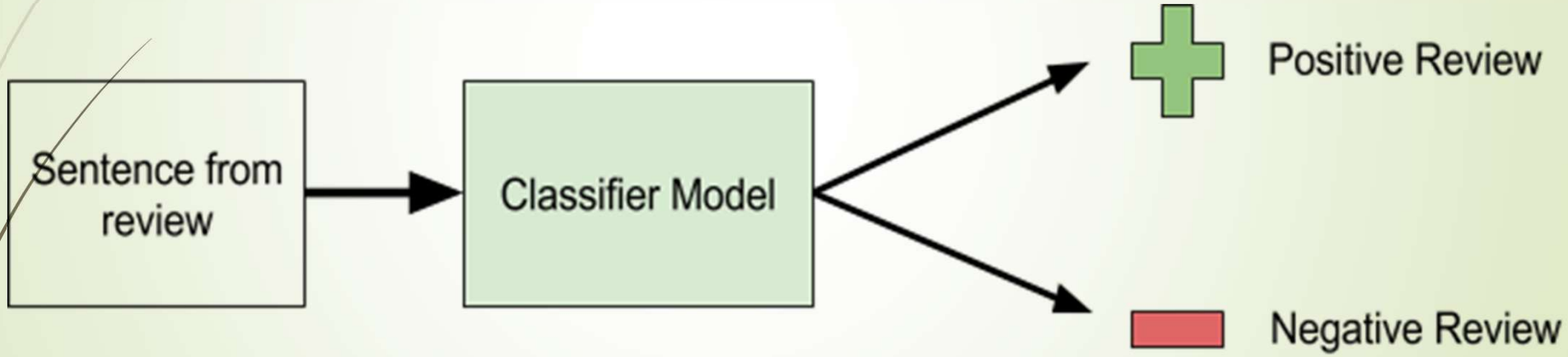
Classification Engine

"Super Duper Cool and Exciting movie"

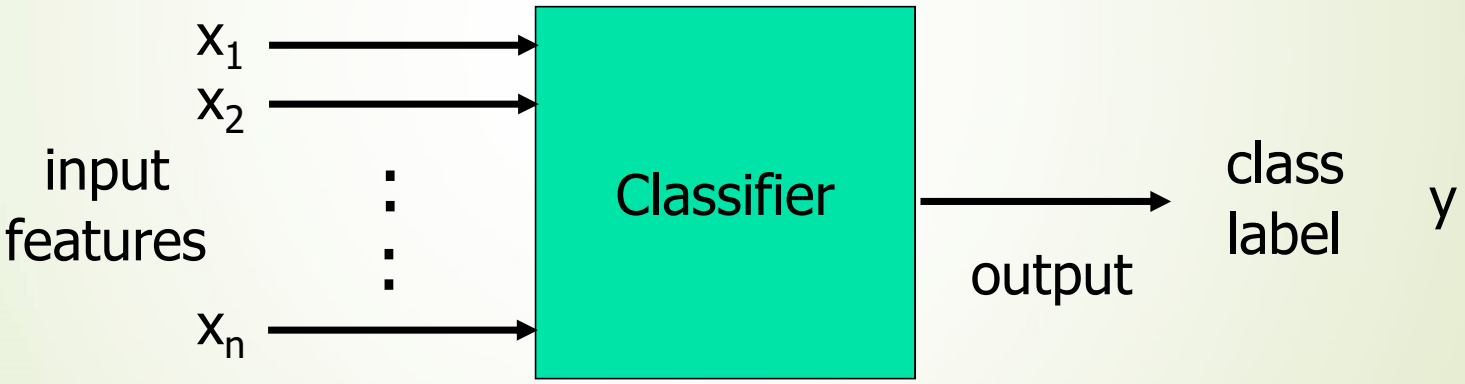


- Taking all the reviews and we're going to break them up into sentences. So each review is composed of multiple sentences and some sentences cover different aspects of the movie.

Classifier


Sentence from review
Classifier Model
Positive Review
Negative Review

Classifier Model



List of positive words	List of negative words
Fantastic , Super, great, awesome, good, amazing, incredible…	Disappointed, horrible, bad, terrible, disgusting, waste…

Simple threshold classifier

Count positive & negative words in sentence

If
number of positive words >
number of negative words:

Then output y = +

Else:

y = −

Sentence from review
Input x

Problems with threshold classifier

- How do we get list of positive/negative words?
- Words have different degrees of sentiment:
  - Great >good
  - How do we weigh different words?
- Single words are not enough:
  - *Good* ➔➔Positive
  - *Not good* ➔➔Negative

Addressed by learning a classifier

Addressed by more elaborate features

A (linear) classifier

- Will use training data to learn a weight for each word

Word	Weight
good	1.0
great	1.5
awesome	2.5
bad	-1.0
terrible	-2.1
awful	-3.3
restaurant, the, we, where, …	0.0
…	…

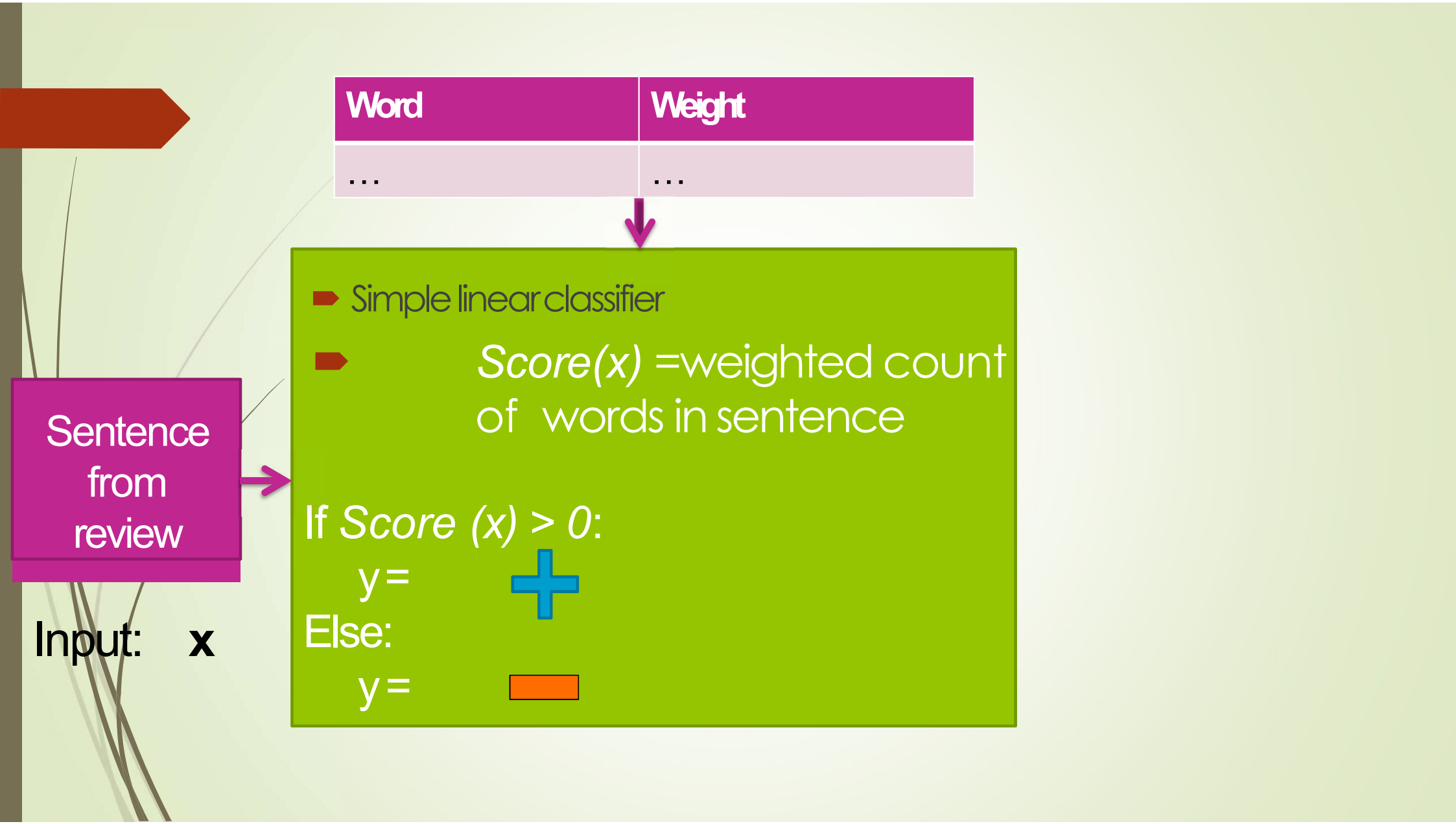
grading a sentence

Word	Weight
good	1.0
great	1.5
awesome	2.5
bad	-1.0
terrible	-2.8
awful	-3.3
Movie , the, we, where, …	0.0
…	…

Input x:

Movie was incredible
the action was awesome, but the sstoryline was terrible.

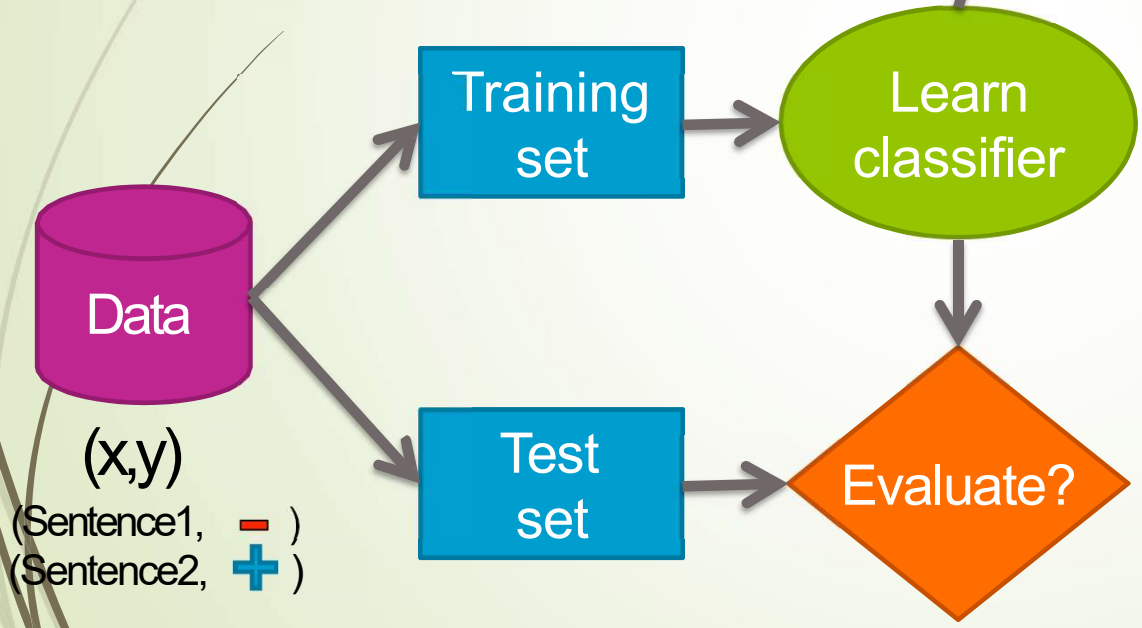
Called a linear classifier, because output is weighted sum of input.


Word
Weight
…
…
Simple linear classifier
Score(x) =weighted count of words in sentence
If Score (x) > 0:
y=
Else:
y=
Sentence from review
Input: x

Training a classifier =Learning the weights

Word	Weight
good	1.0
awesome	2.5
bad	-1.0
awful	-3.3
…	…

These weights are going to be used to score every element in the test set and evaluate how good is our of classification



Classification Error

Movie was great 

Learned classifier

y = 

Test example

Correct	1
Incorrect	0

Let's see if the classifier gets the true label right. (hide that true label).

Classification Error

Movie was OK 

Learned classifier

y = 

Mistake

Test example

y = 

Correct	1
Incorrect	1

Do this for every sentence in the review

Classification error & accuracy

- Error measures of incorrect
  - Best possible value is 0.0
- Accuracy measure
  - Best possible value is 1.0

Error = No. of incorrect

Total No. of Sentences

Accuray = No. of correct

Total No. of Sentences

What if you ignore the sentence, and just guess?

- For binary classification:
  - Half the time, you'll get it right! (on average)

    ➔➔accuracy = 0.5

- For k classes, accuracy = 1/k
  - 0.333 for 3 classes, 0.25 for 4 classes,…

At the very, very, very least,
you should healthily beat random…
Otherwise, it's (usually) pointless…

Is a classifier with 90% accuracy good? Depends…

2010 data shows:
"90% emails sent are spam!"

Predicting every email is spam gets you 90% accuracy!!!

Majority class prediction

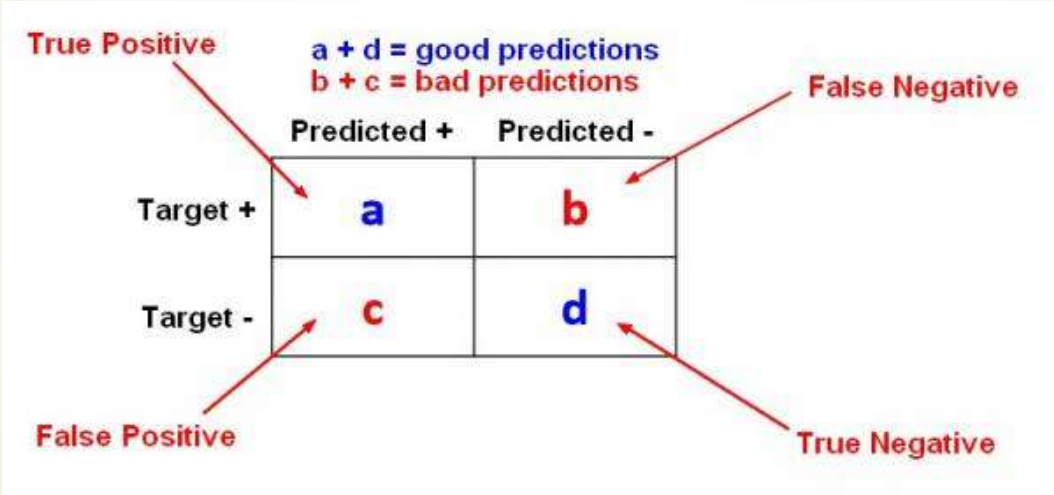
Amazing performance when there is class imbalance
(but silly approach)

- One class is more common than others
- Beats random (if you know the majority class)

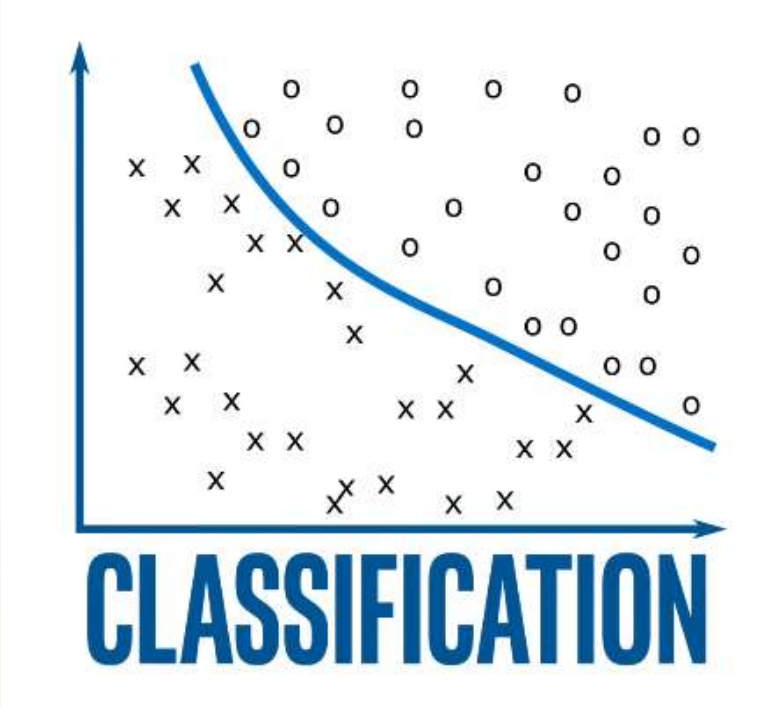
So, always be digging in and asking the hard questions about reported accuracies

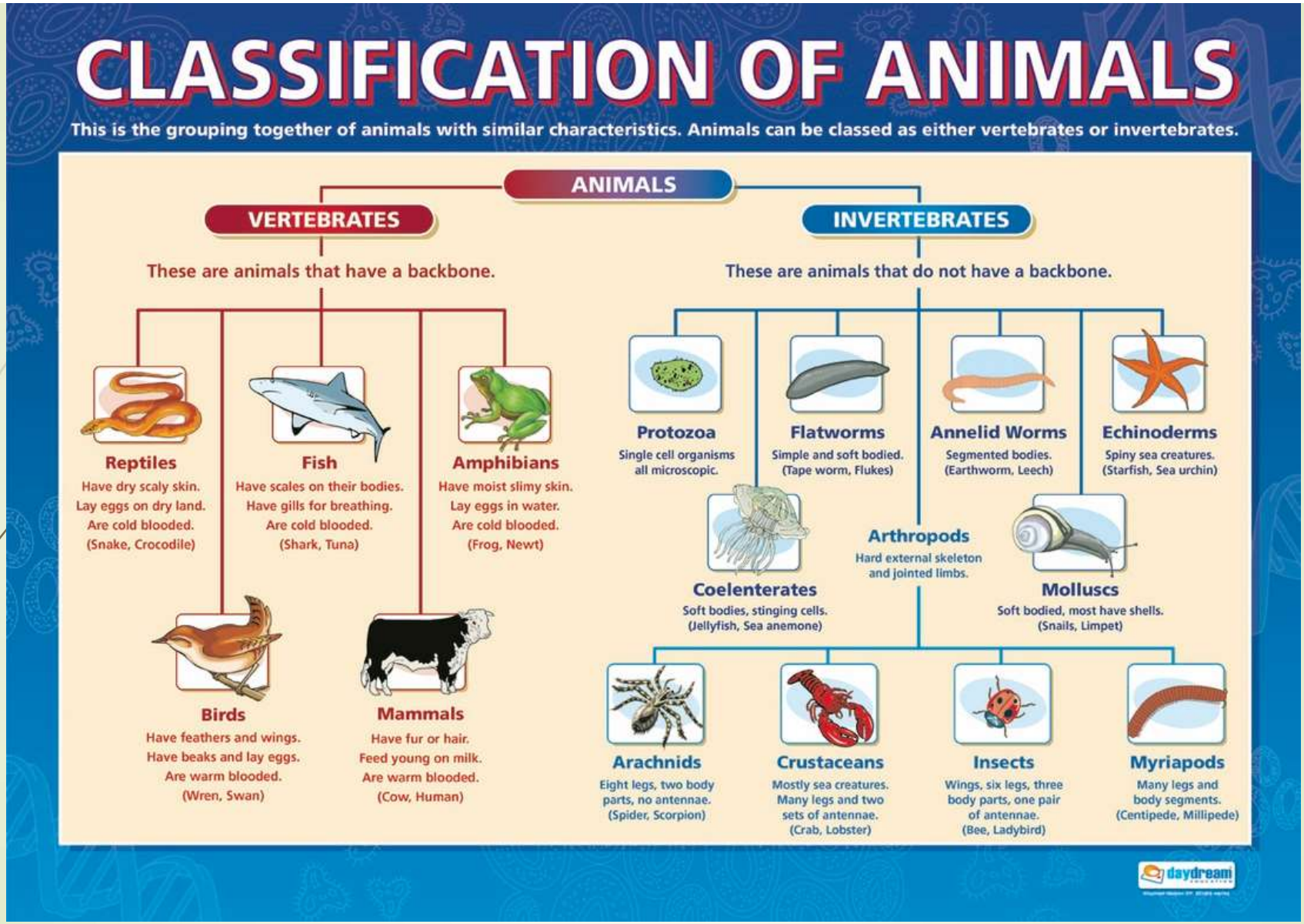
- Is there class imbalance?
- How does it compare to a simple, baseline approach?
  - Random guessing
  - Majority class
  - …
- Most importantly:
  what accuracy does my application need?
  - What is good enough for my user's experience?
  - What is the impact of the mistakes we make?

Confusion matrix – binary classification



Classification





http://yr3wccnws.weebly.com/integrated-studies---term-two/classification-of-living-things-poster-project