

Mini-Challenge 2: Sentiment Classification

Introduction

Daniel Perruchoud
Joe Weibel

30. September 2025



Sentiment Classification

This sound track was beautiful! It paints the senery in your mind so well I would recomend it even to people who hate vid. game music! I have played the game Chrono Cross but out of all of the games I have ever played it has the best music! It backs away from crude keyboarding and takes a fresher step with grate guitars and soulful orchestras. It would impress anyone who cares to listen! ^_^

positive

A complete waste of time. Typographical errors, poor grammar, and a totally pathetic plot add up to absolutely nothing. I'm embarrassed for this author and very disappointed I actually paid for this book.

negative

If you have a kid who likes to move the trains by hand, don't spend the extra money on this. I think it's probably a good product, just not for a 3 yr old.

positive

or

negative

or

neutral

Sentiment Classification

binary

positive

negative

multiclass

positive

neutral

negative

continuous

1.0

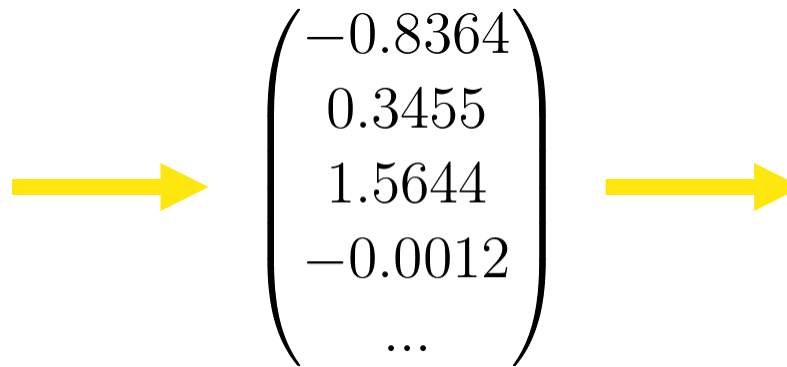
...

0.0

Binary Classification

A complete waste of time.
Typographical errors, poor grammar,
and a totally pathetic plot add up to
absolutely nothing. I'm embarrassed
for this author and very disappointed I
actually paid for this book.

Text



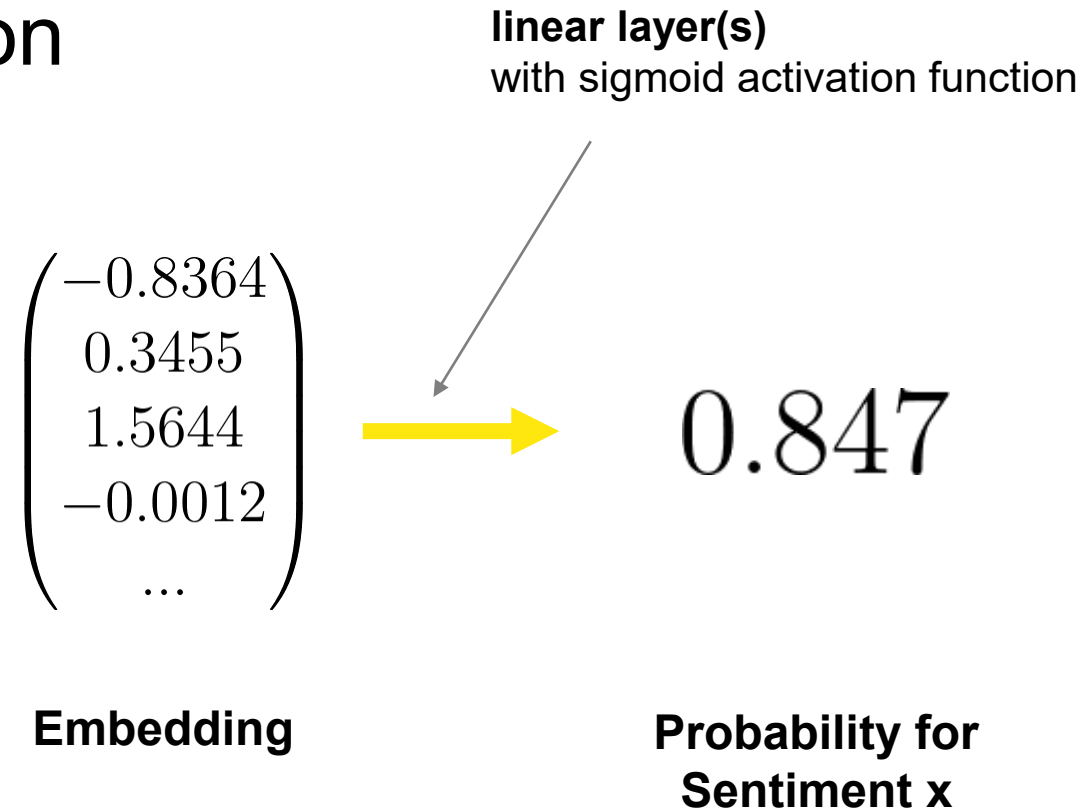
$$\begin{pmatrix} -0.8364 \\ 0.3455 \\ 1.5644 \\ -0.0012 \\ \dots \end{pmatrix}$$

Embedding

0.847

**Probability for
Sentiment x**

Binary Classification



Binary Classification

A complete waste of time.
Typographical errors, poor grammar,
and a totally pathetic plot add up to
absolutely nothing. I'm embarrassed
for this author and very disappointed I
actually paid for this book.

Text

static representations
(word2vec, GloVe, TF-IDF, ...)


$$\begin{pmatrix} -0.8364 \\ 0.3455 \\ 1.5644 \\ -0.0012 \\ \dots \end{pmatrix}$$

Embedding

contextual language model
(BERT, Sentence-Transformer, ...)

Bidirectional Encoder Representations from Transformers (BERT)



	[CLS]	My	name	is	BERT	
768 ↓	-0.8439	-1.3409	1.9044	1.0563	-0.5644	...
	0.3438	-2.0113	3.9534	2.0488	0.9199	...
	1.4931	-0.8347	-0.0011	1.4566	2.0013	...

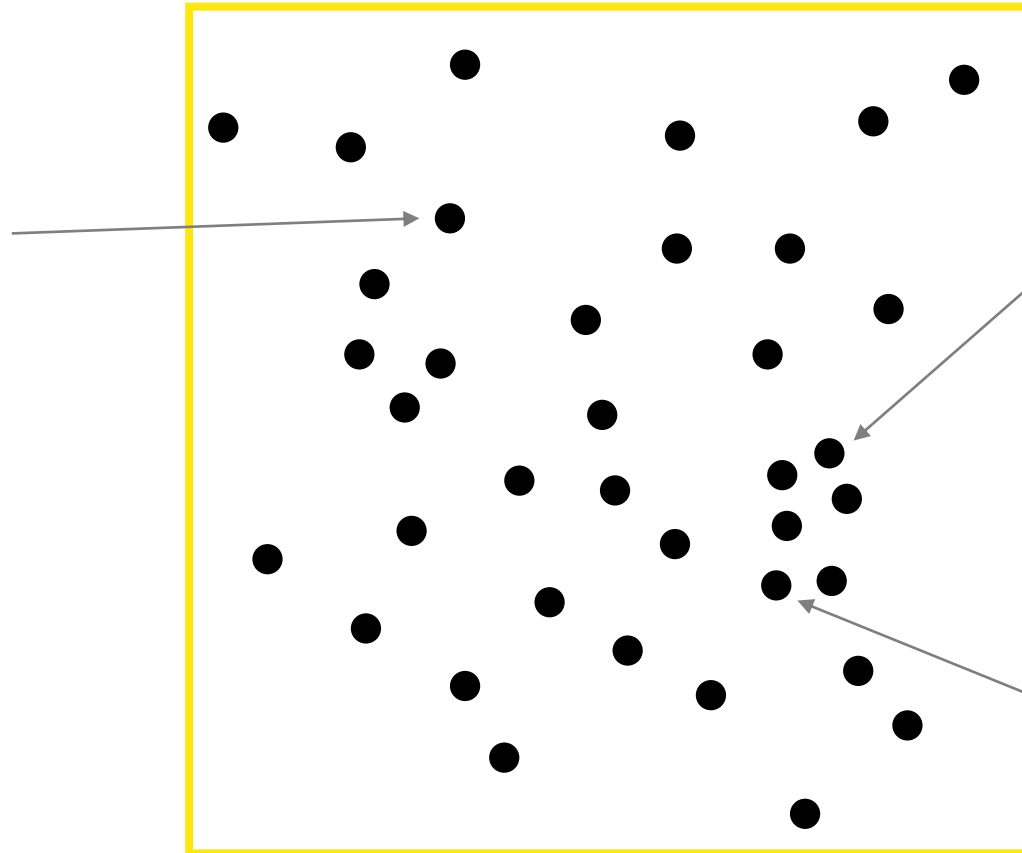
	0.4344	1.0947	-0.4267	-1.5794	0.4545	...

Embeddings with BERT

- represent word and text semantically (not only syntactically)
- contextualize word embeddings (eg. bank)
- not understandable by humans
- basis for different downstream tasks (eg. sentiment classification)



A complete waste of time.
Typographical errors, poor
grammar, and a totally pathetic
plot add up to absolutely
nothing. I'm embarrassed for
this author and very
disappointed I actually paid for
this book.



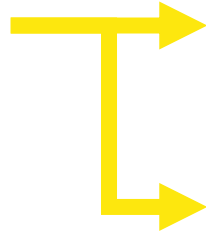
If you have a kid who likes to
move the trains by hand, don't
spend the extra money on this. I
think it's probably a good
product, just not for a 3 yr old.

I am quite sure any of you
actually taking the time to read
this have played the game at
least once, and heard at least a
few of the tracks here. And
whether you were aware of it or
not, Mitsuda's music contributed
greatly to the mood of every
single minute of the whole
game. Composed of 3 CDs and
quite a few songs ...

Data Is All You Need

more is generally better

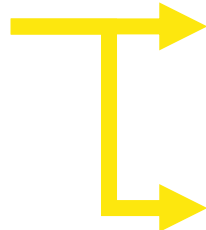
for **Embeddings**



Unlabelled Text Corpus
(books, Wikipedia, ...)

Pretrained models from 🤗 **Hugging Face**

for **Sentiment Classification**



Supervised Learning
with sentiment labels

Semi-Supervised Learning
with weak sentiment labels

Weak Labelling

I am quite sure any of you actually taking the time to read this have played the game at least once, and heard at least a few of the tracks here. And whether you were aware of it or not, Mitsuda's music contributed greatly to the mood of every single minute of the whole game. Composed of 3 CDs and quite a few songs (I haven't an exact count), all of which are heart-rendering and impressively remarkable, this soundtrack is one I assure you you will not forget. It has everything for every listener -- from fast-paced and energetic (Dancing the Tokage or Termina Home), to slower and more haunting (Dragon God), to purely beautifully composed (Time's Scar), to even some fantastic vocals (Radical Dreamers). This is one of the best videogame soundtracks out there, and surely Mitsuda's best ever. ^_^

positive

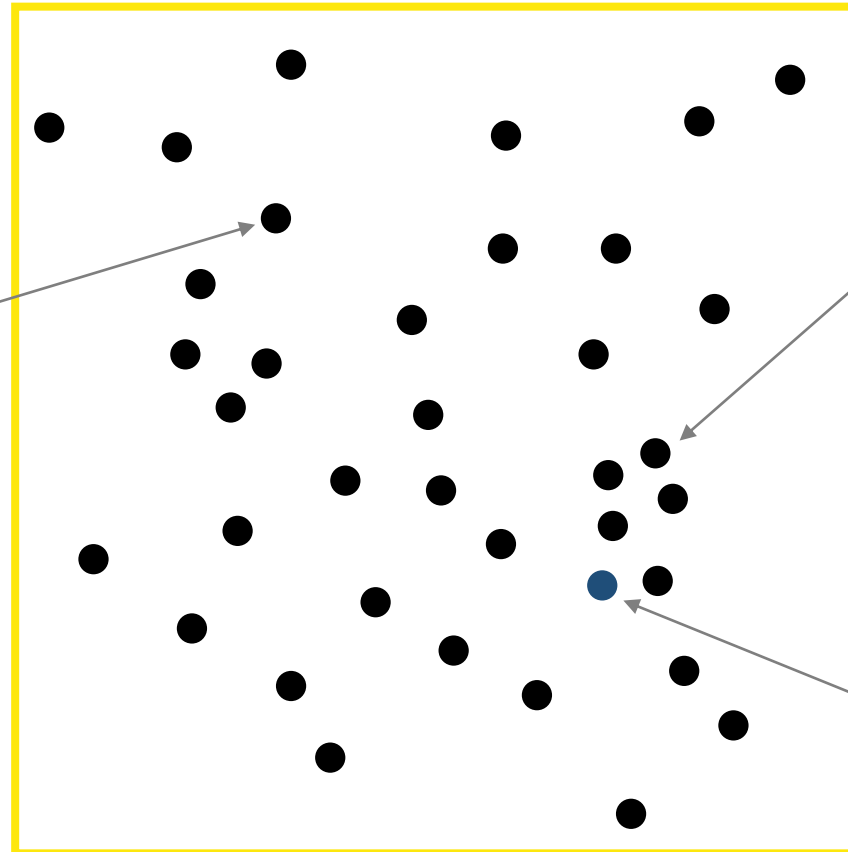
or

negative

Weak Labelling with KNN

A complete waste of time.
Typographical errors, poor
grammar, and a totally pathetic
plot add up to absolutely
nothing. I'm embarrassed for
this author and very
disappointed I actually paid for
this book.

negative



This sound track was beautiful!
It paints the senery in your mind
so well I would recomend it
even to people who hate vid.
game music! I have played ...

positive

I am quite sure any of you
actually taking the time to read
this have played the game at
least once, and heard at least a
few of the tracks here. And
whether you were aware of it or
not, Mitsuda's music contributed
greatly to the mood of every
single minute of the whole
game. Composed of 3 CDs and
quite a few songs ...

positive

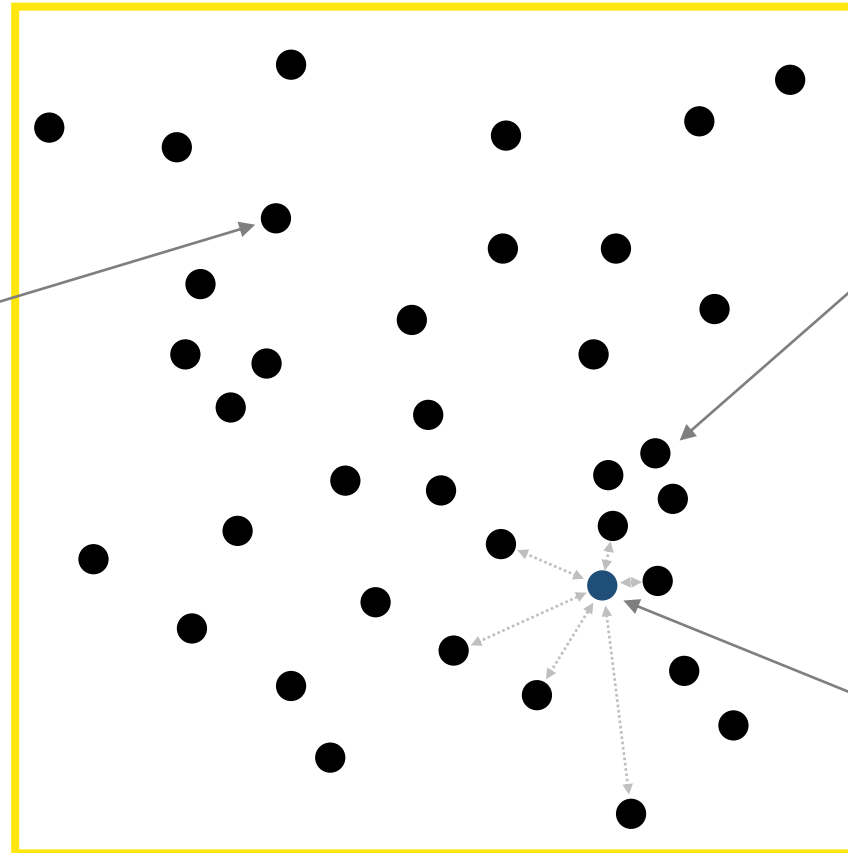
or

negative

Weak Labelling with KNN

A complete waste of time.
Typographical errors, poor
grammar, and a totally pathetic
plot add up to absolutely
nothing. I'm embarrassed for
this author and very
disappointed I actually paid for
this book.

negative



This sound track was beautiful!
It paints the senery in your mind
so well I would recomend it
even to people who hate vid.
game music! I have played ...

positive

I am quite sure any of you
actually taking the time to read
this have played the game at
least once, and heard at least a
few of the tracks here. And
whether you were aware of it or
not, Mitsuda's music contributed
greatly to the mood of every
single minute of the whole
game. Composed of 3 CDs and
quite a few songs ...

positive

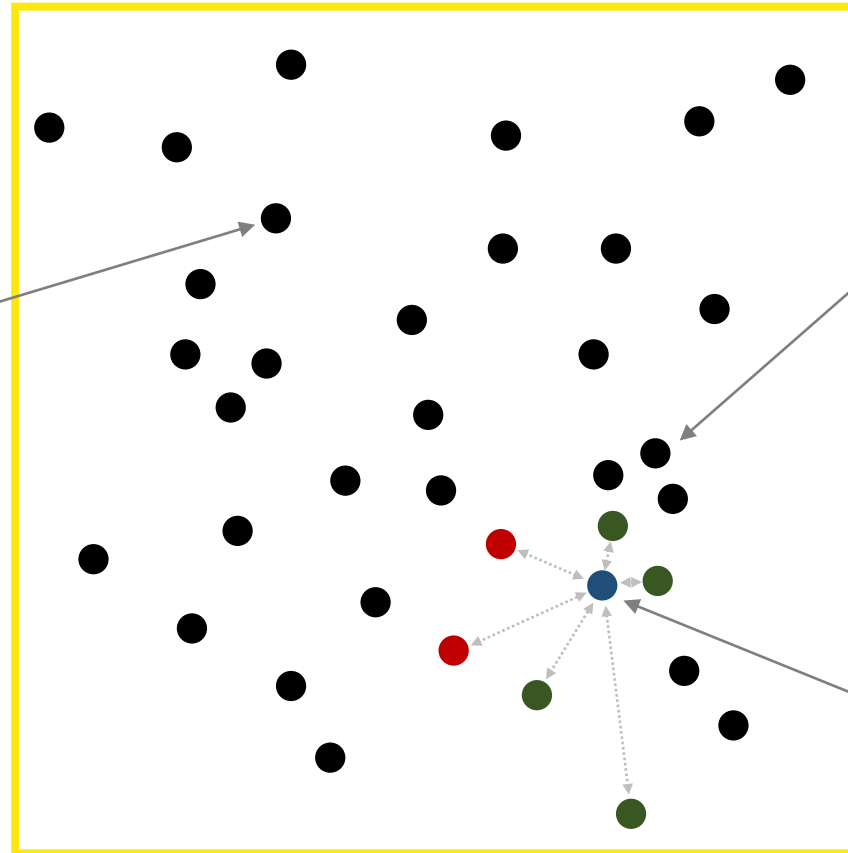
or

negative

Weak Labelling with KNN

A complete waste of time.
Typographical errors, poor
grammar, and a totally pathetic
plot add up to absolutely
nothing. I'm embarrassed for
this author and very
disappointed I actually paid for
this book.

negative



This sound track was beautiful!
It paints the senery in your mind
so well I would recomend it
even to people who hate vid.
game music! I have played ...

positive

I am quite sure any of you
actually taking the time to read
this have played the game at
least once, and heard at least a
few of the tracks here. And
whether you were aware of it or
not, Mitsuda's music contributed
greatly to the mood of every
single minute of the whole
game. Composed of 3 CDs and
quite a few songs ...

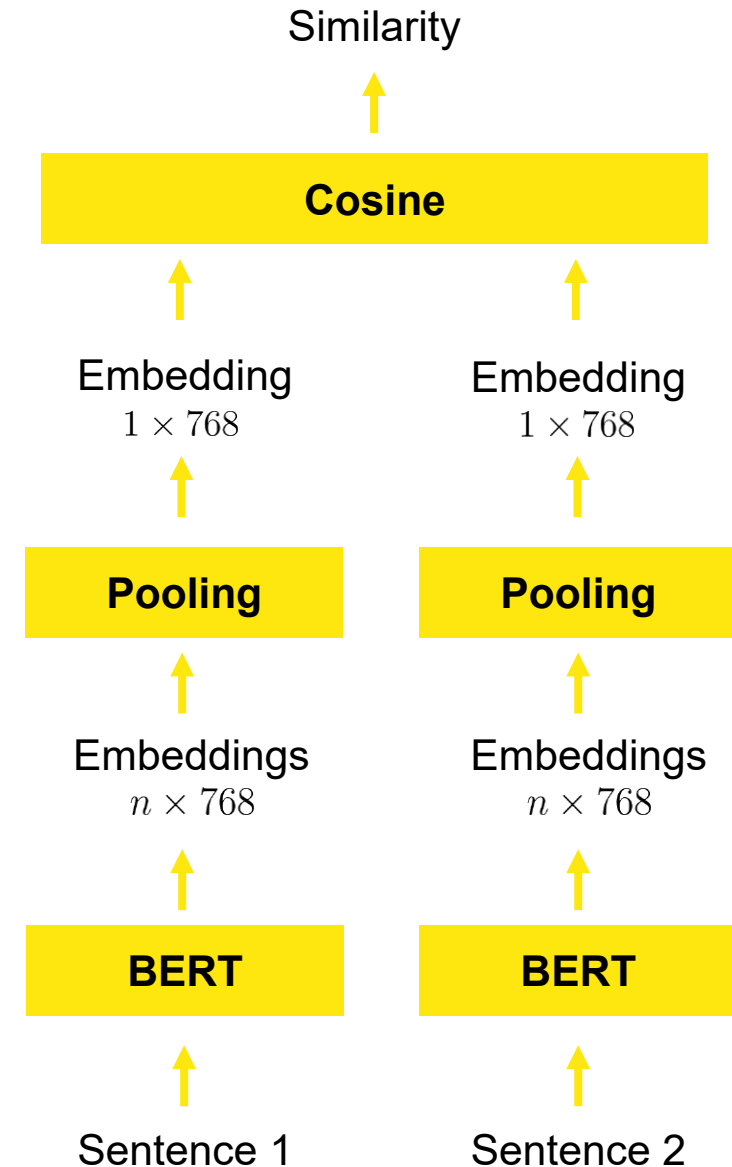
positive

or

negative

Sentence Transformers

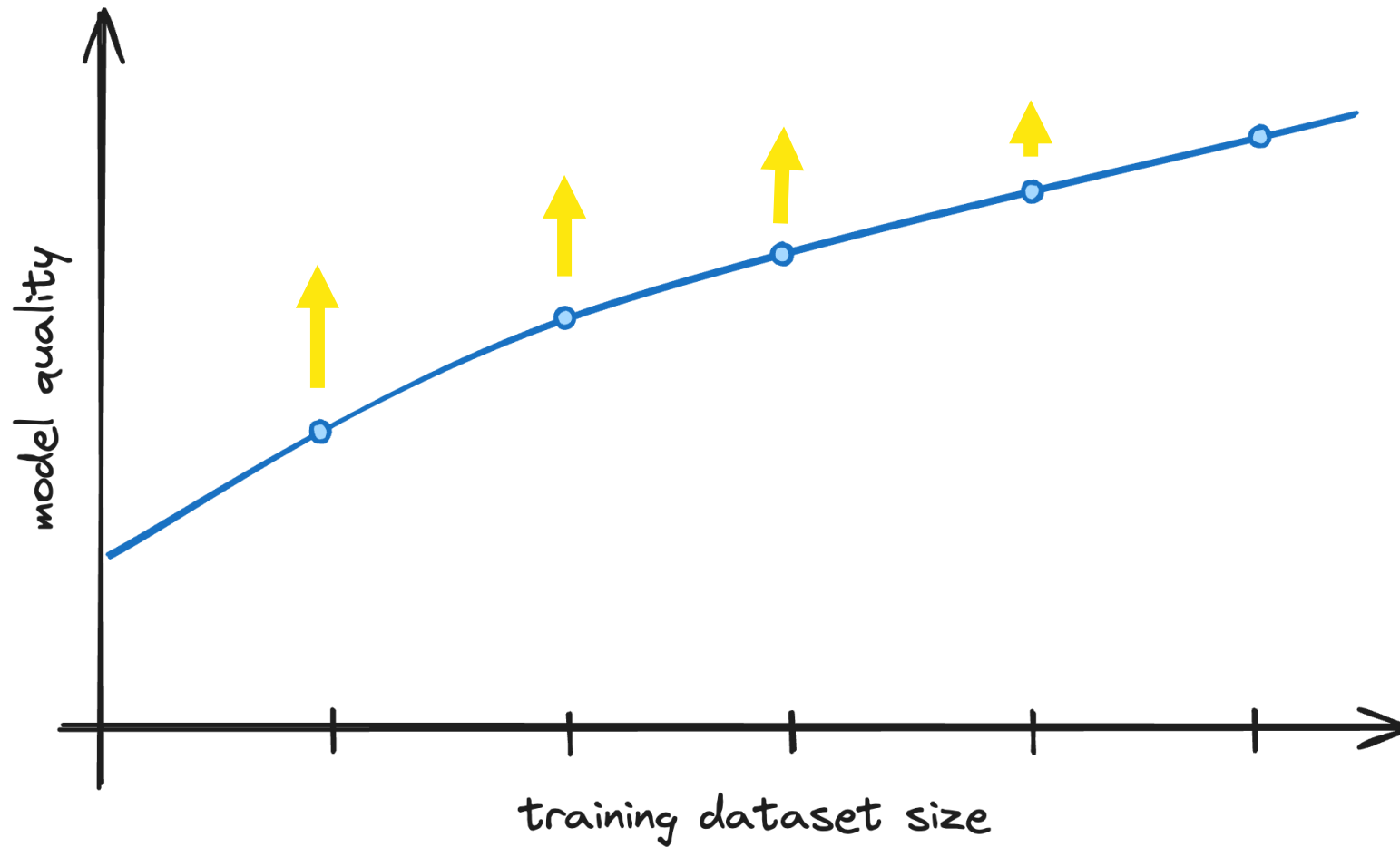
- BERT models not suitable for efficient sentence similarity search
- SBERT focuses on comparable sentence embeddings (cosine similarity)
- Siamese network based on BERT model



Mini Challenge Goals

1. Understand the concept of embeddings
2. Inspect the effect of more training data
3. Show the impact of weak labelling on sentiment classification
4. Analyse the effect of inter-annotator agreement
5. Have fun

Learning Curve



Datasets

- Use Financial Phrasebank dataset
- Number of samples for fine-tuning: 1000 labelled, rest unlabelled
- Build hierarchically nested data splits
- Check for maximum text length
- Balance dataset by undersampling

Proposed Approach

1. Choose the framework for this work
2. Train baseline classification model
3. Calculate and analyze embeddings
4. Try weak labelling techniques
5. Train classification model with additional weak labels
6. Compare results and conclude

Infrastructure

- Sentiment classification model training is computationally expensive
- Possible with Apple Silicon chips or Nvidia GPU with sufficient memory
- Alternatively use CSCS