

Opinion Mining

Analyzing and Aggregating Sentiment

Thomas Vakili, adapted from Rebecka Weegar

Sentiment in text

Sentiment and opinions

An **opinion holder** has an **opinion** about an **object**

- Direct opinions

I think **this watch** is **beautiful**

Sentiment and opinions

An **opinion holder** has an **opinion** about an **object**

- Direct opinions

I think **this watch** is **beautiful**

- Indirect opinions

After **the treatment**, my knee felt **worse**

Sentiment and opinions

An **opinion holder** has an **opinion** about an **object**

- Direct opinions

I think **this watch** is **beautiful**

- Indirect opinions

After **the treatment**, my knee felt **worse**

- Comparisons

KFC is **tastier** than **McDonalds**

Polarity

Opinions exist on a spectrum of polarity:

Negative ← **Neutral** → **Positive**

Examples:

- I love to read
- This is my favourite book
- 10 AM is a good time for a lecture
- Four people were injured in the accident

Subjectivity and objectivity

- Subjective statements concern **internal** attitudes

I like this camera

- Objective statements are **observable**

This phone is black

- Can objective statements convey sentiment?

Subjectivity and objectivity

- Subjective statements concern **internal** attitudes

I like this camera

- Objective statements are **observable**

This phone is black

- Can objective statements convey sentiment?

This phone has a three-day battery life

Who is the opinion holder?

An **opinion holder** has an **opinion** about an **object**

- Direct opinions

  **I** think **this watch** is **beautiful**

- Indirect opinions

After **the treatment**, my knee felt **worse**

- Comparisons

KFC is **tastier** than **McDonalds**

Subjective or objective?

Concerns *internal feelings* but is *objective*

After **the treatment**, my knee felt **worse**

Form is *objective* but the context is *subjective*

KFC is **tastier** than **McDonalds**

Compare this to:

KFC is **more expensive** than **McDonalds**



Comparisons

- **Gradable:** *X is nicer than Y*
- **Non-Gradable:** *X has face recognition, Y is waterproof*
- **Equative:** *X and Y are both very nice*
- **Superlative:** *X is the nicest on the market*
- Comparative sentences use **morphemes** or **syntax**
 - *more/most, -er/-est, less/least*
 - *than* and *as*, indicates what we are comparing objects

Opinion mining

Aggregating many opinions

Facts vs Opinions

- Why don't we just retrieve opinions?
- Search engines retrieve and rank *facts*
 - One fact = Two facts = Three facts
 - Importance is *relevance* to search query
- Individual opinions are not really useful as facts
 - Opinions are more interested when *aggregating them*
 - In fact: even individual opinions need aggregation!

A review

The Guardian reviewed the movie *Wicked*

<https://www.theguardian.com/film/2024/nov/24/wicked-the-film-review-cynthia-erivo-and-ariana-grande-jon-m-chu-part-one-1>

There's some kind of magic afoot. If, like me, you're one of the very few people who haven't already seen the blockbuster stage musical *Wicked* (it's the second-highest-grossing Broadway show of all time, so that's an awful lot of bums on seats), you may approach this shiny, high-energy, relentlessly marketed movie adaptation with low to moderate expectations. There's the unwieldy running time, for a start - two hours and 40 minutes - and the cynical, box-office-gouging decision to carve the story into two films (fans will have to wait almost a year to the day before they get to watch the concluding chapter).

But here's the thing: reservations are soon extinguished and grumbles about the release strategy swiftly quashed. *Wicked* matches its polished razzle-dazzle with real heart. Driven by knockout performances from Cynthia Erivo and Ariana Grande, Jon M Chu's impossibly slick charm assault of an adaptation zips along so enjoyably that you almost wish it were longer (your bladder may disagree). With its all too timely themes of bullying, corrupt leaders and the demonisation of difference, this is a movie that promises a froth of pink and green escapism but delivers considerably more in the way of depth and darkness.

For those who have somehow evaded *Wicked*'s cultural reach over the past couple of decades, here's a brief primer. The film and the stage show are both loosely based on Gregory Maguire's 1995 novel *Wicked: The Life and Times of the Wicked Witch of the West*, which offers an alternative backstory for the Wicked Witch from *The Wizard of Oz*. This film focuses on *Wicked*'s early years and two young witches-to-be: green-skinned outcast Elphaba (Erivo), who will go on to become the Wicked Witch of the West, and the vain, popular Galinda (Grande), who will eventually blossom into Glinda the Good.

Both Elphaba and Galinda are newly arrived at Shiz University (think the student politics of *Mean Girls*'s North Shore High and the curriculum of Hogwarts). Although not enrolled as a student, Elphaba is on site to help her paraplegic younger sister, Nessarose (Marissa Bode). But the formidable teacher Madame Morrible (a haughtily fabulous Michelle Yeoh, breezing over any vocal limitations in a cloud of intimidating glamour) spots potential in Elphaba and offers her one-to-one tutoring in the art of enchantment. To the mutual disgust of both girls, Elphaba and Galinda find themselves assigned as roomies.

It's not just their personalities that clash: the film's colour palette is at first a battleground between the chlorophyll green of Elphaba's skin and the candyfloss pink of Galinda's wardrobe. But visuals that initially seem jarring start to find harmony as the movie progresses. A scene in a forest full of mossy tufts of vegetation and garlands of delicate, rosy blooms is lush and lovely, one of several notable triumphs for the production design department. Jed by Nathan Crowley, whose credits include the similarly lavish *Wonka*. Likewise, Elphaba and Galinda warm to each other and a genuine connection is forged between them.

Both lead actors impress. Erivo is terrific, her rich, velvety voice cracking under the weight of rejections and ridicule suffered by Elphaba; her eyes showing the bruises that her skin cannot. And Grande is supremely well cast. It's not just the voice: the singer has a vocal range so extensive that some of it is only audible to bats, and she uses every last note of it here. But more crucial is her gift for physical comedy - each flouncy hair toss, each ditsy heel kick, is a precision-tooled punchline.

Elsewhere, *Bridgerton*'s Jonathan Bailey, as the shallow and self-absorbed Prince Fiyero, skips away with every one of his scenes - in particular, a dizzyingly complex song-and-dance sequence in the college library. Kudos, too, to the choreographer Christopher Scott for dreaming it all up, and to cinematographer Alice Brooks for capturing the magic.

Does it all work? There are moments that get too caught up in their own whirl of CGI pageantry and empty spectacle. And certainly some scenes could be tightened up a little - it's worth noting that the running time of this first film instalment is longer than the stage version in its entirety. But for the most part, *Wicked* the movie takes flight and lifts our hearts along with it. We're caught in the slipstream of Elphaba and her knobby and uncomfortable-looking broomstick as she whooshes off into the second half of the story.

Document level

In the end, they gave the movie 4/5 stars



<https://www.theguardian.com/film/2024/nov/24/wicked-the-film-review-cynthia-erivo-and-ariana-grande-jon-m-chu-part-one-1>

There's some kind of magic afoot. If, like me, you're one of the very few people who haven't already seen the blockbuster stage musical *Wicked* (it's the second-highest-grossing Broadway show of all time, so that's an awful lot of bums on seats), you may approach this shiny, high-energy, relentlessly marketed movie adaptation with low to moderate expectations. There's the unwieldy running time, for a start - two hours and 40 minutes - and the cynical, box-office-gouging decision to carve the story into two films (fans will have to wait almost a year to the day before they get to watch the concluding chapter).

But here's the thing: reservations are soon extinguished and grumbles about the release strategy swiftly quashed. *Wicked* matches its polished razzle-dazzle with real heart. Driven by knockout performances from Cynthia Erivo and Ariana Grande, Jon M Chu's impossibly slick charm assault of an adaptation zips along so enjoyably that you almost wish it were longer (your bladder may disagree). With its all too timely themes of bullying, corrupt leaders and the demonisation of difference, this is a movie that promises a froth of pink and green escapism but delivers considerably more in the way of depth and darkness.

For those who have somehow evaded *Wicked*'s cultural reach over the past couple of decades, here's a brief primer. The film and the stage show are both loosely based on Gregory Maguire's 1995 novel *Wicked: The Life and Times of the Wicked Witch of the West*, which offers an alternative backstory for the Wicked Witch from *The Wizard of Oz*. This film focuses on *Wicked*'s early years and two young witches-to-be: green-skinned outcast Elphaba (Erivo), who will go on to become the Wicked Witch of the West, and the vain, popular Galinda (Grande), who will eventually blossom into Glinda the Good.

Both Elphaba and Galinda are newly arrived at Shiz University (think the student politics of *Mean Girls*'s North Shore High and the curriculum of Hogwarts). Although not enrolled as a student, Elphaba is on site to help her paraplegic younger sister, Nessarose (Marissa Bode). But the formidable teacher Madame Morrible (a haughtily fabulous Michelle Yeoh, breezing over any vocal limitations in a cloud of intimidating glamour) spots potential in Elphaba and offers her one-to-one tutoring in the art of enchantment. To the mutual disgust of both girls, Elphaba and Galinda find themselves assigned as roomies.

It's not just their personalities that clash: the film's colour palette is at first a battleground between the chlorophyll green of Elphaba's skin and the candyfloss pink of Galinda's wardrobe. But visuals that initially seem jarring start to find harmony as the movie progresses. A scene in a forest full of mossy tufts of vegetation and garlands of delicate, rosy blooms is lush and lovely, one of several notable triumphs for the production design department. Led by Nathan Crowley, whose credits include the similarly lavish *Wonka*. Likewise, Elphaba and Galinda warm to each other and a genuine connection is forged between them.

Both lead actors impress. Erivo is terrific, her rich, velvety voice cracking under the weight of rejections and ridicule suffered by Elphaba; her eyes showing the bruises that her skin cannot. And Grande is supremely well cast. It's not just the voice: the singer has a vocal range so extensive that some of it is only audible to bats, and she uses every last note of it here. But more crucial is her gift for physical comedy - each flouncy hair toss, each ditsy heel kick, is a precision-tooled punchline.

Elsewhere, *Bridgerton*'s Jonathan Bailey, as the shallow and self-absorbed Prince Fiyero, skips away with every one of his scenes - in particular, a dizzyingly complex song-and-dance sequence in the college library. Kudos, too, to the choreographer Christopher Scott for dreaming it all up, and to cinematographer Alice Brooks for capturing the magic.

Does it all work? There are moments that get too caught up in their own whirl of CGI pageantry and empty spectacle. And certainly some scenes could be tightened up a little - it's worth noting that the running time of this first film instalment is longer than the stage version in its entirety. But for the most part, *Wicked* the movie takes flight and lifts our hearts along with it. We're caught in the slipstream of Elphaba and her knobby and uncomfortable-looking broomstick as she whooshes off into the second half of the story.

Paragraph level

The summarizing final paragraph is mostly positive

Does it all work? There are moments that get too caught up in their own whirl of CGI pageantry and empty spectacle. And certainly some scenes could be tightened up a little - it's worth noting that the running time of this first film instalment is longer than the stage version in its entirety. But for the most part, *Wicked* the movie takes flight and lifts our hearts along with it. We're caught in the slipstream of Elphaba and her knobbly and uncomfortable-looking broomstick as she whooshes off into the second half of the story.

<https://www.theguardian.com/film/2024/nov/24/wicked-the-film-review-cynthia-erivo-and-ariana-grande-jon-m-chu-part-one-1>

There's some kind of magic afoot. If, like me, you're one of the very few people who haven't already seen the blockbuster stage musical *Wicked* (it's the second-highest-grossing Broadway show of all time, so that's an awful lot of bums on seats), you may approach this shiny, high-energy, relentlessly marketed movie adaptation with low to moderate expectations. There's the unwieldy running time, for a start - two hours and 40 minutes - and the cynical, box-office-gouging decision to carve the story into two films (fans will have to wait almost a year to the day before they get to watch the concluding chapter).

But here's the thing: reservations are soon extinguished and grumbles about the release strategy swiftly quashed. *Wicked* matches its polished razzle-dazzle with real heart. Driven by knockout performances from Cynthia Erivo and Ariana Grande, Jon M Chu's impossibly slick charm assault of an adaptation zips along so enjoyably that you almost wish it were longer (your bladder may disagree). With its all too timely themes of bullying, corrupt leaders and the demonisation of difference, this is a movie that promises a froth of pink and green escapism but delivers considerably more in the way of depth and darkness.

For those who have somehow evaded *Wicked*'s cultural reach over the past couple of decades, here's a brief primer. The film and the stage show are both loosely based on Gregory Maguire's 1995 novel *Wicked: The Life and Times of the Wicked Witch of the West*, which offers an alternative backstory for the Wicked Witch from *The Wizard of Oz*. This film focuses on *Wicked*'s early years and two young witches-to-be: green-skinned outcast Elphaba (Erivo), who will go on to become the Wicked Witch of the West, and the vain, popular Galinda (Grande), who will eventually blossom into Glinda the Good.

Both Elphaba and Galinda are newly arrived at Shiz University (think the student politics of *Mean Girls*'s North Shore High and the curriculum of Hogwarts). Although not enrolled as a student, Elphaba is on site to help her paraplegic younger sister, Nessarose (Marissa Bode). But the formidable teacher Madame Morrible (a haughtily fabulous Michelle Yeoh, breezing over any vocal limitations in a cloud of intimidating glamour) spots potential in Elphaba and offers her one-to-one tutoring in the art of enchantment. To the mutual disgust of both girls, Elphaba and Galinda find themselves assigned as roomies.

It's not just their personalities that clash: the film's colour palette is at first a battleground between the chlorophyll green of Elphaba's skin and the candyfloss pink of Galinda's wardrobe. But visuals that initially seem jarring start to find harmony as the movie progresses. A scene in a forest full of mossy tufts of vegetation and garlands of delicate, rosy blooms is lush and lovely, one of several notable triumphs for the production design department. Led by Nathan Crowley, whose credits include the similarly lavish *Wonka*. Likewise, Elphaba and Galinda warm to each other and a genuine connection is forged between them.

Both lead actors impress. Erivo is terrific, her rich, velvety voice cracking under the weight of rejections and ridicule suffered by Elphaba; her eyes showing the bruises that her skin cannot. And Grande is supremely well cast. It's not just the voice: the singer has a vocal range so extensive that some of it is only audible to bats, and she uses every last note of it here. But more crucial is her gift for physical comedy - each flouncy hair toss, each ditsy heel kick, is a precision-tooled punchline.

Elsewhere, *Bridgerton*'s Jonathan Bailey, as the shallow and self-absorbed Prince Fiyero, skips away with every one of his scenes - in particular, a dizzyingly complex song-and-dance sequence in the college library. Kudos, too, to the choreographer Christopher Scott for dreaming it all up, and to cinematographer Alice Brooks for capturing the magic.

Does it all work? There are moments that get too caught up in their own whirl of CGI pageantry and empty spectacle. And certainly some scenes could be tightened up a little - it's worth noting that the running time of this first film instalment is longer than the stage version in its entirety. But for the most part, *Wicked* the movie takes flight and lifts our hearts along with it. We're caught in the slipstream of Elphaba and her knobbly and uncomfortable-looking broomstick as she whooshes off into the second half of the story.

Paragraph level

But the opening paragraph is quite critical!

There's some kind of magic afoot. If, like me, you're one of the very few people who hasn't already seen the blockbuster stage musical *Wicked* (it's the second-highest-grossing Broadway show of all time, so that's an awful lot of bums on seats), you may approach this shiny, high-energy, relentlessly marketed movie adaptation with low to moderate expectations. There's the unwieldy running time, for a start - two hours and 40 minutes - and the cynical, box-office-gouging decision to carve the story into two films (fans will have to wait almost a year to the day before they get to watch the concluding chapter).

<https://www.theguardian.com/film/2024/nov/24/wicked-the-film-review-cynthia-erivo-and-ariana-grande-jon-m-chu-part-one-1>

There's some kind of magic afoot. If, like me, you're one of the very few people who hasn't already seen the blockbuster stage musical *Wicked* (it's the second-highest-grossing Broadway show of all time, so that's an awful lot of bums on seats), you may approach this shiny, high-energy, relentlessly marketed movie adaptation with low to moderate expectations. There's the unwieldy running time, for a start - two hours and 40 minutes - and the cynical, box-office-gouging decision to carve the story into two films (fans will have to wait almost a year to the day before they get to watch the concluding chapter).

But here's the thing: reservations are soon extinguished and grumbles about the release strategy swiftly quashed. *Wicked* matches its polished razzle-dazzle with real heart. Driven by knockout performances from Cynthia Erivo and Ariana Grande, Jon M Chu's impossibly slick charm assault of an adaptation zips along so enjoyably that you almost wish it were longer (your bladder may disagree). With its all too timely themes of bullying, corrupt leaders and the demonisation of difference, this is a movie that promises a froth of pink and green escapism but delivers considerably more in the way of depth and darkness.

For those who have somehow evaded *Wicked*'s cultural reach over the past couple of decades, here's a brief primer. The film and the stage show are both loosely based on Gregory Maguire's 1995 novel *Wicked: The Life and Times of the Wicked Witch of the West*, which offers an alternative backstory for the Wicked Witch from *The Wizard of Oz*. This film focuses on *Wicked*'s early years and two young witches-to-be: green-skinned outcast Elphaba (Erivo), who will go on to become the Wicked Witch of the West, and the vain, popular Galinda (Grande), who will eventually blossom into Glinda the Good.

Both Elphaba and Galinda are newly arrived at Shiz University (think the student politics of *Mean Girls*'s North Shore High and the curriculum of Hogwarts). Although not enrolled as a student, Elphaba is on site to help her paraplegic younger sister, Nessarose (Marissa Bode). But the formidable teacher Madame Morrible (a haughtily fabulous Michelle Yeoh, breezing over any vocal limitations in a cloud of intimidating glamour) spots potential in Elphaba and offers her one-to-one tutoring in the art of enchantment. To the mutual disgust of both girls, Elphaba and Galinda find themselves assigned as roomies.

It's not just their personalities that clash: the film's colour palette is at first a battleground between the chlorophyll green of Elphaba's skin and the candyfloss pink of Galinda's wardrobe. But visuals that initially seem jarring start to find harmony as the movie progresses. A scene in a forest full of mossy tufts of vegetation and garlands of delicate, rosy blooms is lush and lovely, one of several notable triumphs for the production design department. Led by Nathan Crowley, whose credits include the similarly lavish *Wonka*. Likewise, Elphaba and Galinda warm to each other and a genuine connection is forged between them.

Both lead actors impress. Erivo is terrific, her rich, velvety voice cracking under the weight of rejections and ridicule suffered by Elphaba; her eyes showing the bruises that her skin cannot. And Grande is supremely well cast. It's not just the voice: the singer has a vocal range so extensive that some of it is only audible to bats, and she uses every last note of it here. But more crucial is her gift for physical comedy - each flouncy hair toss, each ditsy heel kick, is a precision-tooled punchline.

Elsewhere, *Bridgerton*'s Jonathan Bailey, as the shallow and self-absorbed Prince Fiyero, skips away with every one of his scenes - in particular, a dizzyingly complex song-and-dance sequence in the college library. Kudos, too, to the choreographer Christopher Scott for dreaming it all up, and to cinematographer Alice Brooks for capturing the magic.

Does it all work? There are moments that get too caught up in their own whirl of CGI pageantry and empty spectacle. And certainly some scenes could be tightened up a little - it's worth noting that the running time of this first film instalment is longer than the stage version in its entirety. But for the most part, *Wicked* the movie takes flight and lifts our hearts along with it. We're caught in the slipstream of Elphaba and her knobby and uncomfortable-looking broomstick as she whooshes off into the second half of the story.

Sentence level

This sentence clearly expresses
positive sentiment

Both lead actors impress.

<https://www.theguardian.com/film/2024/nov/24/wicked-the-film-review-cynthia-erivo-and-ariana-grande-jon-m-chu-part-one-1>

There's some kind of magic afoot. If, like me, you're one of the very few people who haven't already seen the blockbuster stage musical *Wicked* (it's the second-highest-grossing Broadway show of all time, so that's an awful lot of bums on seats), you may approach this shiny, high-energy, relentlessly marketed movie adaptation with low to moderate expectations. There's the unwieldy running time, for a start - two hours and 40 minutes - and the cynical, box-office-gouging decision to carve the story into two films (fans will have to wait almost a year to the day before they get to watch the concluding chapter).

But here's the thing: reservations are soon extinguished and grumbles about the release strategy swiftly quashed. *Wicked* matches its polished razzle-dazzle with real heart. Driven by knockout performances from Cynthia Erivo and Ariana Grande, Jon M Chu's impossibly slick charm assault of an adaptation zips along so enjoyably that you almost wish it were longer (your bladder may disagree). With its all too timely themes of bullying, corrupt leaders and the demonisation of difference, this is a movie that promises a froth of pink and green escapism but delivers considerably more in the way of depth and darkness.

For those who have somehow evaded *Wicked*'s cultural reach over the past couple of decades, here's a brief primer. The film and the stage show are both loosely based on Gregory Maguire's 1995 novel *Wicked: The Life and Times of the Wicked Witch of the West*, which offers an alternative backstory for the Wicked Witch from *The Wizard of Oz*. This film focuses on *Wicked*'s early years and two young witches-to-be: green-skinned outcast Elphaba (Erivo), who will go on to become the Wicked Witch of the West, and the vain, popular Galinda (Grande), who will eventually blossom into Glinda the Good.

Both Elphaba and Galinda are newly arrived at Shiz University (think the student politics of *Mean Girls*'s North Shore High and the curriculum of Hogwarts). Although not enrolled as a student, Elphaba is on site to help her paraplegic younger sister, Nessarose (Marissa Bode). But the formidable teacher Madame Morrible (a haughtily fabulous Michelle Yeoh, breezing over any vocal limitations in a cloud of intimidating glamour) spots potential in Elphaba and offers her one-to-one tutoring in the art of enchantment. To the mutual disgust of both girls, Elphaba and Galinda find themselves assigned as roomies.

It's not just their personalities that clash: the film's colour palette is at first a battleground between the chlorophyll green of Elphaba's skin and the candyfloss pink of Galinda's wardrobe. But visuals that initially seem jarring start to find harmony as the movie progresses. A scene in a forest full of mossy tufts of vegetation and garlands of delicate, rosy blooms is lush and lovely, one of several notable triumphs for the production design department. Jed by Nathan Crowley, whose credits include the similarly lavish *Wonka*. Likewise, Elphaba and Galinda warm to each other and a genuine connection is forged between them.

Both lead actors impress. Erivo is terrific, her rich, velvety voice cracking under the weight of rejections and ridicule suffered by Elphaba; her eyes showing the bruises that her skin cannot. And Grande is supremely well cast. It's not just the voice: the singer has a vocal range so extensive that some of it is only audible to bats, and she uses every last note of it here. But more crucial is her gift for physical comedy - each flouncy hair toss, each ditsy heel kick, is a precision-tooled punchline.

Elsewhere, *Bridgerton*'s Jonathan Bailey, as the shallow and self-absorbed Prince Fiyero, skips away with every one of his scenes - in particular, a dizzyingly complex song-and-dance sequence in the college library. Kudos, too, to the choreographer Christopher Scott for dreaming it all up, and to cinematographer Alice Brooks for capturing the magic.

Does it all work? There are moments that get too caught up in their own whirl of CGI pageantry and empty spectacle. And certainly some scenes could be tightened up a little - it's worth noting that the running time of this first film instalment is longer than the stage version in its entirety. But for the most part, *Wicked* the movie takes flight and lifts our hearts along with it. We're caught in the slipstream of Elphaba and her knobby and uncomfortable-looking broomstick as she whooshes off into the second half of the story.

Aspect level

In this part, they praise Ariana Grande's vocal range

It's not just the voice: the singer has a vocal range so extensive that some of it is only audible to bats, and she uses every last note of it here.

There's some kind of magic afoot. If, like me, you're one of the very few people who haven't already seen the blockbuster stage musical *Wicked* (it's the second-highest-grossing Broadway show of all time, so that's an awful lot of bums on seats), you may approach this shiny, high-energy, relentlessly marketed movie adaptation with low to moderate expectations. There's the unwieldy running time, for a start - two hours and 40 minutes - and the cynical, box-office-gouging decision to carve the story into two films (fans will have to wait almost a year to the day before they get to watch the concluding chapter).

But here's the thing: reservations are soon extinguished and grumbles about the release strategy swiftly quashed. *Wicked* matches its polished razzle-dazzle with real heart. Driven by knockout performances from Cynthia Erivo and Ariana Grande, Jon M Chu's impossibly slick charm assault of an adaptation zips along so enjoyably that you almost wish it were longer (your bladder may disagree). With its all too timely themes of bullying, corrupt leaders and the demonisation of difference, this is a movie that promises a froth of pink and green escapism but delivers considerably more in the way of depth and darkness.

For those who have somehow evaded *Wicked*'s cultural reach over the past couple of decades, here's a brief primer. The film and the stage show are both loosely based on Gregory Maguire's 1995 novel *Wicked: The Life and Times of the Wicked Witch of the West*, which offers an alternative backstory for the Wicked Witch from *The Wizard of Oz*. This film focuses on *Wicked*'s early years and two young witches-to-be: green-skinned outcast Elphaba (Erivo), who will go on to become the Wicked Witch of the West, and the vain, popular Galinda (Grande), who will eventually blossom into Glinda the Good.

Both Elphaba and Galinda are newly arrived at Shiz University (think the student politics of *Mean Girls*'s North Shore High and the curriculum of Hogwarts). Although not enrolled as a student, Elphaba is on site to help her paraplegic younger sister, Nessarose (Marissa Bode). But the formidable teacher Madame Morrible (a haughtily fabulous Michelle Yeoh, breezing over any vocal limitations in a cloud of intimidating glamour) spots potential in Elphaba and offers her one-to-one tutoring in the art of enchantment. To the mutual disgust of both girls, Elphaba and Galinda find themselves assigned as roomies.

It's not just their personalities that clash: the film's colour palette is at first a battleground between the chlorophyll green of Elphaba's skin and the candyfloss pink of Galinda's wardrobe. But visuals that initially seem jarring start to find harmony as the movie progresses. A scene in a forest full of mossy tufts of vegetation and garlands of delicate, rosy blooms is lush and lovely, one of several notable triumphs for the production design department. Jed by Nathan Crowley, whose credits include the similarly lavish *Wonka*. Likewise, Elphaba and Galinda warm to each other and a genuine connection is forged between them.

Both lead actors impress. Erivo is terrific, her rich, velvety voice cracking under the weight of rejections and ridicule suffered by Elphaba; her eyes showing the bruises that her skin cannot. And Grande is supremely well cast.

It's not just the voice: the singer has a vocal range so extensive that some of it is only audible to bats, and she uses every last note of it here. But more crucial is her gift for physical comedy - each flouncy hair toss, each ditsy heel kick, is a precision-tooled punchline.

Elsewhere, *Bridgerton*'s Jonathan Bailey, as the shallow and self-absorbed Prince Fiyero, skips away with every one of his scenes - in particular, a dizzyingly complex song-and-dance sequence in the college library. Kudos, too, to the choreographer Christopher Scott for dreaming it all up, and to cinematographer Alice Brooks for capturing the magic.

Does it all work? There are moments that get too caught up in their own whirl of CGI pageantry and empty spectacle. And certainly some scenes could be tightened up a little - it's worth noting that the running time of this first film instalment is longer than the stage version in its entirety. But for the most part, *Wicked* the movie takes flight and lifts our hearts along with it. We're caught in the slipstream of Elphaba and her knobby and uncomfortable-looking broomstick as she whooshes off into the second half of the story.

<https://www.theguardian.com/film/2024/nov/24/wicked-the-film-review-cynthia-erivo-and-ariana-grande-jon-m-chu-part-one-1>

Another example: booking hotels



SEK





List your property

Register

Sign in

Stays

Flights

Flight + Hotel

Car rentals

Attractions

Airport taxis



Samarkand





Fri, Mar 14 — Sun, Mar 23

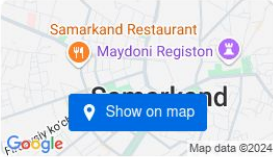


2 adults · 0 children · 1 room



Search

[Home](#) > [Uzbekistan](#) > [Samarkand](#) > Search results



Samarkand Restaurant

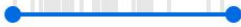
Maydoni Registon

Show on map

Filter by:

Your budget (per night)

SEK 200 – SEK 3,000+



Most used filters in Samarkand

☐ Free Wifi 35

☐ Wonderful: 9+ Based on guest reviews 22

☐ Apartments 5

Samarkand: 36 properties found

List

Grid

Sort by: Top picks for long stays

Very Good: 8+

Private bathroom

Less than 1 km

Based on your filters, 93% of places to stay are unavailable on our site.

Commission paid on bookings and other factors can affect property rankings. Learn about these ranking parameters and how to select and modify them. [Learn more](#)



Cinzano Complex

Samarkand

Show on map

0.5 km from downtown

Double Room

1 full bed

Free cancellation

No prepayment needed – pay at the property

Only 2 rooms left at this price on our site

Wonderful 9.1

229 reviews

Comfort 9.3

9 nights, 2 adults

SEK 4,873

Includes taxes and fees

See availability

20

Strongly aggregated opinion

Wonderful
229 reviews

9.1

Strongly aggregated opinion

Wonderful
229 reviews

9.1

Comfort 9.3

But not so sophisticated...

Tell us a little more about the aspects you experienced during your stay (optional)

Staff



Facilities



Cleanliness



Comfort



Value



Location



Bonus Question – How satisfied were you with your room?



☐ Not Applicable

3. Could you tell us a little more? For example: "room," "breakfast," and "view"

 What did you like?

What went well?

 Long or short, your review helps others.

What if we can't collect this data?

Google reviews

Review summary

Write a review

Add a photo



Rooms · 3,5 ★★★★★

Guests liked the large, clean, updated rooms, though some said maintenance could be improved · Guests liked the beds, though some noted the rooms could be quieter · Rooms had minibars, TVs & kitchenettes

Location · 4,8 ★★★★★

Near public transit & the train station · Near the city center; shopping, sightseeing, restaurants & bars nearby · Easily accessible by car

Service & facilities · 4,7 ★★★★★

Guests appreciated the friendly, professional staff · Guests spoke highly of management & reception · Guests enjoyed the sauna

[View Google reviews](#)

Who is talking about what?

- We can use *Named Entity Recognition (NER)*
- Entities can be: persons, organisations, etc
- Helps us identify what is being discussed!

⚡ Inference API ⓘ

🏷️ Token Classification

⚡ Cold ▾

Examples ▾

I enjoyed the ambience at the Royal Opera but not at the National Museum

Compute

I enjoyed the ambience at the Royal ORG Opera LOC but not at the National Museum LOC

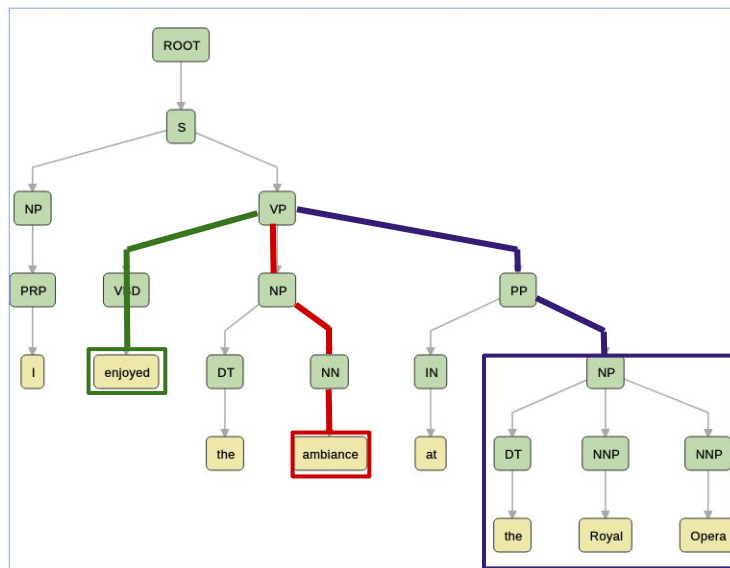
</> View Code ⌚ 0.2s ✓ 0.1s

📄 Minimize

Who is talking about what?

- *Parsing* can link **opinions** to **objects** and **aspects**

Constituency Parse:



Summary

- Opinionated statements have
 - Polarity: negative $\rightarrow$ positive
 - Objectivity: subjective $\rightarrow$ objective
- Opinions – unlike facts – matter most *in aggregate*
- Textual data can contain *many opinions on many levels*
- Opinion mining is when we aggregate opinions
- We can find entities using *named entity recognition (NER)*
- Entities can be linked by, for example, *dependency parsing*

Classifying sentiment

Mining for opinions

What we want



El Grifo Perita - Craft Beer Bar
C. La Fuensanta, 4, 29640 Fuengirola, Málaga, ...

★★★★★ 3 years ago

Very friendly bar 🍺 A lot of really free delicious craft beer, both Spanish and international 🌐 Happy to see some Scandinavian brews 🇳🇴🇩🇪🇸 ...

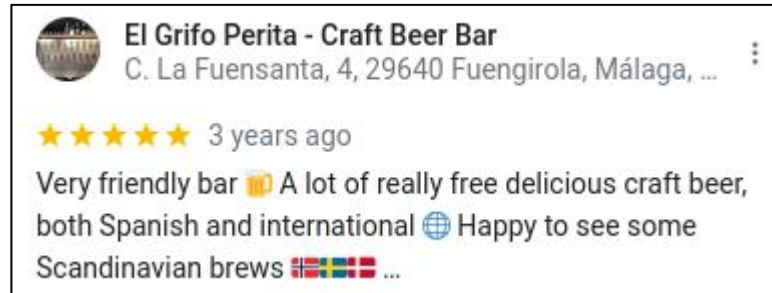
Fuengirola again BUT now we will. We will be back! [More](#)

lovers! [more](#)



Food ★★★★★
Beer ★★★★★
Ambience ★★★★★

Idea: process review one by one



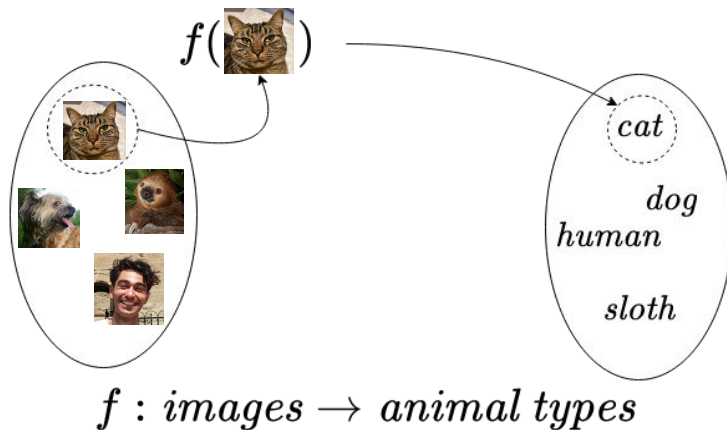
How?

Classification

- Mapping an **input** to a **label**

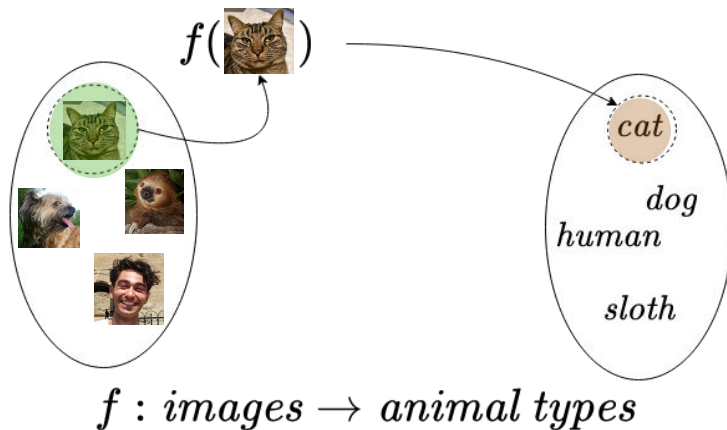
Example: image classification

- Mapping an **input** to a **label**

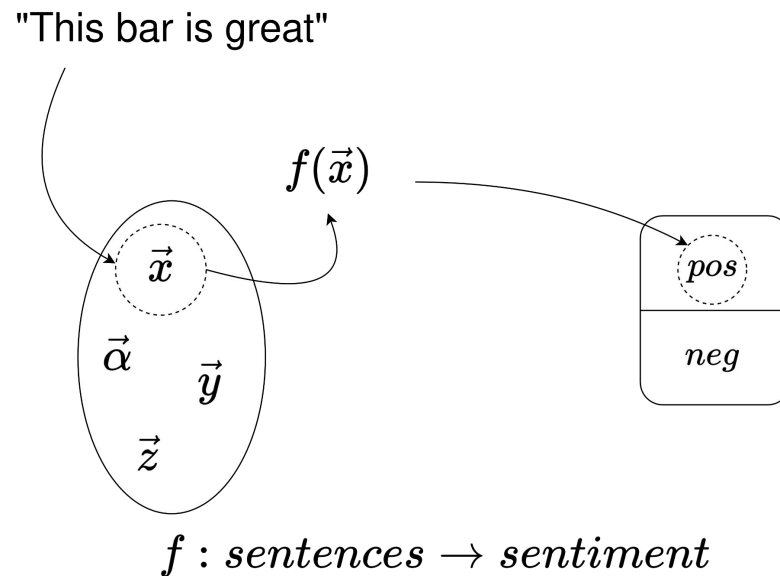


Example: image classification

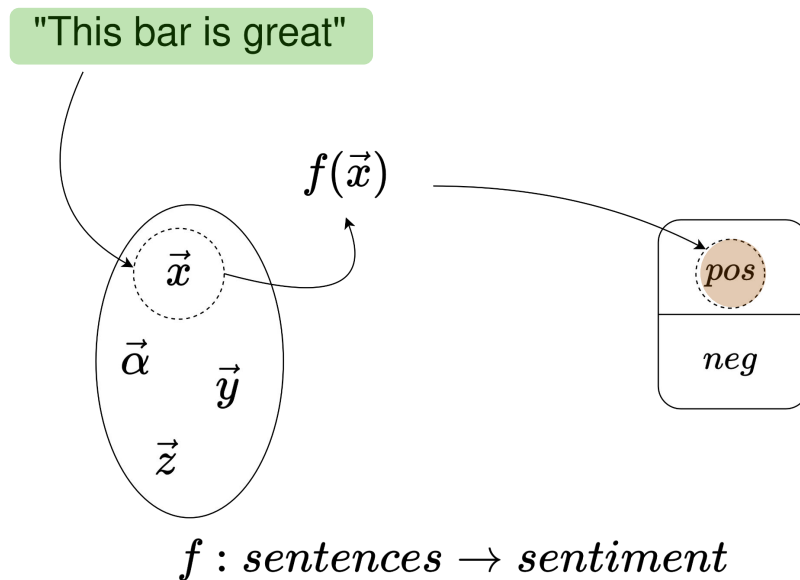
- Mapping an **input** to a **label**



Text classification

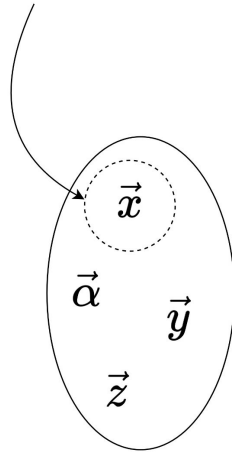


What is the input?



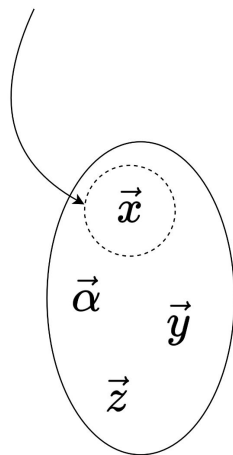
Vectorization

"This bar is great"



Vectorization

"This bar is great"



We talked about
this for *words* in
an earlier lecture

Co-occurrence matrix: vectors

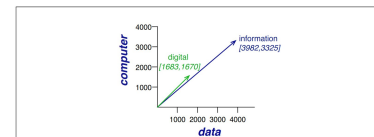


Figure 6.7 A spatial visualization of word vectors for *digital* and *information*, showing just two of the dimensions, corresponding to the words *data* and *computer*.

From [Jurafsky & Martin \(2024\)](#)

47

Word2Vec

- Word2Vec creates vectors that are *learned*
- Vectors are learned using the *skip-gram* algorithm
- Predict a *target word* from its *context*

Target word
All cats use whiskers to sense their surroundings
Context words

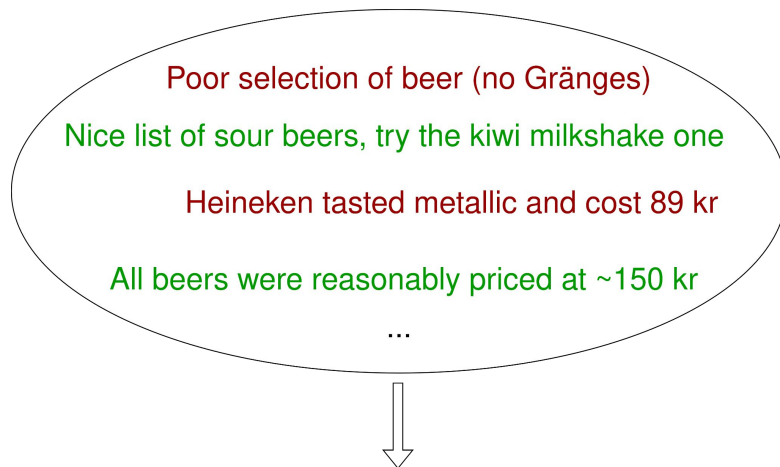
We process the corpus with a *sliding window*.

50

Many ways to vectorize *sequences*

- Word2Vec creates a vector from each word
 - We can average them
- Newer transformer models (F9) build representations
 - Create representations and classify them
- In the lab: tf-idf vectorization
 - *Often works well enough in practice*

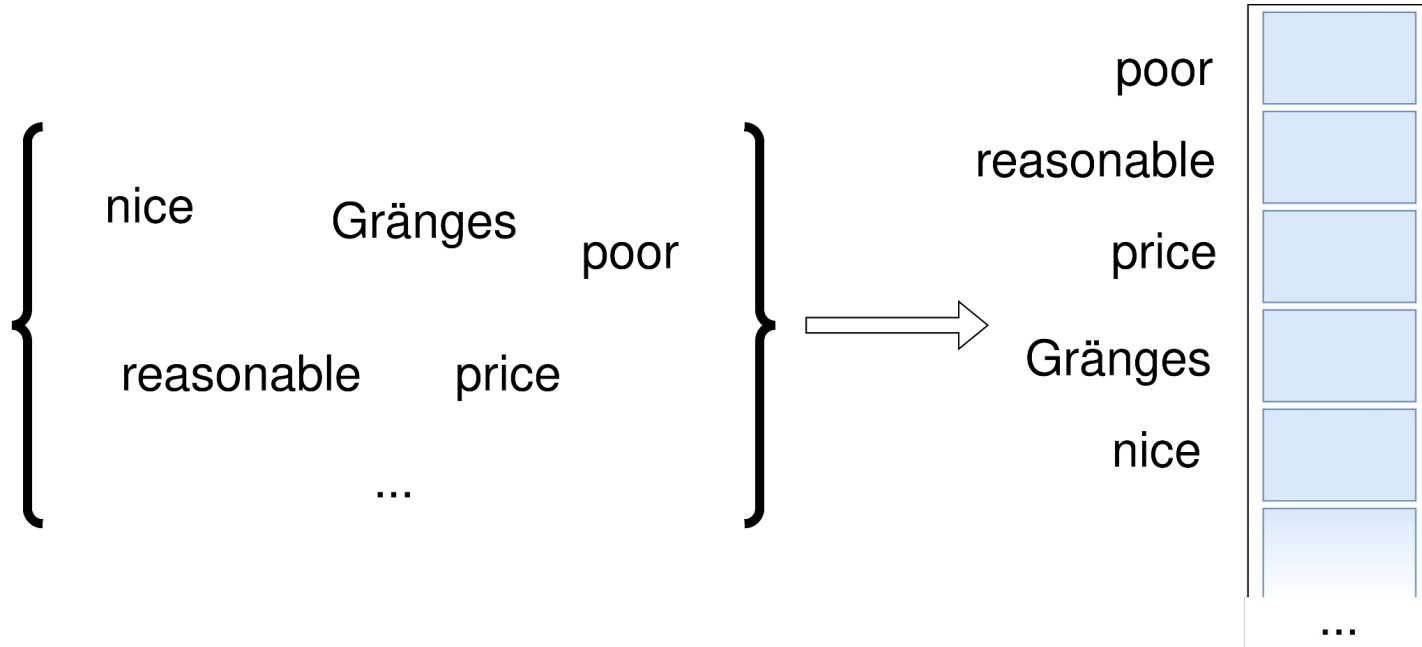
What words do we have?



Reviews for bars

Unique words

Assigning indices



Count vectorization

Poor selection of beer (no Gränges)

Nice list of sour beers, try the kiwi milkshake one

Heineken tasted metallic and cost 89 kr

All beers were reasonably priced at ~150 kr

...

poor

1

reasonable

0

price

0

Gränges

1

nice

0

0

...

Final step

Poor selection of beer (no Gränges)

Nice list of sour beers, try the kiwi milkshake one

Heineken tasted metallic and cost 89 kr

All beers were reasonably priced at ~150 kr

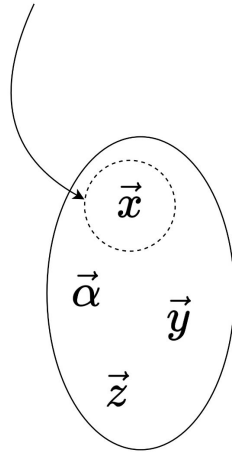
...

With **idf** based on the
training data!

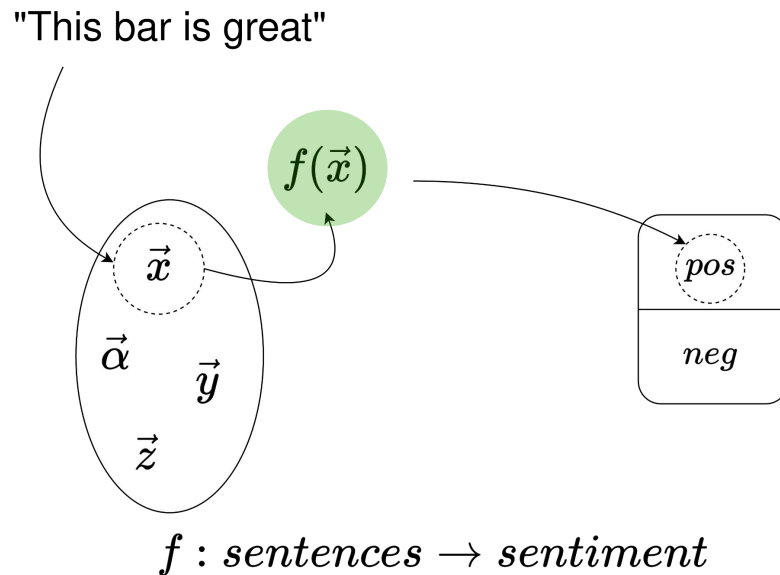
poor	tf-idf(1)
reasonable	tf-idf(0)
price	tf-idf(0)
Gränges	tf-idf(1)
nice	tf-idf(0)
	tf-idf(0)
	...

Now we can vectorize!

"This bar is great"



So how do we classify?



Countless classification techniques

- Fully connected neural network
- Linear classifiers
- In the lab you will try:
 - *Decision trees (DT)*
 - *Support vector machines (SVM)*

Vectorizing our data

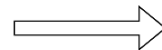
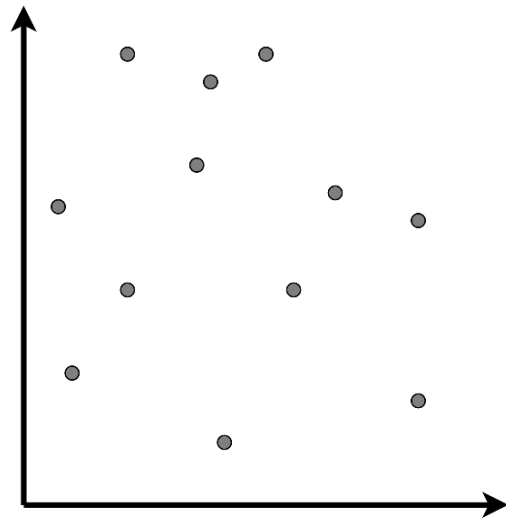
Poor selection of beer (no Gränges)

Nice list of sour beers, try the kiwi milkshake one

Heineken tasted metallic and cost 89 kr

All beers were reasonably priced at ~150 kr

...



We know the classes

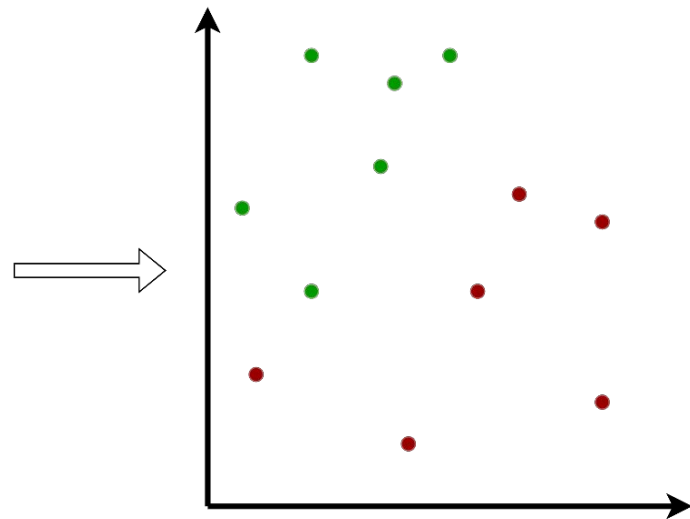
Poor selection of beer (no Gränges)

Nice list of sour beers, try the kiwi milkshake one

Heineken tasted metallic and cost 89 kr

All beers were reasonably priced at ~150 kr

...



SVM: find the boundary

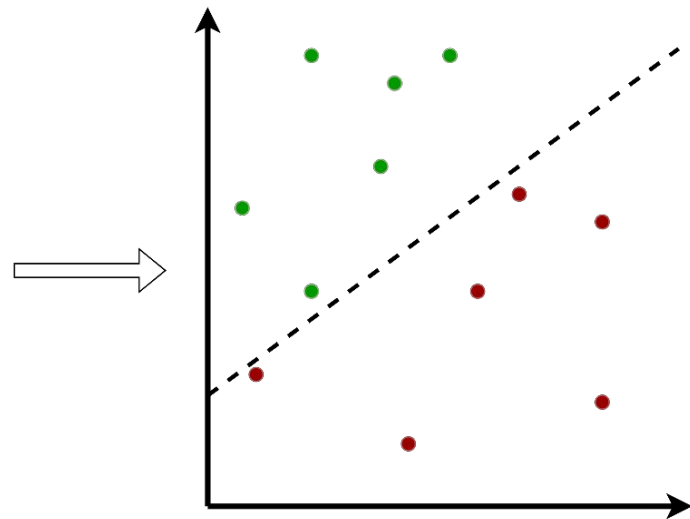
Poor selection of beer (no Gränges)

Nice list of sour beers, try the kiwi milkshake one

Heineken tasted metallic and cost 89 kr

All beers were reasonably priced at ~150 kr

...



We want to *maximize the margin*

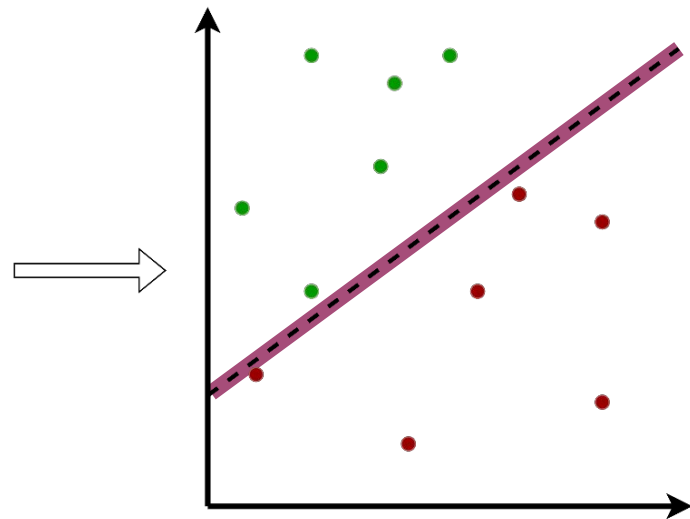
Poor selection of beer (no Gränges)

Nice list of sour beers, try the kiwi milkshake one

Heineken tasted metallic and cost 89 kr

All beers were reasonably priced at ~150 kr

...



There are algorithms to do this efficiently

The *support vectors* in SVM

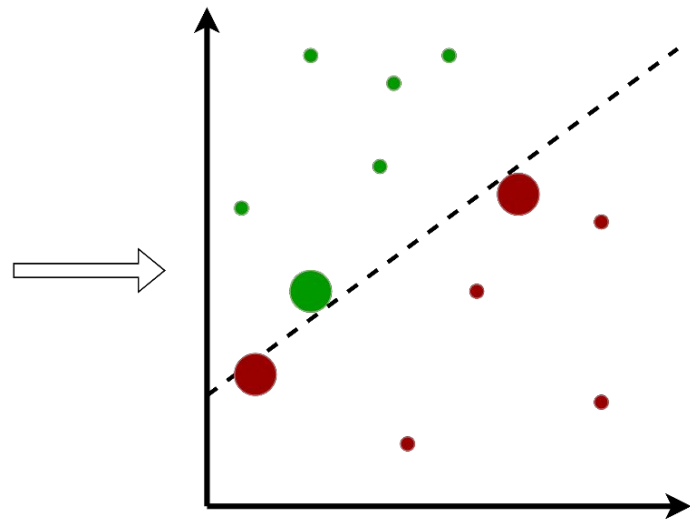
Poor selection of beer (no Gränges)

Nice list of sour beers, try the kiwi milkshake one

Heineken tasted metallic and cost 89 kr

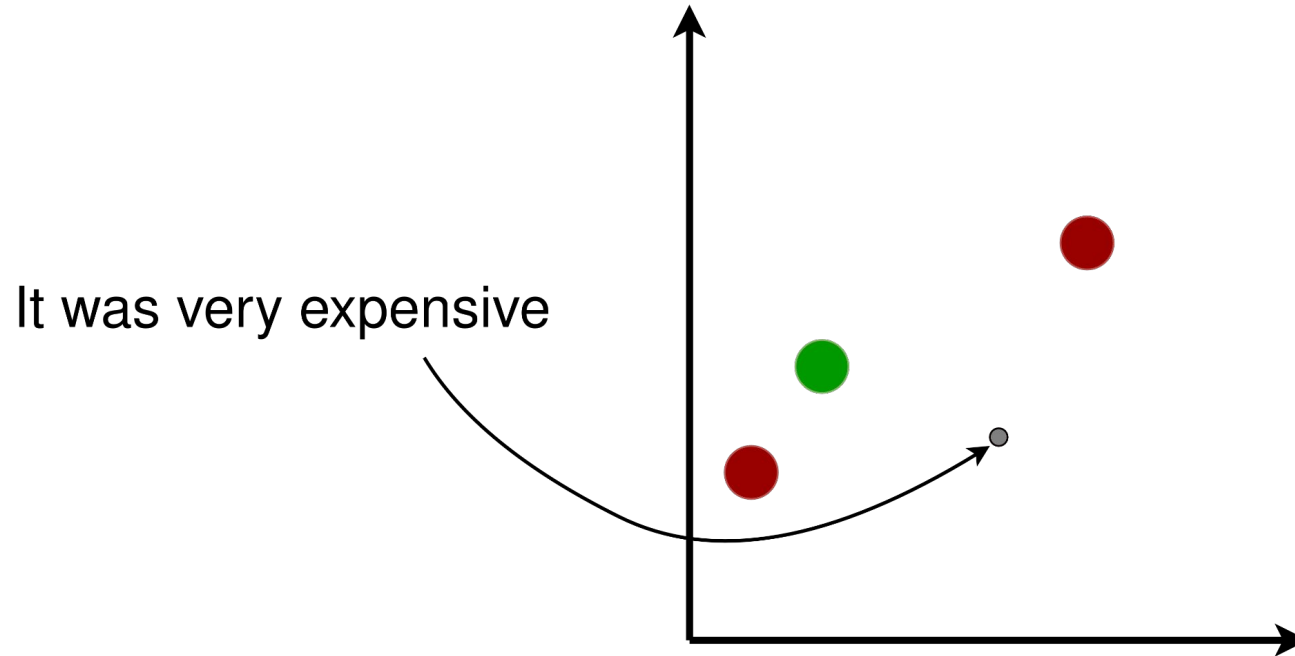
All beers were reasonably priced at ~150 kr

...

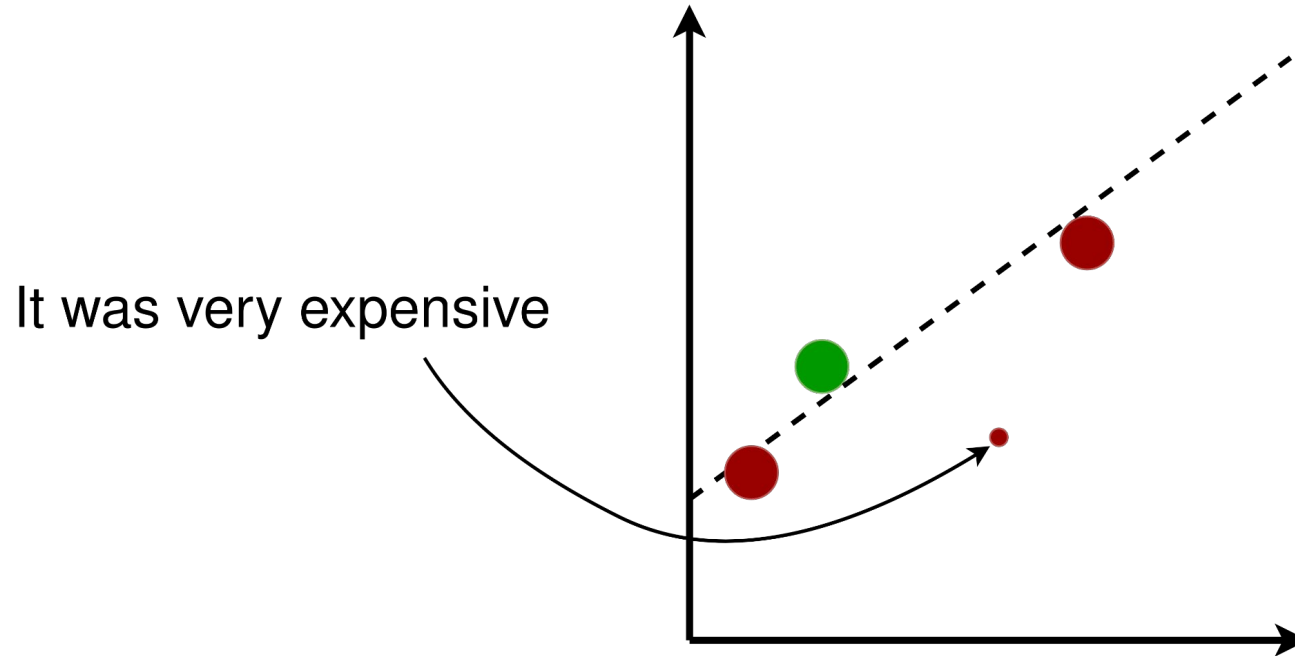


The **support vectors** are the vectors that, if removed, would **change** the **separating hyperplane**

New sample after training

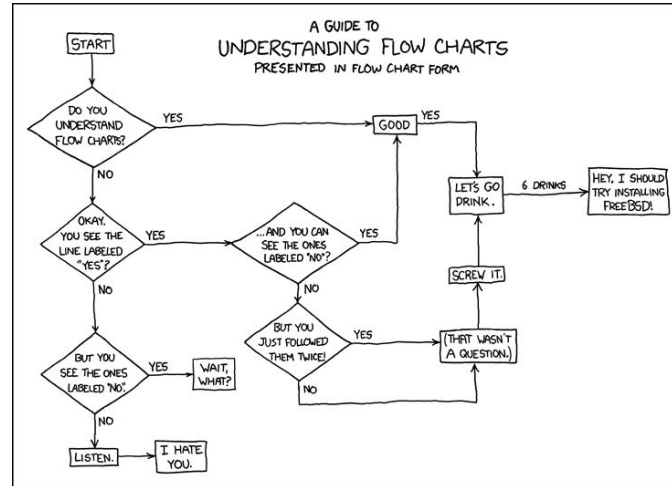


Negative!

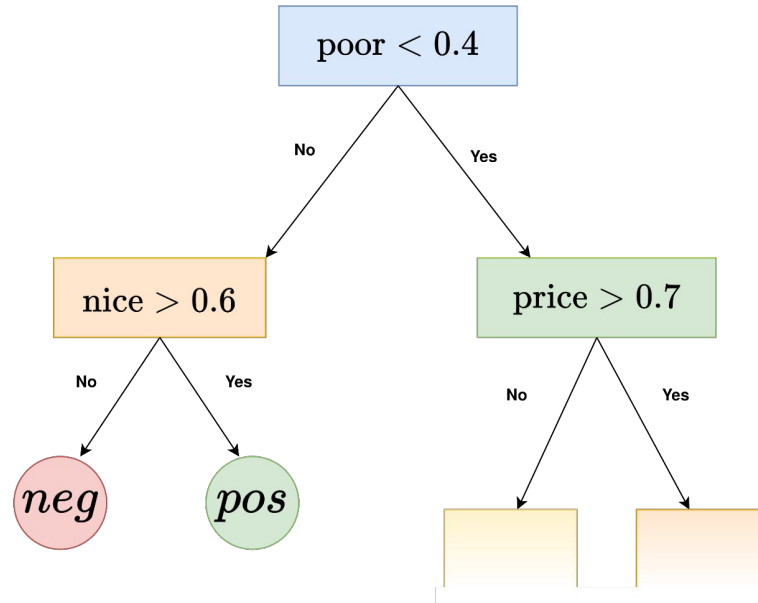


Decision trees

- Decision trees work similarly to flow charts

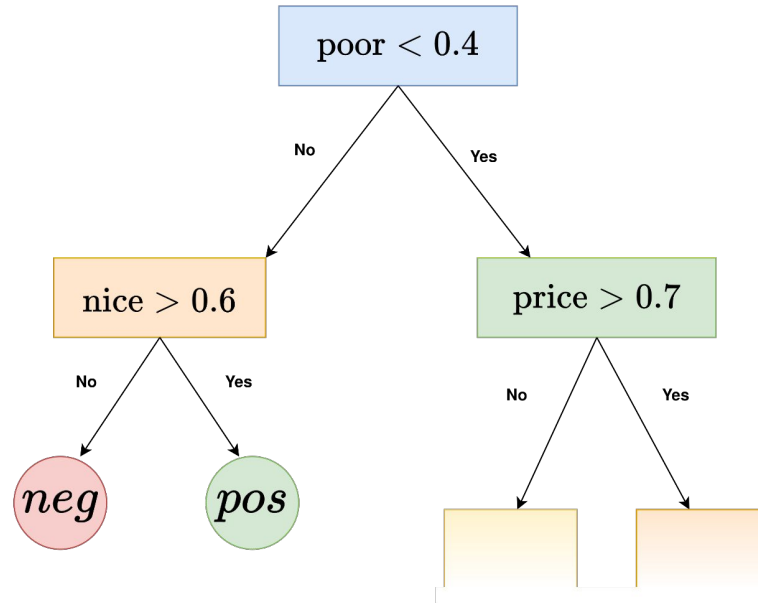


Making decisions



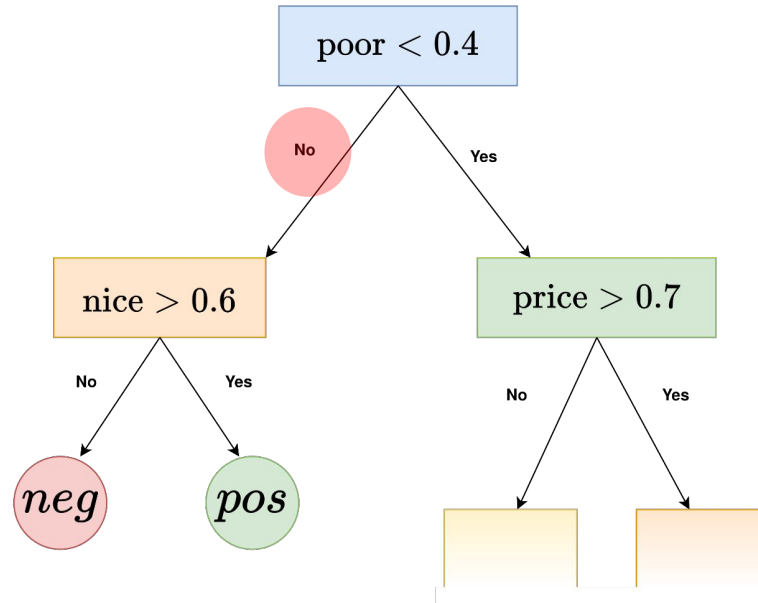
Making decisions

poor	0.5
reasonable	0.2
price	3.3
Gränges	1.5
nice	0.9
...	

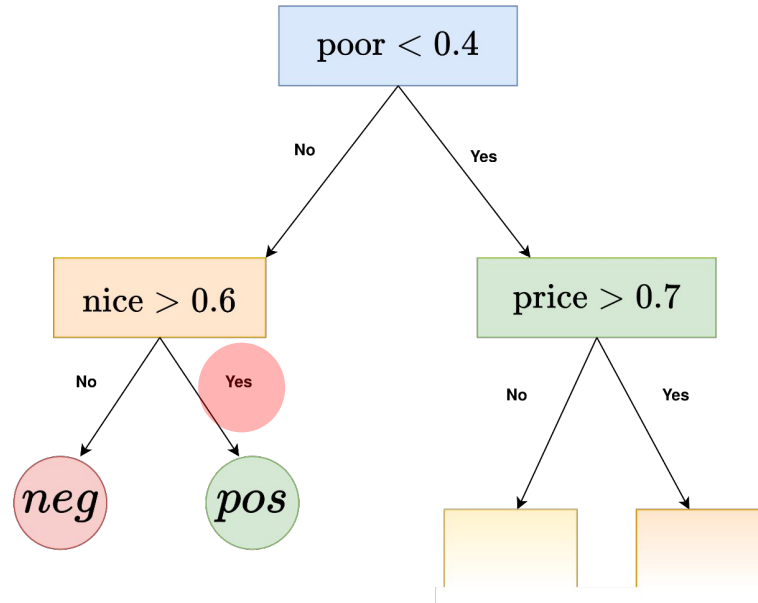


Making decisions

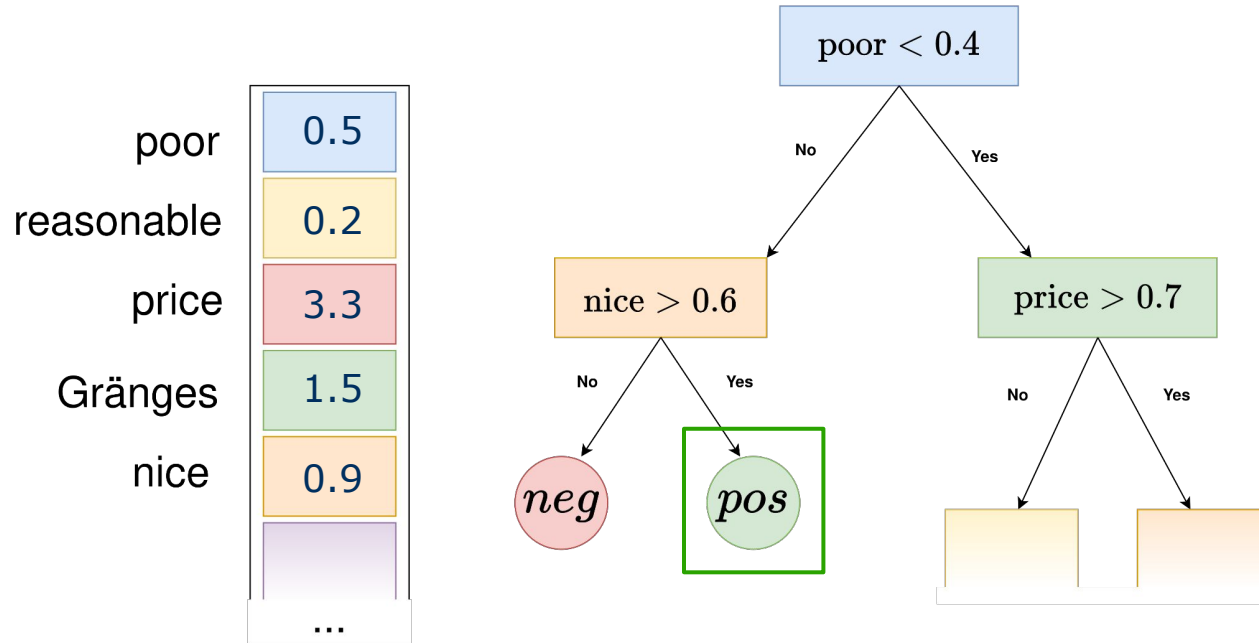
poor	0.5
reasonable	0.2
price	3.3
Gränges	1.5
nice	0.9
...	



Making decisions



Making decisions



The decisions at each node are *learned*!

Summary: classification

- Classifiers deal with *vectorized* data
- Text can be vectorized in a variety of ways
 - In the lab, you'll use a tf-idf vectorizer
 - You can also use pre-trained LLMs (more expensive)
- Classifiers map vectors to *labels*
 - In sentiment analysis: typically negative or positive

Summary: SVMs

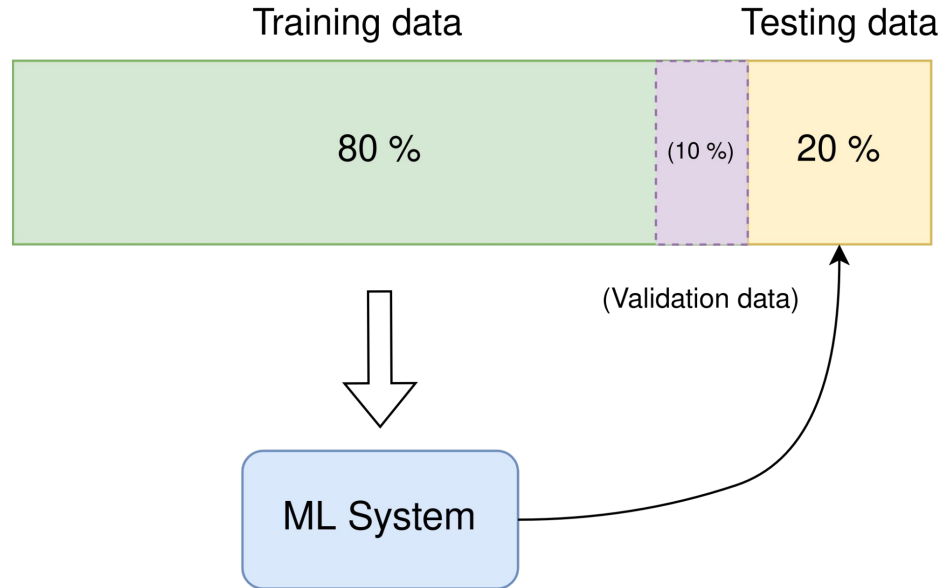
- *Support vector machines (SVMs)* learn a *hyperplane*
 - This hyperplane separates the classes
- The hyperplane is defined by its *support vectors*
- Training algorithms search for the *maximum boundary*

Summary: DTs

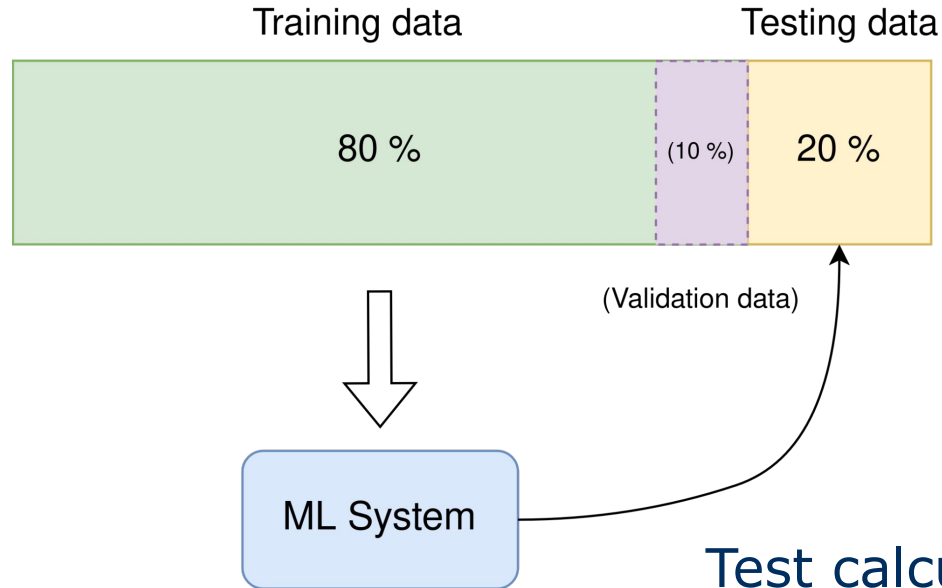
- *Decision trees (DTs)* work like flowcharts
 - They usually learn a series of *binary rules*
 - These rules relate to the different *features*
- New samples are classified using the learned rules

Evaluating classifiers

Training an ML algorithm



Training an ML algorithm



Evaluating classifiers

- We can be wrong in *different ways*
- Accuracy (percentage correct) is imprecise
- Sometimes, our data are imbalanced
- Other times, some errors are riskier

	Predicted <i>pos</i>	Predicted <i>neg</i>
Actually <i>pos</i>	True positive (TP)	False negative (FN)
Actually <i>neg</i>	False positive (FP)	True negative (TN)

Evaluating classifiers

- We can be wrong in *different ways*
- Accuracy (percentage correct) is imprecise
- **Sometimes, our data are imbalanced**
- Other times, some errors are riskier

	Predicted <i>pos</i>	Predicted <i>neg</i>
Actually <i>pos</i>	True positive (TP)	False negative (FN)
Actually <i>neg</i>	False positive (FP)	True negative (TN)

Accuracy with imbalance data

	Predicted <i>pos</i>	Predicted <i>neg</i>
Actually <i>pos</i>	TP = 1	FN = 12
Actually <i>neg</i>	FP = 11	TN = 165

Accuracy with imbalance data

	Predicted <i>pos</i>	Predicted <i>neg</i>
Actually <i>pos</i>	TP = 1	FN = 12
Actually <i>neg</i>	FP = 11	TN = 165

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Accuracy with imbalance data

	Predicted <i>pos</i>	Predicted <i>neg</i>
Actually <i>pos</i>	TP = 1	FN = 12
Actually <i>neg</i>	FP = 11	TN = 165

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{1 + 165}{1 + 165 + 11 + 12} \approx 88\%$$

Accuracy with imbalance data

	Predicted <i>pos</i>	Predicted <i>neg</i>
Actually <i>pos</i>	TP = 1	FN = 12
Actually <i>neg</i>	FP = 11	TN = 165

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{1 + 165}{1 + 165 + 11 + 12} \approx 88\%$$

Accuracy with imbalance data

	Predicted <i>pos</i>	Predicted <i>neg</i>	
Actually <i>pos</i>	TP = 1	FN = 12	= 13
Actually <i>neg</i>	FP = 11	TN = 165	= 176

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{1 + 165}{1 + 165 + 11 + 12} \approx 88\%$$

Recall: how many do we find?

	Predicted <i>pos</i>	Predicted <i>neg</i>
Actually <i>pos</i>	TP = 1	FN = 12
Actually <i>neg</i>	FP = 11	TN = 165

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{1 + 165}{1 + 165 + 11 + 12} \approx 88\%$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{1}{1 + 12}$$

Precision: how sure are we?

	Predicted <i>pos</i>	Predicted <i>neg</i>
Actually <i>pos</i>	TP = 1	FN = 12
Actually <i>neg</i>	FP = 11	TN = 165

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{1 + 165}{1 + 165 + 11 + 12} \approx 88\%$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{1}{1 + 12}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{1}{1 + 11}$$

Better reflection...

	Predicted <i>pos</i>	Predicted <i>neg</i>
Actually <i>pos</i>	TP = 1	FN = 12
Actually <i>neg</i>	FP = 11	TN = 165

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{1 + 165}{1 + 165 + 11 + 12} \approx 88\%$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{1}{1 + 12} \approx 0.076$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{1}{1 + 11} \approx 0.083$$

Weighing it all together: F-score

F-score is the *harmonic mean* of recall and precision

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{1}{1 + 12} \approx 0.076$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{1}{1 + 11} \approx 0.083$$

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}}$$

Weighing it all together: F-score

F₁ is the most common variant

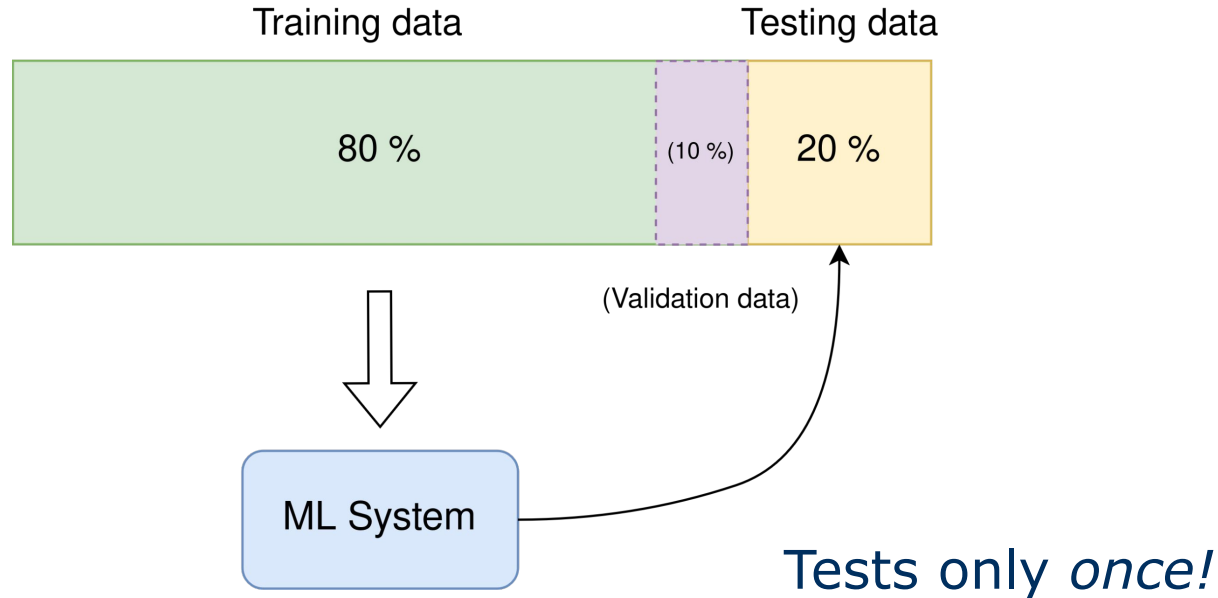
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{1}{1 + 12} \approx 0.076$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{1}{1 + 11} \approx 0.083$$

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}}$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \approx 0.079$$

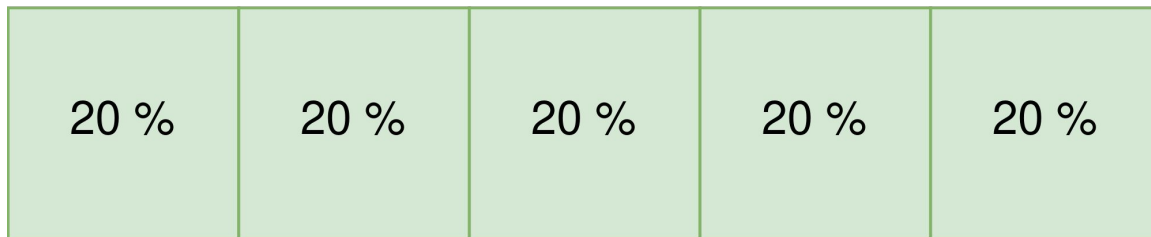
Training an ML algorithm



K-fold cross-validation

Train on $k-1$ **folds** and test on the remaining data

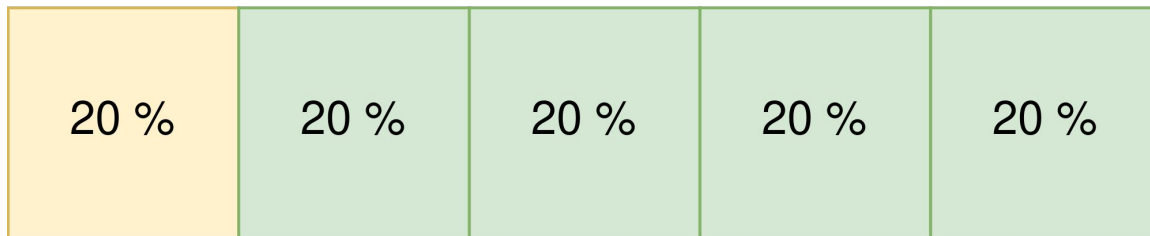
Split training data into 5 *folds*



5-fold cross-validation

Train on 4 **folds** à 20 % and test on the remaining 20 %

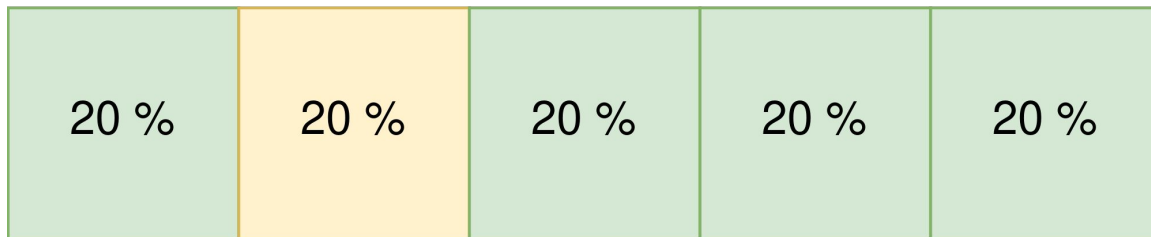
Split training data into 5 *folds*



5-fold cross-validation

Repeat by testing with the next **fold** and training on the rest

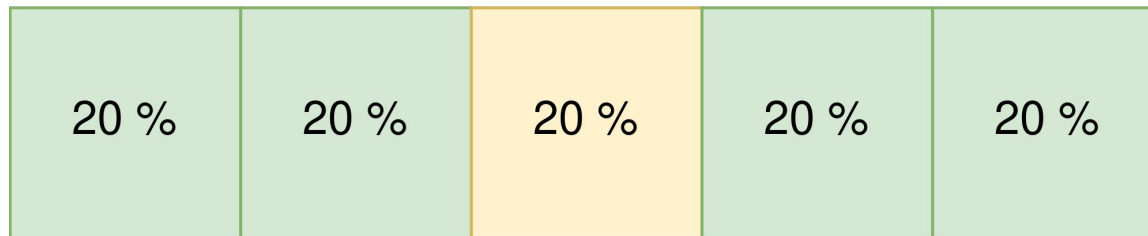
Split training data into 5 *folds*



5-fold cross-validation

etc

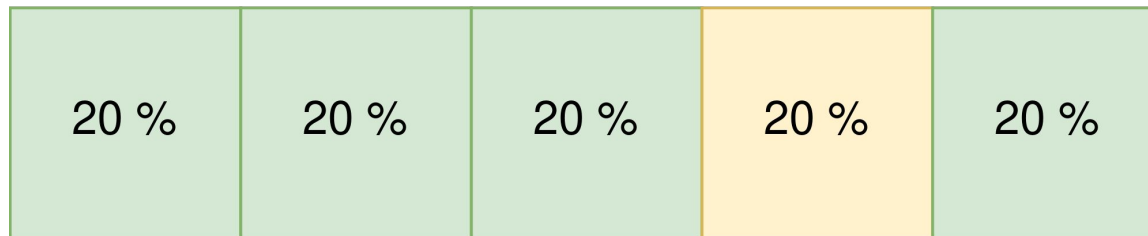
Split training data into 5 *folds*



5-fold cross-validation

etc etc

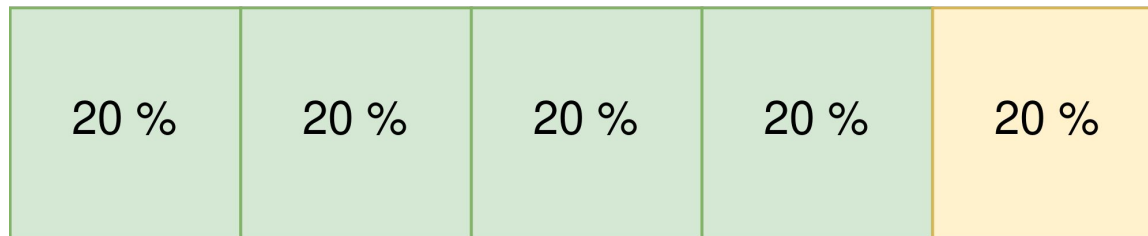
Split training data into 5 *folds*



5-fold cross-validation

etc etc etc

Split training data into 5 *folds*



Why did we do this?

In the end we have 5 different evaluations, which is more robust! We can also average together the metrics.

Split training data into 5 *folds*

20 %	20 %	20 %	20 %	20 %
------	------	------	------	------

Summary

- Classifiers can have a variety of *error types*
 - Accuracy won't always work (e.g., *class imbalance*)
- Common to use *recall, precision, and F_1*
- We train and evaluate using a train and test splits
 - Sometimes, we also use *validation data*
- Even better: we can use k-fold cross-validation
 - More robust and statistically meaningful results
 - However: takes a lot more time to do