Live session
Methods in AI research

Dong Nguyen
10 Sept 2020

Utrecht University

Practicalities

- This session won't be recorded
- Please mute your mic.
- One hour. Afterwards there's an virtual "office hour".
- Structure
  - Discussion topics/questions related to the quiz
  - Remaining topics/questions

- Next week
  - Same as this week!
  - Watch pre-recorded videos, do quiz, read literature
  - Go to the live session

But first…

- Depth vs. breadth: Some topics like machine learning or natural language processing could be complete courses on their own.

- How can we practice for the exam?
  - Quizzes, lab sessions, live sessions, *practice exam*

Let's say you work at a bank. You're asked to make a system to detect whether a credit card transaction is fraudulent or genuine. What kind of features would you use? List at least 5 features

- Characteristics of the transaction
  - Amount, time, location, etc.
- Characteristics of the receiver/sender? Maybe there is some blacklist?
- Deviations
  - E.g. How much does the amount differ from previous/average transactions
  - How far away from the usual area (stolen card?)
- Time between transactions

Let's say you work at a bank. You're asked to make a system to detect whether a credit card transaction is fraudulent or genuine. What kind of features would you use? List at least 5 features

- Characteristics of the transaction
  - Amount, time, location, etc.
- Characteristics of the receiver/sender? Maybe there is some blacklist?
- Deviations
  - E.g. How much does the amount differ from previous/average transactions
  - How far away from the usual area (stolen card?)
- Time between transactions

What features could we use to find bot accounts on Twitter?

If the input features don’t capture the necessary information, even a complex model won’t be able to do well.
So…. the more features the better?

If the input features don’t capture the necessary information, even a complex model won’t be able to do well.
So…. the more features the better?

No:

- More features increases the risk of overfitting
- Sometimes there are features that we don’t want to use (demographics)
- Feature leakage
- Interpretability

Evaluation metrics

Label	Feature 1	Feature 2	Feature 3
A	0.5	1	0
B	-0.1	1	0
A	0.2	0	1
A	0.3	0	0
B	-0.2	1	1

You have the following dataset with 5 instances. You have a classifier that always predicts the label “A”. What is the **accuracy** of this classifier? (provide a value between 0 and 1 (inclusive))

What is the accuracy?

$$\frac{\#\text{correctly labeled instances}}{\#\text{total instances}}$$

Evaluation metrics

Label	Feature 1	Feature 2	Feature 3
A	0.5	1	0
B	-0.1	1	0
A	0.2	0	1
A	0.3	0	0
B	-0.2	1	1

You have the following dataset with 5 instances. You have a classifier that always predicts the label “A”. What is the **accuracy** of this classifier? (provide a value between 0 and 1 (inclusive))

$$\frac{\#\text{correctly labeled instances}}{\#\text{total instances}}$$

What is the accuracy?

3/5 = 0.6

Evaluation metrics

Label	Feature 1	Feature 2	Feature 3
A	0.5	1	0
B	-0.1	1	0
A	0.2	0	1
A	0.3	0	0
B	-0.2	1	1

You have the following dataset with 5 instances. You have a classifier that always predicts the label "A". What is the **precision** of this classifier for class A? (provide a value between 0 and 1 (inclusive))

What fraction of the ones that you have identified belong to that class?

What is the precision for class A?

3/5 = 0.6

Evaluation metrics

Label	Feature 1	Feature 2	Feature 3
A	0.5	1	0
B	-0.1	1	0
A	0.2	0	1
A	0.3	0	0
B	-0.2	1	1

You have the following dataset with 5 instances. You have a classifier that always predicts the label "A". What is the **precision** of this classifier for class A? (provide a value between 0 and 1 (inclusive))

	Truth: A	Truth: B
Predicted: A	3 (TP)	2 (FP)
Predicted: B	0 (FN)	0 (TN)

What is the precision for class A?

$3/5 = 0.6$

$$precision = \frac{\#TP}{\#TP + \#FP}$$

Evaluation metrics

Label	Feature 1	Feature 2	Feature 3
A	0.5	1	0
B	-0.1	1	0
A	0.2	0	1
A	0.3	0	0
B	-0.2	1	1

What fraction of the ones that belong to the class have you identified?

You have the following dataset with 5 instances. You have a classifier that always predicts the label “A”. What is the **recall** of this classifier for class A? (provide a value between 0 and 1 (inclusive))

What is the recall for class A?

1

Evaluation metrics

Label	Feature 1	Feature 2	Feature 3
A	0.5	1	0
B	-0.1	1	0
A	0.2	0	1
A	0.3	0	0
B	-0.2	1	1

You have the following dataset with 5 instances. You have a classifier that always predicts the label "A". What is the **recall** of this classifier for class A? (provide a value between 0 and 1 (inclusive))

	Truth: A	Truth: B
Predicted: A	3 (TP)	2 (FP)
Predicted: B	0 (FN)	0 (TN)

What is the recall for class A?

1

$$recall = \frac{\#TP}{\#TP + \#FN}$$

1) A task for which recall is more important than precision.

- Fire alarm
- Medical screening
- Rain forecast
- Identifying possible fraudulent bank transactions
- Airport security (checking luggage)

2) A task for which precision is more important than recall.

- Recommendation systems (Spotify, YouTube)
- Recognizing recyclable plastic

1) A task for which recall is more important than precision.

- Fire alarm
- Medical screening
- Rain forecast
- Identifying possible fraudulent bank transactions
- Airport security (checking luggage)

2) A task for which precision is more important than recall.

- Recommendation systems (Spotify, YouTube)
- Recognizing recyclable plastic

It can depend on the task:

- Hashtag suggestion vs. automatic tagging hashtags
- Web search: finding similar images/retrieving a fact vs. retrieving all relevant docs (e.g. legal search)

Evaluation metrics: F1 score

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

It's the harmonic mean between precision and recall

Macro F1 average:
Calculate metrics for each class, and aggregate by taking an (unweighted) average

F1 A: 0.6
F1 B: 0.2
F1 C: 0.4

Macro F1: (0.6 + 0.2+0.4)/3 = 0.4

Evaluation metrics: F1 score

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

```
true =      [0, 1, 2, 0, 1, 2]
predicted = [0, 2, 1, 0, 1, 1]
```

Class 0:

TP = 2
FP = 0
FN = 0

Precision = 1
Recall = 1
F1score = 1

Evaluation metrics: F1 score

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

```
true =      [0, 1, 2, 0, 1, 2]
predicted = [0, 2, 1, 0, 1, 1]
```

Class 0:

TP = 2
FP = 0
FN = 0

Precision = 1
Recall = 1
F1score = 1

Class 1:

TP = 1
FP = 2
FN = 1

Precision = 1/3
Recall = ½
F1score = 0.4

Class 2:

TP = 0
FP = 1
FN = 2

Precision = 0
Recall = 0
F1score = 0

Evaluation metrics: F1 score

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

```
true =      [0, 1, 2, 0, 1, 2]
predicted = [0, 2, 1, 0, 1, 1]
```

Class 0:

TP = 2
FP = 0
FN = 0

Precision = 1
Recall = 1
F1score = 1

Class 1:

TP = 1
FP = 2
FN = 1

Precision = 1/3
Recall = ½
F1score = 0.4

Class 2:

TP = 0
FP = 1
FN = 2

Precision = 0
Recall = 0
F1score = 0

Macro average:
Calculate metrics for each class, and aggregate by taking an (unweighted) average

F1 score macro average:
(1+0.4+0)/3 = 0.466

Evaluation metrics: F1 score

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

```
true =      [0, 1, 2, 0, 1, 2]
predicted = [0, 2, 1, 0, 1, 1]
```

Class 0:

TP = 2
FP = 0
FN = 0

Precision = 1
Recall = 1
F1score = 1

Class 1:

TP = 1
FP = 2
FN = 1

Precision = 1/3
Recall = ½
F1score = 0.4

Class 2:

TP = 0
FP = 1
FN = 2

Precision = 0
Recall = 0
F1score = 0

Micro average: Calculate F1 by counting total nr of true positives, false negatives and false positives

Precision micro: TP/(TP + FP) = 3/(3 + 3) = 0.5
Recall micro = TP/(TP + FN) = 3/(3+3) = 0.5
F1 score micro = 2*0.5*0.5/(0.5+0.5) = 0.5

As conversational agents are getting better and better, it may sometimes not always be obvious that you're communicating with a conversational agent instead of a real person. An example could be when you're communicating through an online interface to receive help about a product you bought.

As conversational agents are getting better and better, it may sometimes not always be obvious that you're communicating with a conversational agent instead of a real person. An example could be when you're communicating through an online interface to receive help about a product you bought.

California bot law (2019)

- https://www.newyorker.com/tech/annals-of-technology/will-californias-new-bot-law-strengthen-democracy
- https://www.natlawreview.com/article/2019-bot-odyssey

Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases (Luo et al. 2019)

Results suggest that undisclosed chatbots are as effective as proficient workers and four times more effective than inexperienced workers in engendering customer purchases. However, a disclosure of chatbot identity before the machine–customer conversation reduces purchase rates by more than 79.7%.

Intent classification

Intent: what task/action is the user trying to accomplish?

Domain	Intent	Query
BANKING	TRANSFER	*move 100 dollars from my savings to my checking*
	TRANSFER	*put $20000 into my checking account from my savings account*
	TRANSFER	*send fifty dollars from me to carrie*
BANKING	TRANSACTIONS	*how much did my last purchase cost*
	TRANSACTIONS	*what were my last five transactions on my visa card*
	TRANSACTIONS	*what did i buy last*
BANKING	BALANCE	*what's my current bank account total*
	BALANCE	*how much do i have in my savings account*
	BALANCE	*how much is in my pnc account*
BANKING	FREEZE ACCOUNT	*may you stop a paymet on my account*
	FREEZE ACCOUNT	*please put a hold on my retirement account right now*
	FREEZE ACCOUNT	*how can i stop transactions on my account*
BANKING	PAY BILL	*can you give me a hand paying my water bill*
	PAY BILL	*pay my gas bill with my checking account*
	PAY BILL	*use my park bank account to pay my electric bill*
BANKING	BILL BALANCE	*what's my bill for water and electricity*
	BILL BALANCE	*read my bill balances*
	BILL BALANCE	*how much do i owe for my gas and phone bills*
BANKING	BILL DUE	*tell me the last day i can pay my gas bill*
	BILL DUE	*when do i have to pay my internet*
	BILL DUE	*when does my gas bill need paid by*

An Evaluation Dataset for Intent Classification and Out-of-Scope Prediction, Larson et al. 2019
https://github.com/clinc/oos-eval/blob/master/supplementary.pdf

Entropy

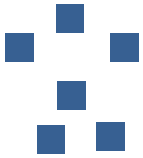
Entropy:

$$H(S) = -\sum_i p_i \log_2 p_i$$

p_i: the probability of class i (i.e. the fraction of instances of class i in S)

Entropy comes from information theory

"the amount of randomness"
"the average number of yes/no questions to guess a draw from S"



Entropy = 0

Entropy = 1

Entropy

Entropy:

$$H(S) = -\sum_i p_i \log_2 p_i$$

p_i: the probability of class i (i.e. the fraction of instances of class i in S)

Entropy comes from information theory

"the amount of randomness"
"the average number of yes/no questions to guess a draw from S"

	p
A	0.5
B	0.5
C	0

Question: Is it A?

On average we need 1 question

-0.5*log2(0.5)-0.5*log2(0.25)
-0*log2(0) = 1

Entropy

Entropy:

$$H(S) = -\sum_i p_i \log_2 p_i$$

p_i: the probability of class i (i.e. the fraction of instances of class i in S)

Entropy comes from information theory

"the amount of randomness"
"the average number of yes/no questions to guess a draw from S"

	p
A	0.5
B	0.25
C	0.25

What strategy would use you to guess my draw? (A,B, or C).

Entropy

Entropy:

$$H(S) = -\sum_i p_i \log_2 p_i$$

p_i: the probability of class i (i.e. the fraction of instances of class i in S)

Entropy comes from information theory

"the amount of randomness"
"the average number of yes/no questions to guess a draw from S"

	p
A	0.5
B	0.25
C	0.25

Heads? → A 1 * 0.5 = 0.5
Tails? → Heads? → B 2 * 0.25 = 0.5
Tails? → C 2 * 0.25 = 0.5

On average we need 1.5 questions

-0.5*log2(0.5)-0.25*log2(0.25)
-0.25*log2(0.25) = 1.5

Programming tips?

There are many tutorials/videos online.
Two frameworks that I like:

- Scikit-learn (any ML except deep learning)
- Keras (deep learning)

Start with tutorials and existing datasets.

Loss function

We'll get back to this in Lecture 4!

Decision trees:

- What are we trying to minimize? (e.g. misclassification rate)