Prediction vs. inference dilemma

MACHINE LEARNING FOR BUSINESS

Karolis Urbonas
Head of Machine Learning & Science, AWS

Inference vs. prediction dilemma

Inference or causal models:

- The goal is to understand the drivers of a business outcome
- Inference focused models are interpretable
- Less accurate than **prediction** models

Prediction:

- The prediction itself is the main goal
- Are not easily interpretable i.e. work like "black-boxes"
- Much more accurate than **inference** models

Start with the business question

- "What are the main **drivers** of fraud?"
  - Inference
- "**How much** conditions X impact heart attack risk?"
  - Inference
- "**Which** transactions are **likely** fraudulent?"
  - Prediction
- "Is the patient **at risk** of having a heart attack?"
  - Prediction

Modeling data structure

	Transaction data A	Transaction data B	Transaction data C	Transaction data D
Transaction 1				
Transaction 2				
Transaction 3				
Transaction ...				
Transaction N				

Fraud probability

Target variable

Target variable

	Transaction data A	Transaction data B	Transaction data C	Transaction data D
Transaction 1				
Transaction 2				
Transaction 3				
Transaction ...				
Transaction N				

Fraud probability

Input features

Data about transactions that the business collected (input features)

Target variable

	Transaction data A	Transaction data B	Transaction data C	Transaction data D	Fraud probability
Transaction 1					
Transaction 2					
Transaction 3					
Transaction ...					
Transaction N					

Using input features

Use these data points

	Transaction data A	Transaction data B	Transaction data C	Transaction data D
Transaction 1				
Transaction 2				
Transaction 3				
Transaction ...				
Transaction N				

Fraud probability

Predicting target variable

Use these data points → **To predict this**

	Transaction data A	Transaction data B	Transaction data C	Transaction data D	Fraud probability
Transaction 1					
Transaction 2					
Transaction 3					
Transaction ...					
Transaction N					

Inference model focus

Which of these affect the fraud probability the most?

	Transaction data A	Transaction data B	Transaction data C	Transaction data D
Transaction 1				
Transaction 2				
Transaction 3				
Transaction ...				
Transaction N				

Fraud probability

Prediction model focus

	Transaction data A	Transaction data B	Transaction data C	Transaction data D
Transaction 1				
Transaction 2				
Transaction 3				
Transaction ...				
Transaction N				

Get the most accurate probability this is fraud

Fraud probability

Let's practice!

MACHINE LEARNING FOR BUSINESS

Inference (causal) models

MACHINE LEARNING FOR BUSINESS

Karolis Urbonas
Head of Machine Learning & Science, Amazon

What is causality?

- Identify causal relationship of how much certain actions affect an outcome of interest
- Answers the "why?" questions
- Optimizes for model interpretability vs. performance
- Models try to detect patterns in observed data (observational) and draw causal conclusions

Experiments vs. observations

- Experiments are designed and causal conclusions are guaranteed e.g. in A/B tests
- When experiments are impossible (unethical, too expensive, both) - the models are used (also called observational studies) to calculate effect of certain inputs on desired outcomes
- Experiments are **always** preferred over observational studies whenever possible

Best practices

1. Do experiments wherever you can

2. If running experiments all the time is too expensive, run them periodically (quarterly, annually) and use it as benchmark

3. If there are no way to run any experiments, build a causal model. This will require an advanced methodology

Inference model example

	Last month spend	Recency in days	Average cart value	Store visits per year
Customer 1	845 USD	20	340 USD	32
Customer 2	205 USD	1	100 USD	25
Customer 3	0 USD	55	70 USD	14
Customer ...	...	...	...	...
Customer N	43	114.5	134	61.2

Next month spend
585 USD
150 USD
20 USD
...
69 USD

Inference - training

Use these data points

	Last month spend	Recency in days	Average cart value	Store visits per year
Customer 1	845 USD	20	340 USD	32
Customer 2	205 USD	1	100 USD	25
Customer 3	0 USD	55	70 USD	14
Customer ...	...	...	...	...
Customer N	43	114.5	134	61.2

Next month spend
585 USD
150 USD
20 USD
...
69 USD

Inference - learning

Use these data points → **Learn rules to predict this**

	Last month spend	Recency in days	Average cart value	Store visits per year	Next month spend
Customer 1	845 USD	20	340 USD	32	585 USD
Customer 2	205 USD	1	100 USD	25	150 USD
Customer 3	0 USD	55	70 USD	14	20 USD
Customer ...	...	...	...	...	...
Customer N	43	114.5	134	61.2	69 USD

Inference - regression coefficients

Coefficients

Last month spend	Recency in days	Average cart value	Store visits per year
0.58	-0.03	0.28	0.18

Inference - interpretation

Coefficients

Last month spend	Recency in days	Average cart value	Store visits per year
0.58	-0.03	0.28	0.18

How much 1 incremental USD spent in the last month results in predicted next month spend. Here, the customers who on average spent 1 USD **more** in the last month, will spend 0.58 USD more in the next month compared to customers with 1 USD **less** last month.

Let's practice!

MACHINE LEARNING FOR BUSINESS

Prediction models (supervised learning)

MACHINE LEARNING FOR BUSINESS

Karolis Urbonas
Head of Machine Learning & Science, Amazon

Supervised vs. Unsupervised

Supervised models

> Predicting **class/type** of an outcome (e.g. subscription cancellation, fraud, purchase) - **CLASSIFICATION**
>
> Predicting **quantity** of an outcome (e.g. dollars spent, hours played) - **REGRESSION**

Unsupervised models

> Clustering - grouping observations into similar groups or clusters (e.g. customer or market segmentation)

Supervised learning types

Classification - Target variable is categorical (discrete) (class of outcome) (**classification**)

> Will the customer cancel a service subscription?
>
> Is this transaction fraudulent?
>
> What is the profession of this user?

Regression - Target variable is continuous (amount of outcome) (**regression**)

> Number of product purchases next month
>
> Number of gaming hours next year
>
> Dollars spent on insurance

Data collection

Machine learning teams should collect all available data to predict desired outcome with the highest degree of accuracy e.g. in case of purchase predictions:

> Customer information
>
> Purchase history, cancellations, order amount
>
> Browsing history, logs, errors
>
> Device details and location
>
> Product/service usage frequency
>
> And others...

Classification example

	Past fraud count	Time of transaction	Declined in T-30 days	Amount	Fraud
Transaction 1	20	3 am	Yes	5.25 USD	Yes
Transaction 2	1	9 pm	Yes	19.5 USD	Yes
Transaction 3	0	9.30 am	No	500 USD	No
Transaction ...					
Transaction N					

Classification - training

Use these data points

	Past fraud count	Time of transaction	Declined in T-30 days	Amount
Transaction 1	20	3 am	Yes	5.25 USD
Transaction 2	1	9 pm	Yes	19.5 USD
Transaction 3	0	9.30 am	No	500 USD
Transaction ...				
Transaction N				

Fraud
Yes
Yes
No

Classification - learning

Use these data points

Learn rules to predict this on unseen data

	Past fraud count	Time of transaction	Declined in T-30 days	Amount	Fraud
Transaction 1	20	3 am	Yes	5.25 USD	Yes
Transaction 2	1	9 pm	Yes	19.5 USD	Yes
Transaction 3	0	9.30 am	No	500 USD	No
Transaction ...					
Transaction N					

Classification - unseen data

New unseen data

	Past fraud count	Time of transaction	Declined in T-30 days	Amount
Transaction 1	0	12.30 am	Yes	200 USD
Transaction 2	3	1 am	Yes	19.52 USD
Transaction 3	0	10 am	No	500 USD

Fraud probability

Classification - prediction

New unseen data

	Past fraud count	Time of transaction	Declined in T-30 days	Amount
Transaction 1	0	12.30 am	Yes	200 USD
Transaction 2	3	1 am	Yes	19.52 USD
Transaction 3	0	10 am	No	500 USD

Predict fraud probability

Fraud probability
23 %
85 %
0 %

Regression example

	Last month spend	Recency in days	Average cart value	Store visits per year		Next month spend
Customer 1	845 USD	20	340 USD	32		585 USD
Customer 2	205 USD	1	100 USD	25		150 USD
Customer 3	0 USD	55	70 USD	14		20 USD
Customer ...	...	...	...	...		...
Customer N	43	114.5	134	61.2		69 USD

Regression - training

Use these data points

	Last month spend	Recency in days	Average cart value	Store visits per year
Customer 1	845 USD	20	340 USD	32
Customer 2	205 USD	1	100 USD	25
Customer 3	0 USD	55	70 USD	14
Customer ...	...	...	...	...
Customer N	43	114.5	134	61.2

Next month spend
585 USD
150 USD
20 USD
...
69 USD

Regression - learning

Use these data points → **Learn rules to predict this on unseen data**

	Last month spend	Recency in days	Average cart value	Store visits per year	Next month spend
Customer 1	845 USD	20	340 USD	32	585 USD
Customer 2	205 USD	1	100 USD	25	150 USD
Customer 3	0 USD	55	70 USD	14	20 USD
Customer ...	...	...	...	...	...
Customer N	43	114.5	134	61.2	69 USD

Regression - unseen data

New unseen data

	Last month spend	Recency in days	Average cart value	Store visits per year
Customer 1	400 USD	10	200	12
Customer 2	5000 USD	60	1000	6
Customer 3	10 USD	25	200	12

Next month spend

Regression - prediction

New unseen data

	Last month spend	Recency in days	Average cart value	Store visits per year
Customer 1	400 USD	10	200	12
Customer 2	5000 USD	60	1000	6
Customer 3	10 USD	25	200	12

Predict next month spend

Next month spend
287 USD
3152 USD
62 USD

Regression - actual prediction

New unseen data

	Last month spend	Recency in days	Average cart value	Store visits per year
Customer 1	400 USD	10	200	12
Customer 2	5000 USD	60	1000	6
Customer 3	10 USD	25	200	12

Predict next month spend

Next month spend
287 USD
3152 USD
62 USD

These are **real** predictions based from a linear regression model!

Let's practice!

MACHINE LEARNING FOR BUSINESS

Prediction models (unsupervised learning)

MACHINE LEARNING FOR BUSINESS

Karolis Urbonas
Head of Machine Learning & Science, Amazon

What is unsupervised machine learning?

Unsupervised models

> **Clustering** - grouping observations into similar groups or clusters (e.g. customer or market segmentation)
>
> **Anomaly detection** - detecting which observations fall out of the discovered "regular pattern" and use it as an input in supervised learning or a business input
>
> **Recommender engines** - e.g. recommending products or services to customers based on their similarity to other customers e.g. Netflix movie recommendations

Clustering example - segmentation

	Annual spend	Recency in days	Store visits per year
Customer 1	8450 USD	20	32
Customer 2	2050 USD	1	25
Customer 3	450 USD	55	14
Customer ...	...	...	...
Customer N	628 USD	114.5	61.2

Segmentation - data

	Monetary value	Recency	Frequency
Customer 1	8450 USD	20	32
Customer 2	2050 USD	1	25
Customer 3	450 USD	55	14
Customer ...	...	...	...
Customer N	628 USD	114.5	61.2

Segmentation - training

Use these data points

	Monetary value	Recency	Frequency
Customer 1	8450 USD	20	32
Customer 2	2050 USD	1	25
Customer 3	450 USD	55	14
Customer ...	...	...	...
Customer N	628 USD	114.5	61.2

Segmentation - discover

Use these data points

	Monetary value	Recency	Frequency
Customer 1	8450 USD	20	32
Customer 2	2050 USD	1	25
Customer 3	450 USD	55	14
Customer ...	...	...	...
Customer N	628 USD	114.5	61.2

To identify similar segments of customers

Segmentation - analyze

Use these data points

	Monetary value	Recency	Frequency
Customer 1	8450 USD	20	32
Customer 2	2050 USD	1	25
Customer 3	450 USD	55	14
Customer ...	...	...	...
Customer N	628 USD	114.5	61.2

To identify similar segments of customers

	Monetary value	Recency	Frequency
1	4788	76	74
2	8872	21	34
3	1312	29	21

Let's practice!

MACHINE LEARNING FOR BUSINESS