Nonparametric modelling for spaced repetition scheduling

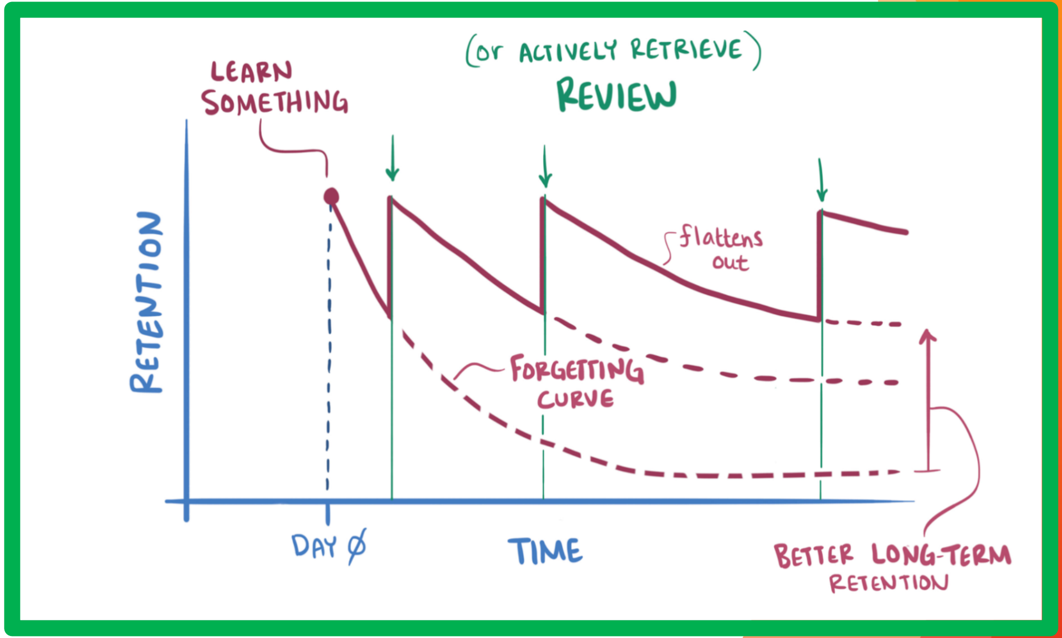
Francesco Giammaria, Giacomo Randazzo, Matteo S. Savino

Politecnico di Milano - 17/12/2020

Spaced Repetition

Spaced repetition is a method for **memorizing concepts**:

- No cramming, reviews spaced through time
- <u>Increasing durations</u> between reviews as one learns the item
- Software schedules each review



duolingo and Half-Life Regression

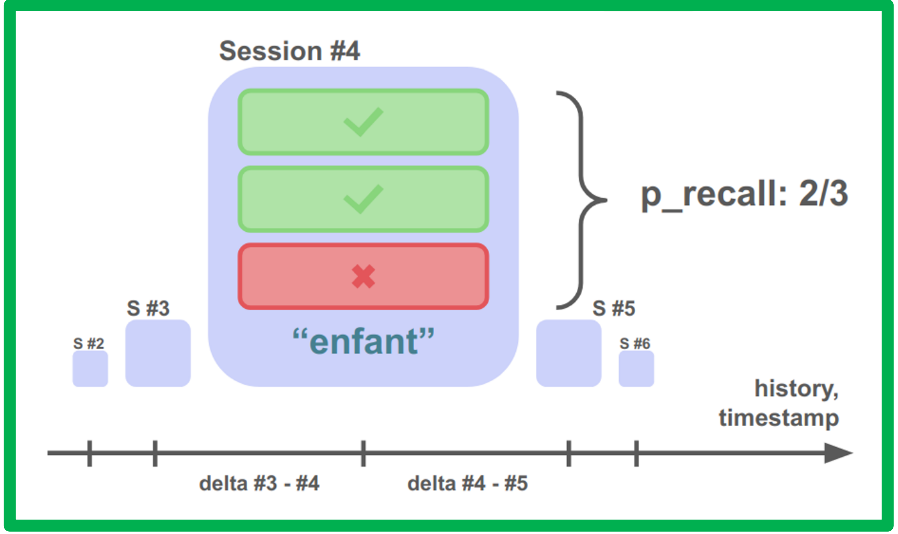
Duolingo is a language learning app:

- Relies on spaced repetition under the hood
- Half-Life Regression model to estimate the user's probability of recalling an item at any point in time after the last review

Half-Life Regression paper + dataset:

- 2 weeks of real usage data
- 115'000 users
- 13 million word recall probabilities

Our Data



- **p_recall** - proportion of exercises from this lesson/practice where the word/lexeme was correctly recalled
- **timestamp** - UNIX timestamp of the current lesson/practice
- **delta** - time (in seconds) since the last lesson/practice that included this word/lexeme
- **user_id** - student user ID who did the lesson/practice (anonymized)
- **learning_language** - language being learned
- **ui_language** - user interface language (presumably native to the student)
- **lexeme_id** - system ID for the lexeme tag (i.e., word)
- **lexeme_string** - lexeme tag (see below)
- **history_seen** - total times user has seen the word/lexeme prior to this lesson/practice
- **history_correct** - total times user has been correct for the word/lexeme prior to this lesson/practice
- **session_seen** - times the user saw the word/lexeme during this lesson/practice
- **session_correct** - times the user got the word/lexeme correct during this lesson/practice

	p_recall	timestamp	delta	user_id	learning_language	ui_language	lexeme_id	lexeme_string	history_seen	history_correct	session_seen	session_correct
1	1.0000000	1362076081	27649635	u:FO	de	en	76390c1350a8dac31186187e2fe1e178	lernt/lernen<vblex><pri><p3><sg>	6	4	2	2
2	0.5000000	1362076081	27649635	u:FO	de	en	7dfd7086f3671685e2cf1c1da72796d7	die/die<det><def><f><sg><nom>	4	4	2	1
3	1.0000000	1362076081	27649635	u:FO	de	en	35a54c25a2cda8127343f6a82e6f6b7d	mann/mann<n><m><sg><nom>	5	4	1	1
4	0.5000000	1362076081	27649635	u:FO	de	en	0cf63ffe3dda158bc3dbd55682b355ae	frau/frau<n><f><sg><nom>	6	5	2	1

Goals

- **Main goals**: Duolingo's HLR model is parametric, they make a strong assumption that, by looking at the data, might not be valid. We'll build a nonparametric model with the same goals.
- **Extensions**: It could be interesting to see if the some languages or words are easier to learn in Duolingo compared to others or if some users learn more quickly than others. We'll do this by comparing appropriate survival functions of items in memory.

Model (1)

HLR formula: $$p = 2^{-\Delta/h}$$

We introduced some new variables:

`delta` → `delta_day`

`p_history` = `history_correct / history_seen`

`lexeme_length` = length of the word

Our GAM model

```
model_gam <- gam(p_recall ~
                   s(p_history, bs='cr') +
                   s(delta_day, bs='cr', k=3) +
                   s(lexeme_length,bs='cr') +
                   learning_language_3,
                 family="binomial", method="REML")
```

Model (1) Results

```r
model_gam <- gam(p_recall ~
                   s(p_history, bs='cr') +
                   s(delta_day, bs='cr', k=3) +
                   s(lexeme_length,bs='cr') +
                   learning_language_3,
                 family="binomial", method="REML")
```

MAE: **0.167**

Model (1) (on a validation set, different from the paper's test set)

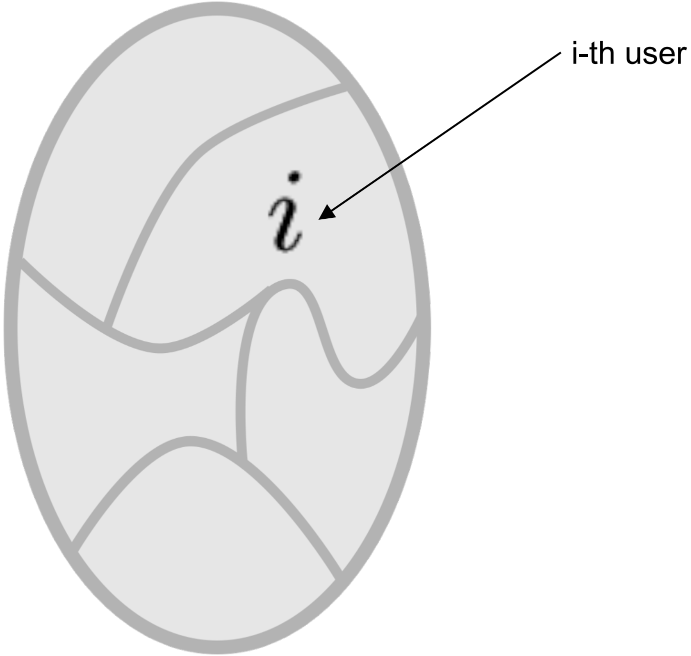
Model	MAE↓	AUC↑
HLR	**0.128***	0.538*
HLR -lex	**0.128***	0.537*
HLR -h	0.350	0.528*
HLR -lex-h	0.350	0.528*
Leitner	0.235	**0.542***
Pimsleur	0.445	0.510*
LR	0.211	0.513*
LR -lex	0.212	0.514*
Constant $\bar{p} = 0.859$	0.175	n/a

(Test set) Results from the Half-Life Regression paper.

Model (2)

i i-th user

avg_user_p



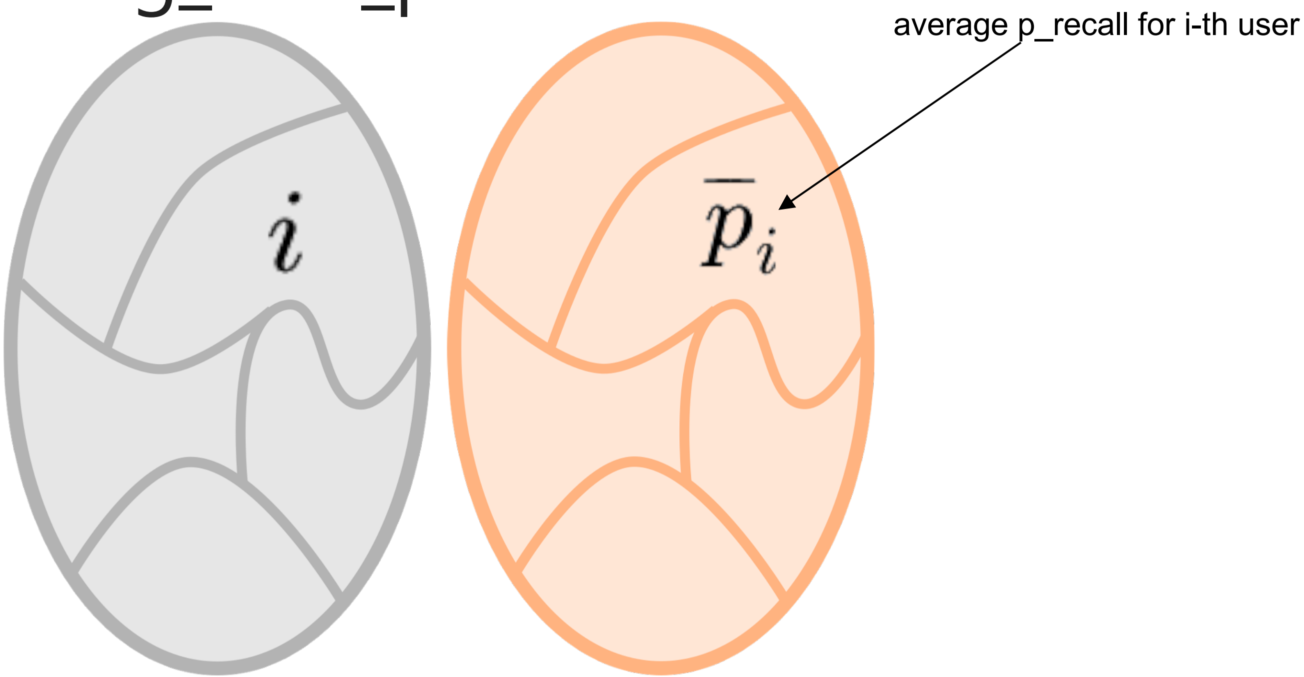
Training set

Model (2)

i i-th user

$\bar{p}_i$ average p_recall for i-th user

avg_user_p



Training set

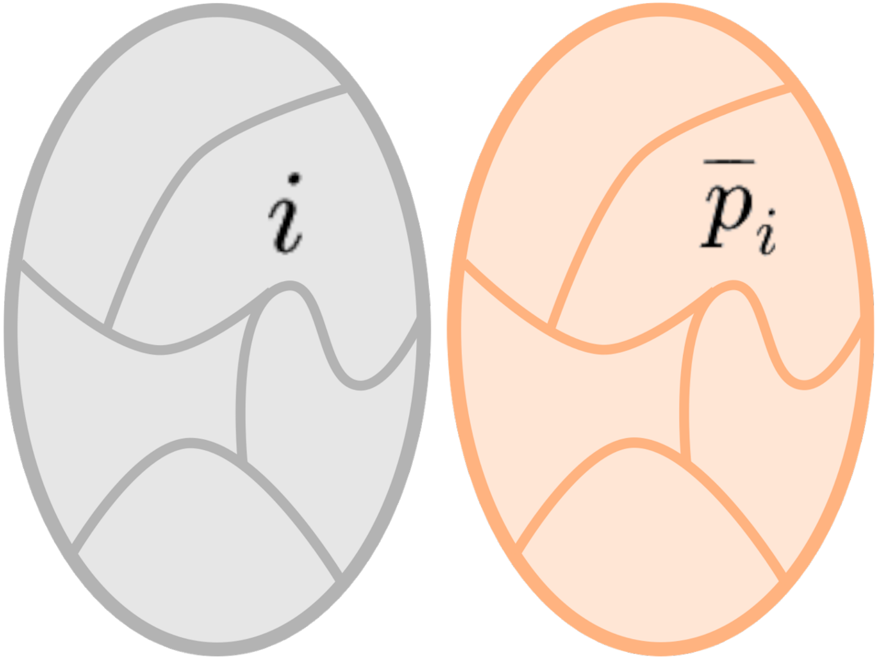
Model (2)

avg_user_p

i i-th user

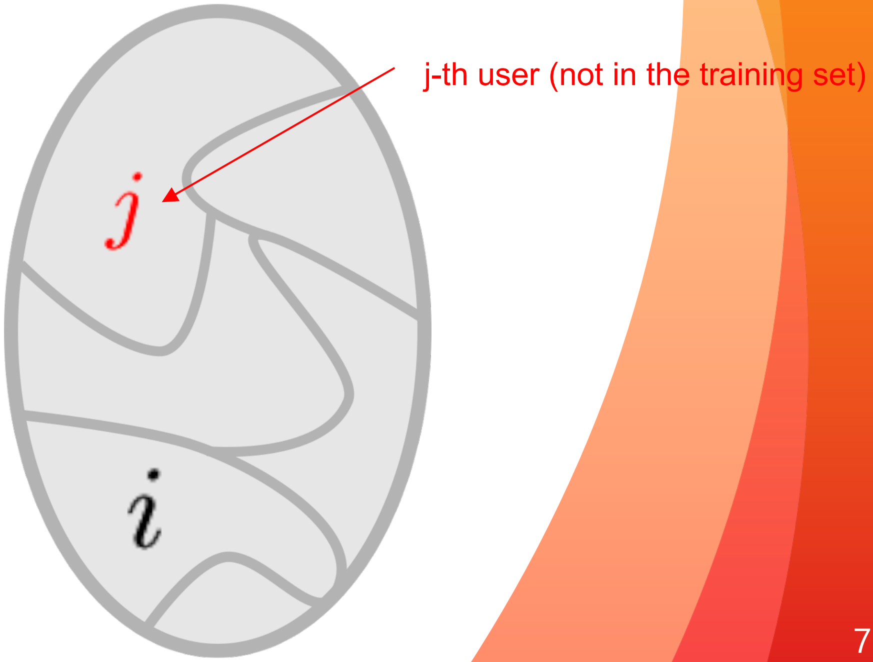
$\bar{p}_i$ average p_recall for i-th user

j j-th user (not in the training set)



Training set

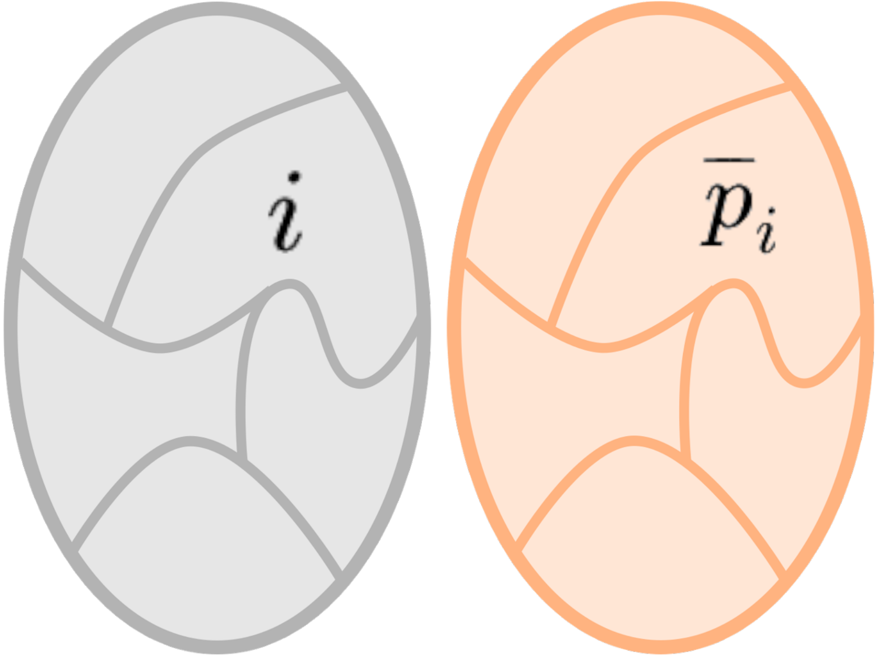
Validation (and test) set



Model (2)

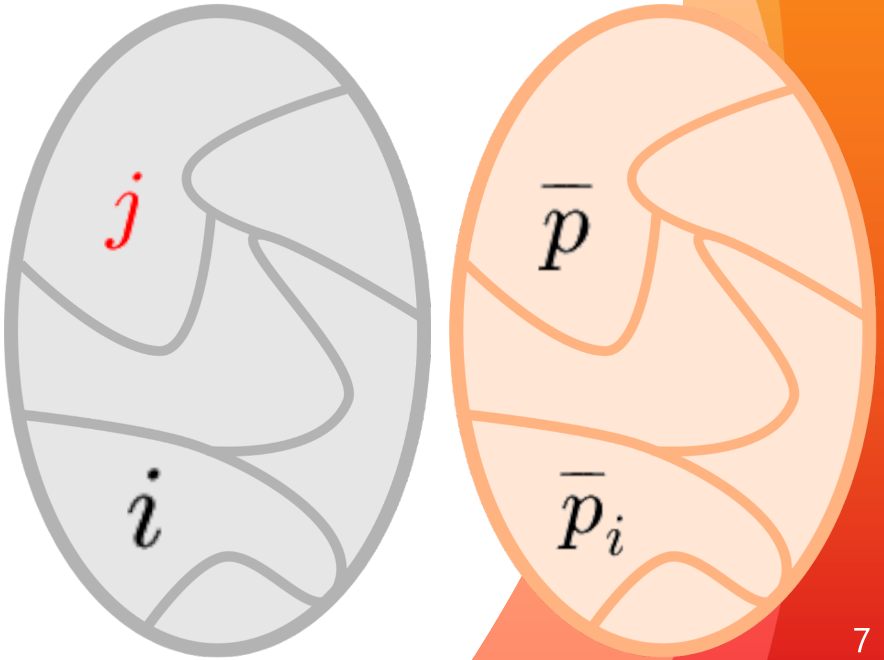
avg_user_p

i	i-th user
$\bar{p}_i$	average p_recall for i-th user
j	j-th user (not in the training set)
$\bar{p}$	global average p_recall



Training set

Validation (and test) set



Model (2) Results

```
glm(p_recall ~ avg_user_p,
    family="binomial")
```

Permutation Test: $H_0 : \beta_{\text{avg\_user\_p}} = 0 \quad vs \quad H_1 : \beta_{\text{avg\_user\_p}} \neq 0$

p-value = 0

MAE: **0.122**
AUC: **0.625**
Model (2) (on a validation set, different from the paper's test set)

VS

MAE: **0.128**
AUC: **0.538**
HLR model (on paper's test set)

Future

Explore the GAM approach further

Try to exploit word tag data

Continue to study languages and words and try to understand if some of them are easier to learn by comparing appropriate survival functions

THANK YOU !!!

POLITECNICO
MILANO 1863

References

- Settles, Meeder (2016) A Trainable Spaced Repetition Model for Language Learning. In Proceedings of the Association for Computational Linguistics (ACL), pages 1848-1858.
- Settles, Burr (2017) Replication Data for: A Trainable Spaced Repetition Model for Language Learning. https://doi.org/10.7910/DVN/N8XJME, Harvard Dataverse, V1
- Matuschak and Nielsen (2019) How can we develop transformative tools for thought? https://numinous.productions/ttft, San Francisco