# Teaching a robot to interpret natural language navigation instructions

Ryan Eloff

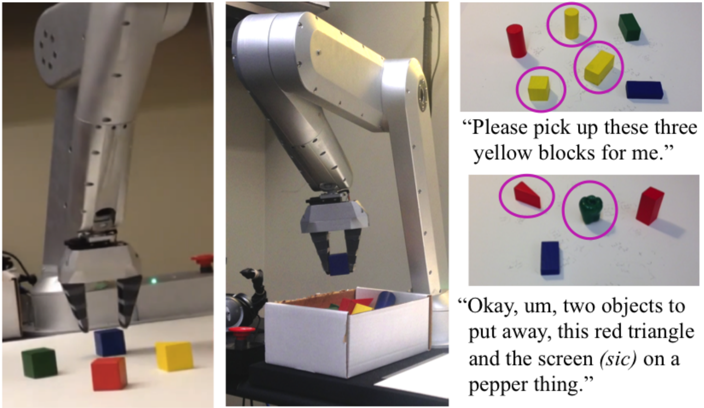
Supervisor: Dr. H. Kamper
Department of Electrical & Electronic Engineering

Maties Machine Learning
Stellenbosch University

23 February 2018

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

# Natural language for human-robot interaction




(a) (b) (c)

The Gambit platform which learns to identify and pick up objects from a users speech and gestures.[1]

[1] C. Matuszek, L. Bo, L. Zettlemoyer, *et al.*, "Learning from unscripted deictic gesture and language for human-robot interactions," in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

# Natural language for human-robot interaction


A robotic wheelchair that learns about the environment from a narrated tour.[2]

[2] M. R. Walter, S. Hemachandra, B. Homberg, *et al.*, "A framework for learning semantic maps from grounded natural language descriptions," *The International Journal of Robotics Research*, 2014.

# Natural language for human-robot interaction

- Still face challenges in building robots that can understand our written natural language instructions.

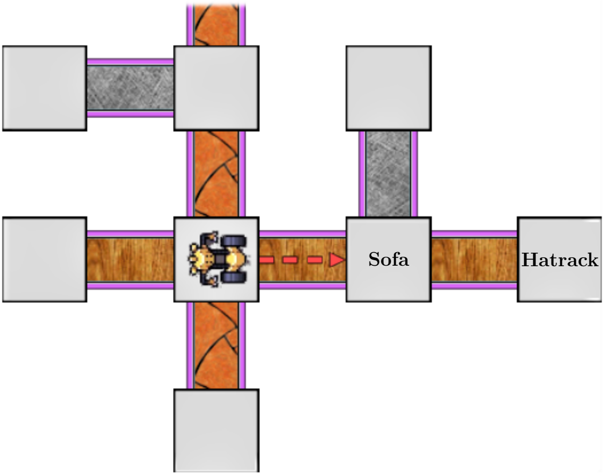
# Natural language for human-robot interaction

- Still face challenges in building robots that can understand our written natural language instructions.
- Most existing systems are only available in most commonly used languages: English, German, Spanish ...

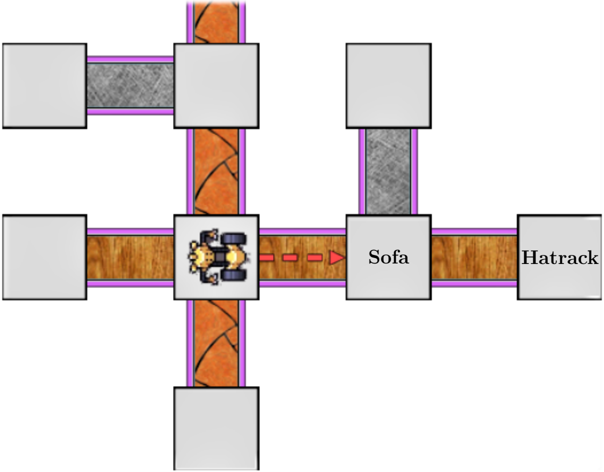
# Natural language for human-robot interaction

- Still face challenges in building robots that can understand our written natural language instructions.
- Most existing systems are only available in most commonly used languages: English, German, Spanish ...
- These technologies are not available for low-resource languages.
- This is a problem in South Africa!

# Natural language instructions for route direction

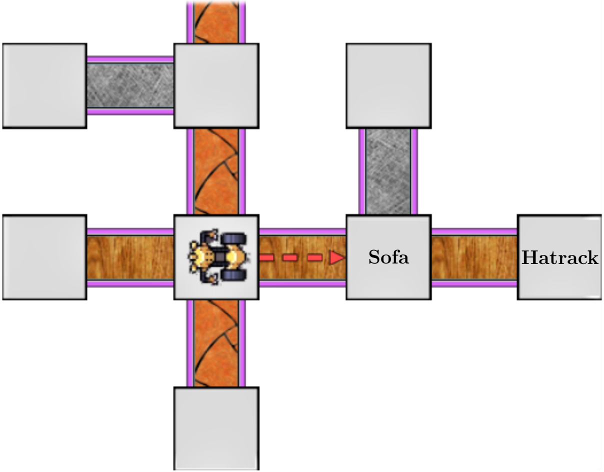


# Natural language instructions for route direction




- "Walk forward once"
- "Move to the sofa"
- "Go down till the chair"
- "Go forward one segment to the intersection with a bare concrete floor"

# Natural language instructions for route direction

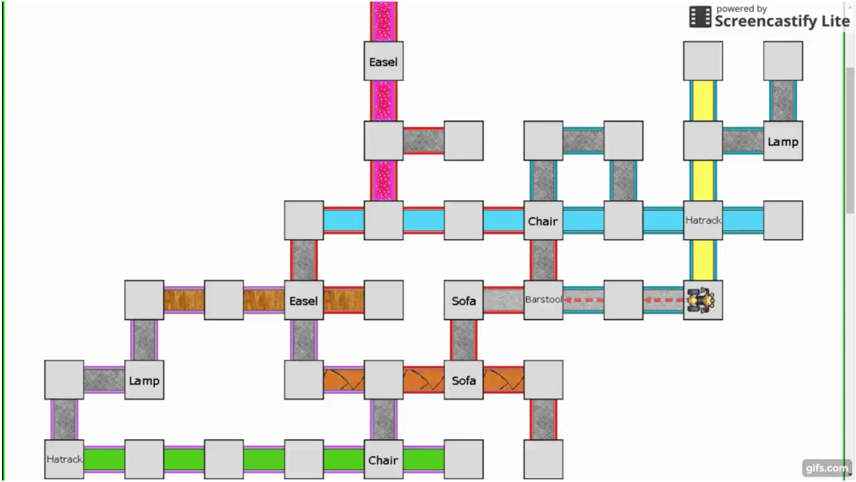



- "Walk forward once"
- "Move to the sofa"
- "Go down till the chair"
- "Go forward one segment to the intersection with a bare concrete floor"

Can we build a language independent system for interpreting written natural language navigation instructions that is also less dependent on data?

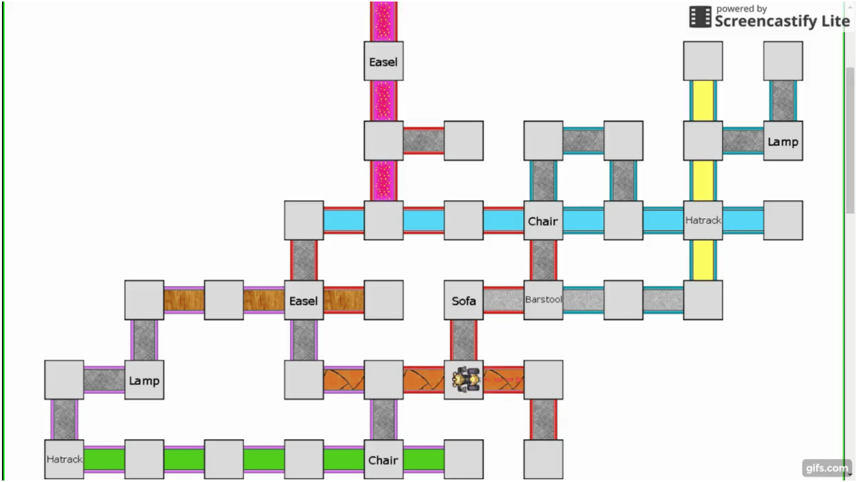
# Evaluation, virtual environments, & web application

# Example instruction data



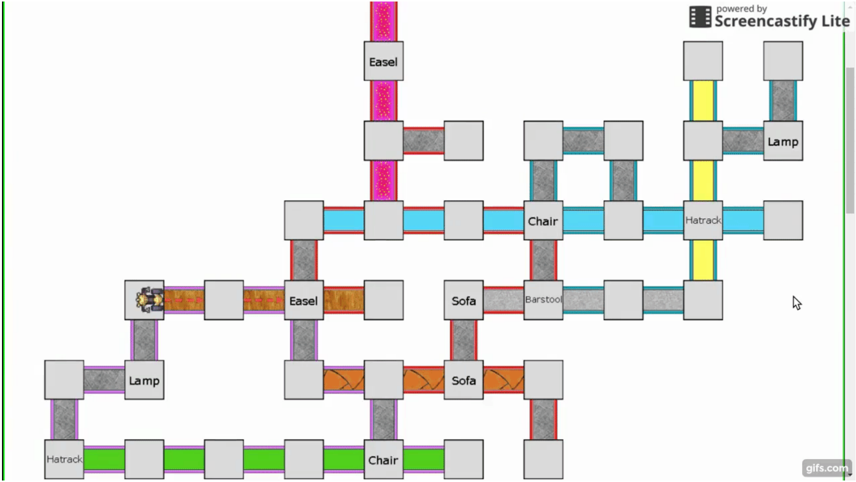

- Example 1: "Walk two steps forward"

# Example instruction data



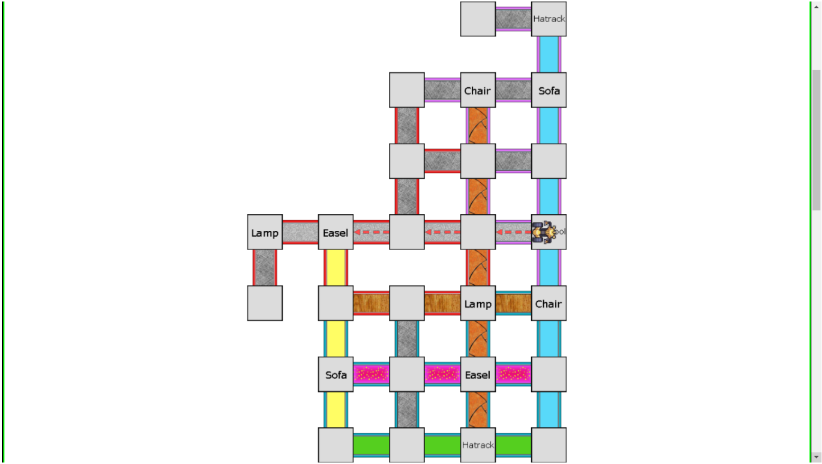

- Example 2: "Move forward and turn right"

# Example instruction data




- Example 3: “Move to the easel”

# Unseen instruction problem




- Unseen: “Walk three steps to the easel”

# General model

# General model

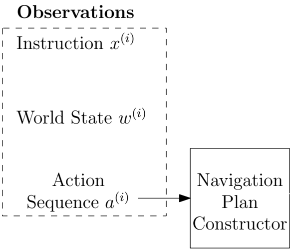
**Observations**

Instruction $x^{(i)}$

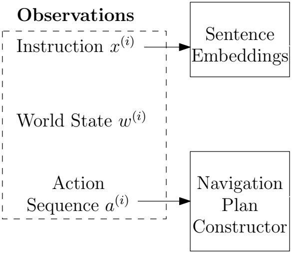
World State $w^{(i)}$

Action Sequence $a^{(i)}$

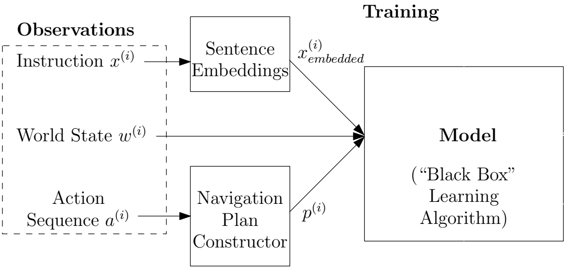
# General model



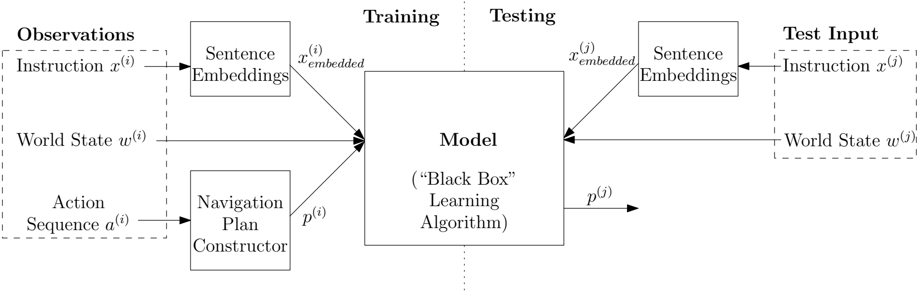
# General model



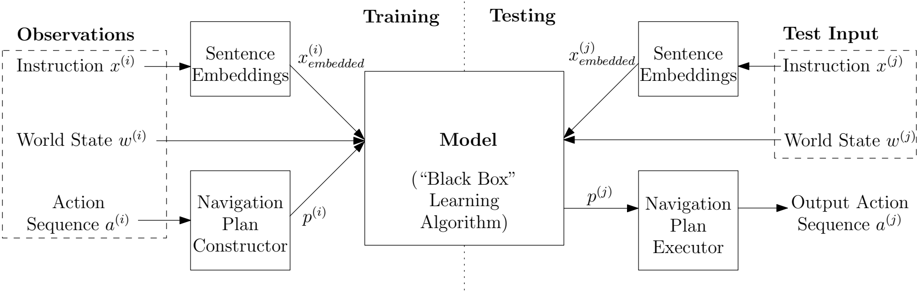
# General model



# General model

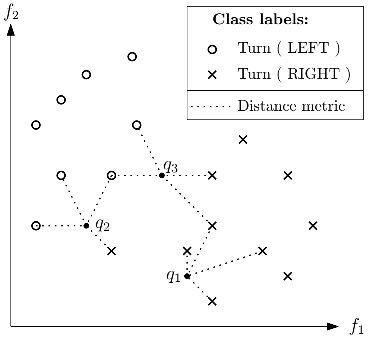


# General model



# Proposed models

k-Nearest neighbours (kNN) models:

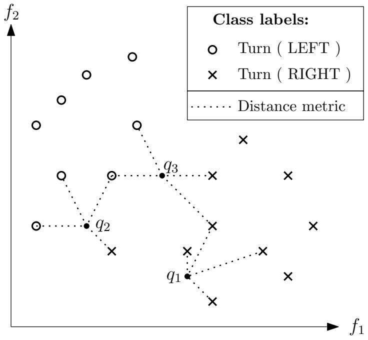



Proposed distance metrics:

- Levenshtein distance
- Euclidean distance

# Proposed models

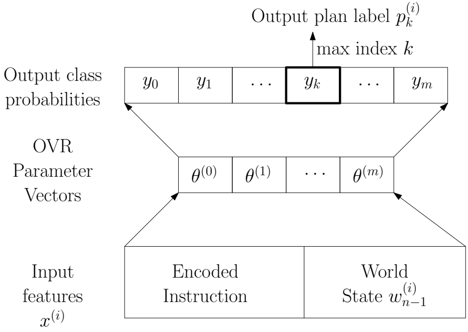
k-Nearest neighbours (kNN) models:




Proposed distance metrics:

- Levenshtein distance
- Euclidean distance

Basic one-vs-rest (OVR) logistic regression model:



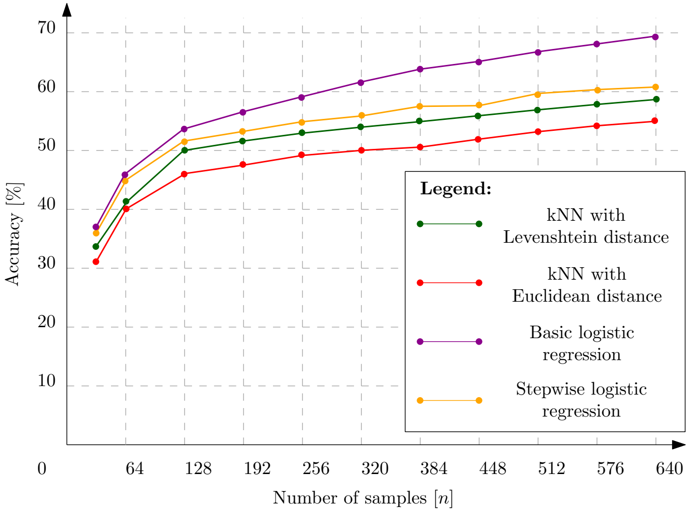
# Results: Overall model performance

- Overall performance of the proposed models and that of previous work:

| Model | $F_1$ | Accuracy |
|---|---|---|
| kNN with Levenshtein distance | 77.10 | 55.25 |
| kNN with Euclidean distance | 75.44 | 51.65 |
| Basic logistic regression | **78.18** | **57.32** |
| Stepwise logistic regression | 72.58 | 44.36 |
| Chen and Mooney (2011) | 68.37 | 54.40 |
| Chen (2012) | 69.43 | 57.28 |
| Mei et al. (2016) | - | **69.98** |

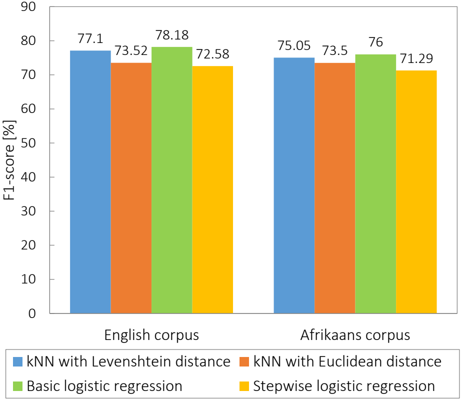
# Results: Performance on small datasets

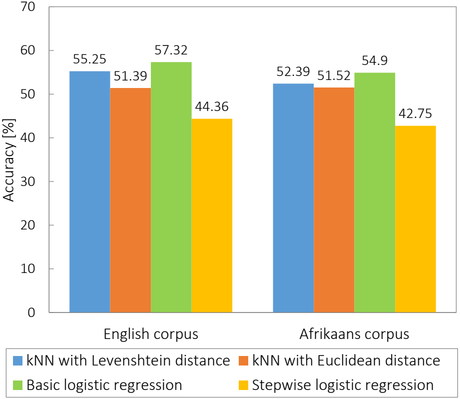
- Average strict accuracy of the proposed models when trained on different sample sizes $n$:



# Results: Performance on an Afrikaans corpus

- Proposed models evaluated on an Afrikaans translation of the benchmark English corpus*:




* English corpus translated to Afrikaans using Google's state-of-the-art Neural Machine Translation system.

# Contributions

- Simple machine learning models used for the novel purpose of learning from limited datasets, which is competitive with complex systems of previous work.

# Contributions

- Simple machine learning models used for the novel purpose of learning from limited datasets, which is competitive with complex systems of previous work.
- Web application has data collection capabilities which is ideal for future research on human-robot interaction.

# Contributions

- Simple machine learning models used for the novel purpose of learning from limited datasets, which is competitive with complex systems of previous work.
- Web application has data collection capabilities which is ideal for future research on human-robot interaction.
- These models are the first steps and foundation for future research on more complex models for limited data. (and speech!!)

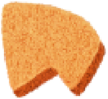
# Contributions

- Simple machine learning models used for the novel purpose of learning from limited datasets, which is competitive with complex systems of previous work.
- Web application has data collection capabilities which is ideal for future research on human-robot interaction.
- These models are the first steps and foundation for future research on more complex models for limited data. (and speech!!)
- First Afrikaans system for Robotic instruction following. The system may also be trained in any language, so it is ideal for low-resource languages.

# My current work

- One-shot learning of small multimodal datasets consisting of images combined with speech captions.
- One-shot learning example:[3]

This is a "dax".



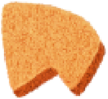
- Goal: Develop systems that learn and think like people.

[3] R. Feinman and B. M. Lake, "Learning inductive biases with simple neural networks," *arXiv:1802.02745*, 2018.

# My current work

- One-shot learning of small multimodal datasets consisting of images combined with speech captions.
- One-shot learning example:[3]

This is a "dax".



Which of these is a "dax"?

(1)

(2)

(3)

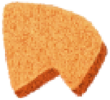
- Goal: Develop systems that learn and think like people.

[3] R. Feinman and B. M. Lake, "Learning inductive biases with simple neural networks," *arXiv:1802.02745*, 2018.

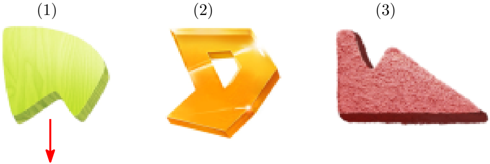
# My current work

- One-shot learning of small multimodal datasets consisting of images combined with speech captions.
- One-shot learning example:[3]

This is a "dax".



Which of these is a "dax"?




- Goal: Develop systems that learn and think like people.

[3] R. Feinman and B. M. Lake, "Learning inductive biases with simple neural networks," *arXiv:1802.02745*, 2018.

# Some other slides ...

# Word embeddings with latent semantic analysis

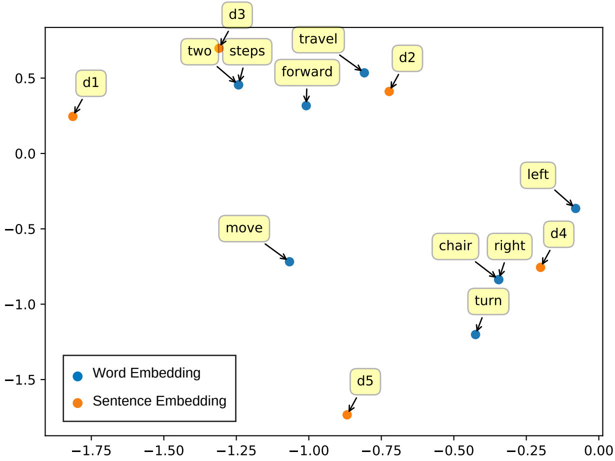
Example instructions:

- $d_1$ : *Move forward two steps*
- $d_2$ : *Travel forward*
- $d_3$ : *Travel two steps*
- $d_4$ : *Turn left*
- $d_5$ : *Turn right and move to the chair*

# Word embeddings with latent semantic analysis

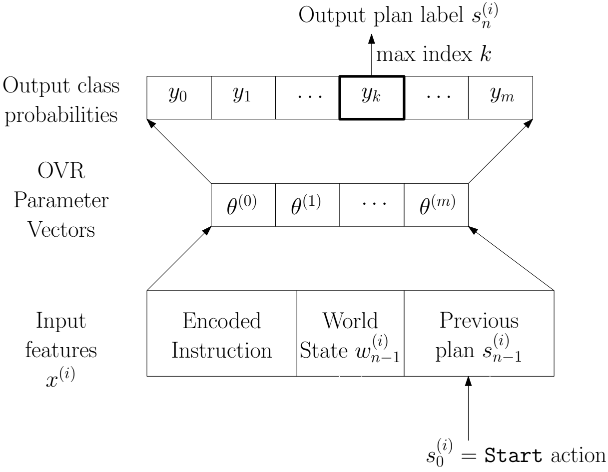
Example instructions:

- $d_1$ : *Move forward two steps*
- $d_2$ : *Travel forward*
- $d_3$ : *Travel two steps*
- $d_4$ : *Turn left*
- $d_5$ : *Turn right and move to the chair*



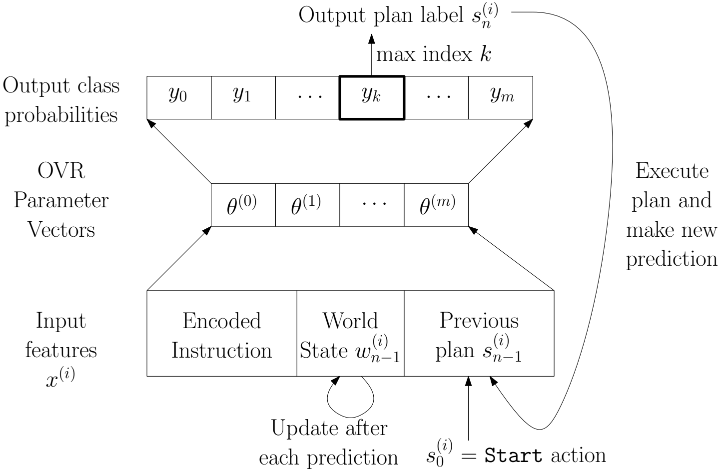
# Stepwise logistic regression

Stepwise one-vs-rest (OVR) logistic regression model:



# Stepwise logistic regression

Stepwise one-vs-rest (OVR) logistic regression model:



# Stepwise logistic regression

Stepwise one-vs-rest (OVR) logistic regression model: