CSI5180. Machine Learning for Bioinformatics Applications

Rule Learning

by

Marcel Turcotte

Version **November 21, 2019**

Preamble

Preamble

Rule Learning

Chances are that you have never heard the term **rule learning** despite the fact that it is **one of the oldest paradigms in machine learning**. Particularly now, the emphasis is on developing machine learning algorithms with exceptionally high "accuracy". We have deep learning algorithms with superhuman powers classifying images, detecting cancer from medical images, or defeating the world champions of Go, one of the most challenging games. In this lecture, we focus on a set of methods putting the emphasis on **interpretability** rather than numerical performance.

General objective :

- **Explain** rule learning in your own words

Learning objectives

- **Justify** the need (or not) for interpretability
- **Explain** rule learning in your own words

Reading:

- Fürnkranz, D. Gamberger, and N. Lavrač. *Foundations of Rule Learning.* Cognitive Technologies. Springer Berlin Heidelberg, 2012.
- King, R. D. et al. The automation of science. *Science* **324**, 8589 (2009).
- Sparkes, A. et al. Towards Robot Scientists for autonomous scientific discovery. *Autom Exp* **2**, 1 (2010).
- King, R. D., Schuler Costa, V., Mellingwood, C. & Soldatova, L. N. Automating Sciences: Philosophical and Social Dimensions. *IEEE Technology and Society Magazine* **37**, 4046 (2018).

Plan

1. Preamble
2. Introduction
3. Building blocks
4. Science (fiction)
5. Current research
6. Prologue

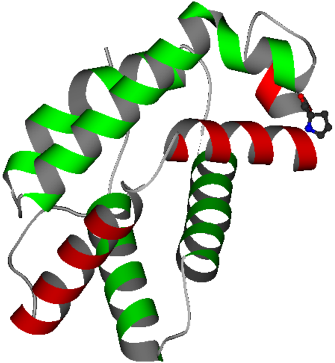
2020

- Make this the last lecture of the term.

Introduction

Rule Learning,
a vast and diverse continent
that you may never have heard of.

Globin-like



```
fold('Globin-like', X) :-
    adjacent(X, A, B, 1, h, h),
    has_pro(B).
```

Flavodoxin, Rossman-fold, TIM-barrel

```
fold('Flavodoxin-like',A) :-
  nb_alpha(A, B), nb_beta(A,B), interval_l(B ≤ 6).

fold('NAD(P)-binding Rossmann-fold domains',A) :-
  nb_alpha(A,B), nb_beta(A,B), interval(5 ≤ B ≤ 7).

fold('beta/alpha (TIM)-barrel',A) :-
  nb_alpha(A,B), nb_beta(A,B), interval(8 ≤ B ≤ 16).
```

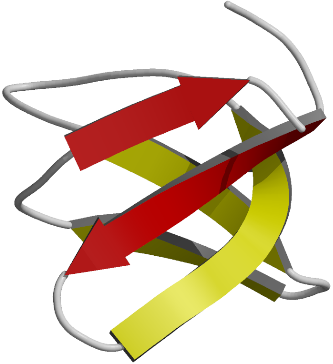
- The number of strands is the **same** as the number of helices, however, that number is variable.

Beta-grasp

```
fold('beta-Grasp',A) :-
  adjacent(A,B,C,2,e,h),
  adjacent(A,C,D,1,h,e),
  coil(C,D,3).
```

- This rule effectively describes a relation involving **three** secondary structure elements, β_2-α_1-β_3, although **no triple relationship was explicitly encoded** in the background knowledge.

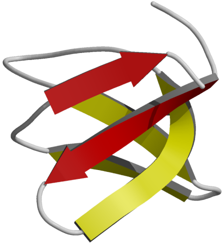
SH3



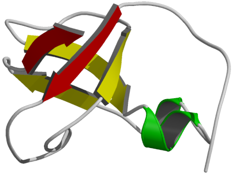
```
fold(A, 'SH3-like barrel') :-
    number_strands(4 =< A =< 7),
    sheet(A, B, anti),
    has_n_strands(B, 5),
    strand(A, C, B,  1),
    strand(A, D, B, -1),
    antiparallel(C, D).
```

- The **first** and the **last** are anti-parallel!

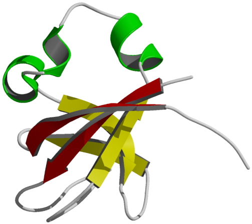
SH3



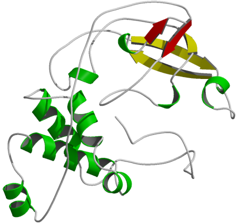
(1bia)



(d1bb)



(d1pht)



(2ahj)

“Inductive” Logic Programming

Examples:
Phycocyanin adopts a globin fold.
Hemoglobin adopts a globin fold.
Oct-1 POU Homeodomain is not a globin.

$+$

Background:
The second helix in phycocyanin contains a proline.
To calculate the hydrophobic moment . . .

$\Downarrow$

Hypothesis:
The first helix is followed by another one that contains a proline.

Keywords

- **Knowledge** discovery

Keywords

- **Knowledge** discovery
  - Can **expert-like** knowledge be discovered automatically?

Keywords

- **Knowledge** discovery
  - Can **expert-like** knowledge be discovered automatically?
- **Background** knowledge

Keywords

- **Knowledge** discovery
  - Can **expert-like** knowledge be discovered automatically?
- **Background** knowledge
  - How can we make effective use of **accumulated knowledge**?

Keywords

- **Knowledge** discovery
  - Can **expert-like** knowledge be discovered automatically?
- **Background** knowledge
  - How can we make effective use of **accumulated knowledge**?
- **Relational** information

Keywords

- **Knowledge** discovery
  - Can **expert-like** knowledge be discovered automatically?
- **Background** knowledge
  - How can we make effective use of **accumulated knowledge**?
- **Relational** information
  - Can we learn **complex interactions** between sub-structures?

Keywords

- **Knowledge** discovery
  - Can **expert-like** knowledge be discovered automatically?
- **Background** knowledge
  - How can we make effective use of **accumulated knowledge**?
- **Relational** information
  - Can we learn **complex interactions** between sub-structures?
- **Interpretability**

Keywords

- **Knowledge** discovery
  - Can **expert-like** knowledge be discovered automatically?
- **Background** knowledge
  - How can we make effective use of **accumulated knowledge**?
- **Relational** information
  - Can we learn **complex interactions** between sub-structures?
- **Interpretability**
  - How can we make **hypotheses** easily amenable to **human interpretation**?

Building blocks

Foundation

- These algorithms are based on **formal logic**, a sub-branch of mathematics.

Foundation

- These algorithms are based on **formal logic**, a sub-branch of mathematics.
  - **Propositional (zero-order) logic**

Foundation

- These algorithms are based on **formal logic**, a sub-branch of mathematics.
  - **Propositional (zero-order) logic**
    - “If it’s raining then it’s cloudy”

Foundation

- These algorithms are based on **formal logic**, a sub-branch of mathematics.
  - **Propositional (zero-order) logic**
    - “If it’s raining then it’s cloudy”
  - **First-order (predicate) logic**

Foundation

- These algorithms are based on **formal logic**, a sub-branch of mathematics.
  - **Propositional (zero-order) logic**
    - “If it’s raining then it’s cloudy”
  - **First-order (predicate) logic**
    - “there exists x such that x is Socrates and x is a man”

Foundation

- These algorithms are based on **formal logic**, a sub-branch of mathematics.
  - **Propositional (zero-order) logic**
    - “If it’s raining then it’s cloudy”
  - **First-order (predicate) logic**
    - “there exists x such that x is Socrates and x is a man”
- J.W. Lloyd, *Logic for learning: Learning comprehensible theories from structured data*, Cognitive Technologies, Springer Berlin Heidelberg, 2003.

Foundation

- These algorithms are based on **formal logic**, a sub-branch of mathematics.
  - **Propositional (zero-order) logic**
    - "If it's raining then it's cloudy"
  - **First-order (predicate) logic**
    - "there exists x such that x is Socrates and x is a man"
- J.W. Lloyd, *Logic for learning: Learning comprehensible theories from structured data*, Cognitive Technologies, Springer Berlin Heidelberg, 2003.
- Fürnkranz, D. Gamberger, and N. Lavrač. *Foundations of Rule Learning*. Cognitive Technologies. Springer Berlin Heidelberg, 2012.

Task - concept (classification)

Given:

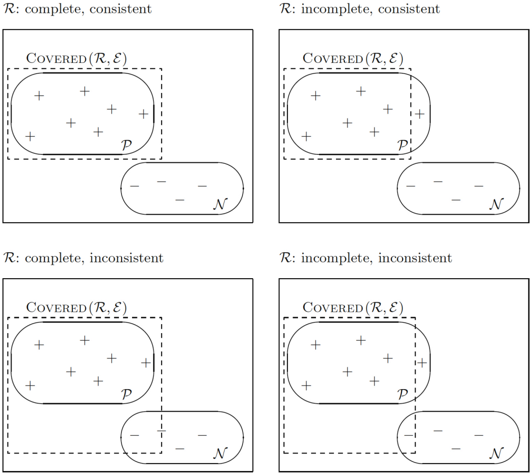
- A **data description language**
- A **target concept**
- A **hypothesis description language**
- A **coverage function**, covered($\mathbf{r}$, e)
- A **class attribute**, C
- A **set of positive examples**, $\mathcal{P}$
- A **set of negative examples**, $\mathcal{N}$

Find:

- A **hypothesis** which is:
  - **complete**, covers all the examples, and
  - **consistent**, predicts the correct class for all the examples.

Adapted from [Fürnkranz et al., 2012] Figure 2.2.

Completeness and consistency



Source: [Fürnkranz et al., 2012] Figure 2.3.

Definitions

- An **instance** is **covered** by a **rule**, if the **description** of the instance **satisfies** the **conditions** of the rule.

Definitions

- An **instance** is **covered** by a **rule**, if the **description** of the instance **satisfies** the **conditions** of the rule.
- An **example** is **correctly covered** by a **rule**, if it is **covered** and the **class of the rule** is the **same** as the **class of the example**.

Representation

- **Propositional (attribute-value) rules.**

Representation

- **Propositional (attribute-value) rules**.
  - The rules have the form:

Representation

- **Propositional (attribute-value) rules.**
  - The rules have the form:
    - IF *Conditions* THEN c

Representation

- **Propositional (attribute-value) rules.**
  - The rules have the form:
    - IF *Conditions* THEN *c*
    - where *Conditions* is a conjunction (and) of simple tests (properties of the instance) and *c* is a class.

Representation

- **Propositional (attribute-value) rules.**
  - The rules have the form:
    - IF *Conditions* THEN c
    - where *Conditions* is a conjunction (and) of simple tests (properties of the instance) and c is a class.
  - Corresponds to the **implication** in **propositional logic**, $c \leftarrow \mathit{Conditions}$.

$$\text{SportsCar} \leftarrow \text{HasChildren} = \text{No} \land \text{Sex} = \text{Male}$$

Representation

- **Propositional (attribute-value) rules**.
  - The rules have the form:
    - IF *Conditions* THEN c
    - where *Conditions* is a conjunction (and) of simple tests (properties of the instance) and c is a class.
  - Corresponds to the **implication** in **propositional logic**, $c \leftarrow Conditions$.

$$\text{SportsCar} \leftarrow \text{HasChildren} = \text{No} \wedge \text{Sex} = \text{Male}$$

- Alternatively, **first-order logic** can be used to represent the data, the background knowledge, and the hypotheses.

Representation

- **Propositional (attribute-value) rules**.
  - The rules have the form:
    - IF *Conditions* THEN c
    - where *Conditions* is a conjunction (and) of simple tests (properties of the instance) and c is a class.
  - Corresponds to the **implication** in **propositional logic**, $c \leftarrow \mathit{Conditions}$.

$$\text{SportsCar} \leftarrow \text{HasChildren} = \text{No} \land \text{Sex} = \text{Male}$$

- Alternatively, **first-order logic** can be used to represent the data, the background knowledge, and the hypotheses.
  - **first-order learning**, **relational learning** or **inductive logic programming**

```
daughter(X,Y) :- female(X), parent(Y,X).
```

Overfitting

- Rule learning systems are also susceptible to **overfitting**.
  - **Completeness** and **consistency** are too strong requirements in the presence of noise.
    - The systems are then forced to learn too **specific rules**.
    - These **criteria are relaxed**, allowing the systems to tolerate a small number of errors.

Progol

Given. The logic programs B and E
where,
B is the background knowledge, and
E is a set of examples (E^+ and E^-)

Find. Hypothesis H, from a predefined language $\mathcal{L}$, such that,

$$B \wedge H \models E$$

and

$$|B \wedge H| < |B \wedge E|$$

Where $||$ is some measure of **complexity** (simplicity)

Progol's algorithm

1. If $E = \emptyset$ return B
2. Select the first positive example in E
3. $\boxed{\text{Construct the “most specific” clause } (\bot)}$
4. $\boxed{\text{General to specific search}}$
5. Add the “**best**” clause to B
6. Remove all examples entailed (covered) by B
7. Goto 1

Step 3 - Constructing ⊥

```
[Generalizing fold('Globin',d1scta_).]
[Most specific clause is]

fold('Globin-like',A) :-
        adjacent(A,B,C,1,h,h),
        adjacent(A,C,D,2,h,h),
        adjacent(A,D,E,3,h,h),
        adjacent(A,E,F,4,h,h),
        adjacent(A,F,G,5,h,h),
        len_interval('$sk0'=<A=<'$sk2'),
        nb_alpha_interval('$sk0'=<A=<'$sk2'),
        nb_beta_interval('$sk0'=<A=<'$sk2'),
        coil(B,C,1), coil(C,D,3), coil(D,E,2),
        coil(E,F,2), coil(F,G,1),
        unit_len(B,hi), unit_len(D,hi),
        unit_len(F,lo), unit_len(G,hi),
        unit_aveh(F,hi),
        unit_hmom(F,lo), unit_hmom(G,lo),
        has_pro(C), has_pro(G).
```

Step 4 – General to Specific Search

The search starts with the most general clause: "everything is a Globin".

```
[C:-8,13,20,0 fold('Globin', X).]
```

Step 4 – General to Specific Search

The search starts with the most general clause: "everything is a Globin".

```
[C:-8,13,20,0 fold('Globin', X).]
```

The clause is specialized: "every domain such that the first helix is followed by another helix".

```
[C:-6,13,17,0 fold('Globin', X) :- adjacent(X,A,B,1,h,h).]
```

Step 4 - General to Specific Search

The search starts with the most general clause: "everything is a Globin".

```
[C:-8,13,20,0 fold('Globin', X).]
```

The clause is specialized: "every domain such that the first helix is followed by another helix".

```
[C:-6,13,17,0 fold('Globin', X) :- adjacent(X,A,B,1,h,h).]
```

The clause is specialized again: "every domain such that the first helix is followed by another helix and another helix".

```
[C:-2,13,12,0 fold('Globin', X) :- adjacent(X,A,B,1,h,h),
                                   adjacent(X,B,C,2,h,h).]

...
```

Step 4 – General to Specific Search

The hypothesis which has the highest score is reported.

```
f=8,p=13,n=1,h=0
[Result of search is]

fold('Globin', X) :-
        adjacent(X,A,B,1,h,h),
        adjacent(X,B,C,2,h,h),
        len(135 =< X =< 166).
```

Applications in bioformatics

- Drug structure-activity
- Mutagenesis
- Predicting protein secondary structure
- Protein fold
- Gene function
- Sorting peptides
- Many more

Implementations

- **Propositional (zero-order) logic**
  - CN2, RIPPER, PRIM, Opus, Apriori
- **First-order (predicate) logic**
  - Foil, Duce, Cigol, Progol, Aleph

Summary

- **Rule learning** systems are based on **formal logic**
- **Expressive** - they have the ability to learn complex relationships
- **Human readable** representations
- Can make use of **accumulated knowledge**

Science (fiction)

Robot scientist

- In a series of publications, **Ross King** and colleagues have described the **Robot scientist**:

Robot scientist

- In a series of publications, **Ross King** and colleagues have described the **Robot scientist**:
  - Ross D. King, Vlad Schuler Costa, Chris Mellingwood, and Larisa N. Soldatova, Automating sciences: Philosophical and social dimensions, *IEEE Technol. Soc. Mag.* **37**;1, 4046, 2018.

Robot scientist

- In a series of publications, **Ross King** and colleagues have described the **Robot scientist**:
  - Ross D. King, Vlad Schuler Costa, Chris Mellingwood, and Larisa N. Soldatova, Automating sciences: Philosophical and social dimensions, *IEEE Technol. Soc. Mag.* **37**;1, 4046, 2018.
  - Sparkes, A. et al. Towards Robot Scientists for autonomous scientific discovery. *Autom Exp* **2**:1, 2010.

Robot scientist

*"The question of whether it is possible to **automate the scientific process** is of both great **theoretical** interest and increasing **practical** importance because, in many scientific areas, **data are being generated much faster than they can be effectively analysed**."*

Robot scientist

"The question of whether it is possible to ***automate the scientific process*** *is of both great* ***theoretical*** *interest and increasing* ***practical*** *importance because, in many scientific areas,* ***data are being generated much faster than they can be effectively analysed****."*

- Ross D King, Kenneth E Whelan, Ffion M Jones, Philip G K Reiser, Christopher H Bryant, Stephen H Muggleton, Douglas B Kell, and Stephen G Oliver, Functional genomic hypothesis generation and experimentation by a robot scientist, *Nature* **427**:6971, 24752, 2004.

Closed-loop machine learning

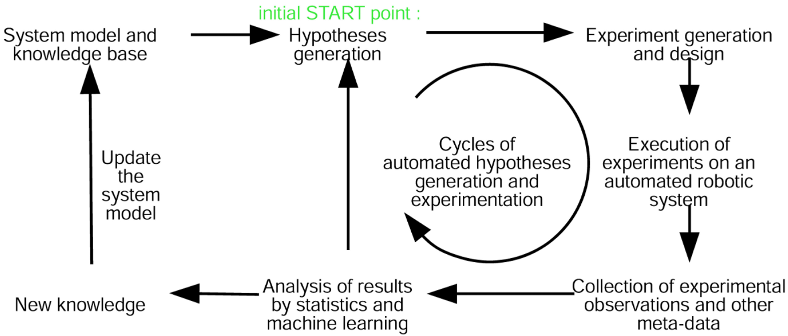


Figure 1 Hypothesis-driven closed-loop learning. Diagram showing how iterative cycles of hypothesis-driven experimentation allow for the autonomous generation of new scientific knowledge.

Source: [Sparkes et al., 2010] Figure 1

Robot scientist

- “The system **automatically originates hypotheses** to explain observations,”

Source: [King et al., 2004]

Robot scientist

- “The system **automatically originates hypotheses** to explain observations,”
- “**devises experiments** to test these hypotheses,”

Source: [King et al., 2004]

Robot scientist

- “The system **automatically originates hypotheses** to explain observations,”
- “**devises experiments** to test these hypotheses,”
- “**physically runs** the experiments using a laboratory robot,”

Source: [King et al., 2004]

Robot scientist

- “The system **automatically originates hypotheses** to explain observations,”
- “**devises experiments** to test these hypotheses,”
- “**physically runs** the experiments using a laboratory robot,”
- “**interprets the results** to falsify hypotheses inconsistent with the data,”

Source: [King et al., 2004]

Robot scientist

- “The system **automatically originates hypotheses** to explain observations,”
- “**devises experiments** to test these hypotheses,”
- “**physically runs** the experiments using a laboratory robot,”
- “**interprets the results** to falsify hypotheses inconsistent with the data,”
- “and then **repeats** the cycle.”

Source: [King et al., 2004]

Prototype

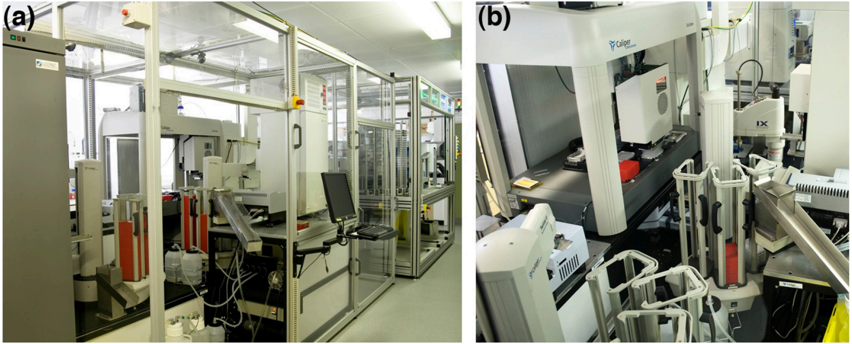


Figure 2 Adam's laboratory robotic system. (a) An external view of Adam's laboratory robotic system, also showing Eve's on the far right, and (b) a view looking down through the middle of Adam's robotic system, again with Eve's beyond.

Source: [Sparkes et al., 2010] Figure 2

Experiment

- "[T]he determination of **gene function** using **deletion mutants** of yeast (*Saccharomyces cerevisiae*) and auxotrophic growth experiments."

Source: [King et al., 2004]

Experiment

- "[T]he determination of **gene function** using **deletion mutants** of yeast (*Saccharomyces cerevisiae*) and auxotrophic growth experiments."
- At the time, 30% of the genes in *Saccharomyces cerevisiae* had no known function.

Source: [King et al., 2004]

Mechanisms

- “The model infers (deduces) that **a knockout mutant will grow if, and only if, a path can be found from the input metabolites to the three aromatic amino acids**. This allows the model to compute the **phenotype** of a particular knockout or to be used to infer missing reactions that could explain an observed phenotype (abduction).”

Source: [King et al., 2004]

Mechanisms

- "The model infers (deduces) that **a knockout mutant will grow if, and only if, a path can be found from the input metabolites to the three aromatic amino acids**. This allows the model to compute the **phenotype** of a particular knockout or to be used to infer missing reactions that could explain an observed phenotype (abduction)."
- Abduction "starts with an observation or set of observations then seeks to find the simplest and most likely explanation for the observations." [Wikipedia,2019-11-21]

Source: [King et al., 2004]

Mechanisms

- “The model infers (deduces) that **a knockout mutant will grow if, and only if, a path can be found from the input metabolites to the three aromatic amino acids**. This allows the model to compute the **phenotype** of a particular knockout or to be used to infer missing reactions that could explain an observed phenotype (abduction).”
- Abduction “starts with an observation or set of observations then seeks to find the simplest and most likely explanation for the observations.” [Wikipedia,2019-11-21]
- **ASE-Progol**, where ASE = Active Selection of Experiments.

Source: [King et al., 2004]

Conclusions

- "We **show** that an intelligent experiment selection strategy is **competitive with human performance** and significantly outperforms, with a **cost decrease** of 3-fold and 100-fold (respectively), both cheapest and random-experiment selection."

Source: [King et al., 2004]

Conclusions

- "We **show** that an intelligent experiment selection strategy is **competitive with human performance** and significantly outperforms, with a **cost decrease** of 3-fold and 100-fold (respectively), both cheapest and random-experiment selection."
- "The model correctly predicted at least 98.5% of the experiments (...)"

Source: [King et al., 2004]

Conclusions

- "We **show** that an intelligent experiment selection strategy is **competitive with human performance** and significantly outperforms, with a **cost decrease** of 3-fold and 100-fold (respectively), both cheapest and random-experiment selection."
- "The model correctly predicted at least 98.5% of the experiments (. . .)"
- "Nevertheless, the Robot Scientist has currently only been demonstrated to rediscover the role of genes of known function;"

Source: [King et al., 2004]

Conclusions

- "We **show** that an intelligent experiment selection strategy is **competitive with human performance** and significantly outperforms, with a **cost decrease** of 3-fold and 100-fold (respectively), both cheapest and random-experiment selection."
- "The model correctly predicted at least 98.5% of the experiments (...)"
- "Nevertheless, the Robot Scientist has currently only been demonstrated to rediscover the role of genes of known function;"
- "Moreover, the application of the Robot Scientist to functional genomics provides further evidence that some aspects of scientific reasoning can be formalized and efficiently automated."

Source: [King et al., 2004]

Current research

Current topics

- Stochastic logic programs
- Predicate invention
- Deep Relational Machines (DRM)

Prologue

Summary

- Rule learning systems are based on **formal logic**.

Summary

- Rule learning systems are based on **formal logic**.
- The resulting rules are **easily understandable by humans**.

Summary

- Rule learning systems are based on **formal logic**.
- The resulting rules are **easily understandable by humans**.
- But also, these systems are ideally suited for **reasoning**, thus providing a foundation for **automated scientific discovery**.

Next module

- **Graph** Learning

Collaborators

- **Imperial Cancer Research Fund, Biomolecular Modelling Laboratory**
  - Michael J.E. Sternberg, head of the group
- **University of York, Department of Computer Science**
  - Stephen H. Muggleton, chair in Machine Learning
- **Industrial collaborators**
  - Mansoor Saqi, Bioinformatics at **Glaxo-Wellcome**
  - Chris Rawlings, Bioinformatics at **Smithkline Beecham**

References

Fürnkranz, J., Gamberger, D., and Lavrač, N. (2012).
Foundations of Rule Learning.
Cognitive Technologies. Springer Berlin Heidelberg.

Ghahramani, Z. (2015).
Probabilistic machine learning and artificial intelligence.
Nature, 521(7553):452–9.

King, R. D., Costa, V. S., Mellingwood, C., and Soldatova, L. N. (2018).
Automating sciences: Philosophical and social dimensions.
IEEE Technol. Soc. Mag., 37(1):40–46.

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L. N., Sparkes, A., Whelan, K. E., and Clare, A. (2009a).
The automation of science.
Science, 324(5923):85–9.

References

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L. N., Sparkes, A., Whelan, K. E., and Clare, A. (2009b).
Make way for robot scientists.
Science, 325(5943):945.

King, R. D., Whelan, K. E., Jones, F. M., Reiser, P. G. K., Bryant, C. H., Muggleton, S. H., Kell, D. B., and Oliver, S. G. (2004).
Functional genomic hypothesis generation and experimentation by a robot scientist.
Nature, 427(6971):247–52.

Lloyd, J. (2003).
Logic for Learning: Learning Comprehensible Theories from Structured Data.
Cognitive Technologies. Springer Berlin Heidelberg.

References

Sparkes, A., Aubrey, W., Byrne, E., Clare, A., Khan, M. N., Liakata, M., Markham, M., Rowland, J., Soldatova, L. N., Whelan, K. E., Young, M., and King, R. D. (2010).
Towards robot scientists for autonomous scientific discovery.
Autom Exp, 2:1.

Marcel **Turcotte**
Marcel.Turcotte@uOttawa.ca

School of Electrical Engineering and **Computer Science** (EECS)
University of Ottawa