



# Information Lattice Learning (ILL)

Haizi Yu

# Music Quiz 1



Soprano

Alto

Tenor

Bass

A musical score for four voices: Soprano, Alto, Tenor, and Bass. The time signature is 4/4. The Soprano part is in treble clef, Alto in treble clef, Tenor in bass clef, and Bass in bass clef. The music consists of a single melodic line for each voice, with some rests and phrasing slurs. The key signature is one sharp (F#).

Soprano

Alto

Tenor

Bass

A musical score for four voices: Soprano, Alto, Tenor, and Bass. The time signature is 4/4. The Soprano part is in treble clef, Alto in treble clef, Tenor in bass clef, and Bass in bass clef. The music is more complex than the first score, featuring many accidentals (sharps and naturals) and a dense accompaniment in the lower parts. The key signature is one sharp (F#).

# Music Quiz 2



# Music Quiz 3



# Music Quiz 4



What makes X X?

(What makes Bach's chorales Bach's chorales?)

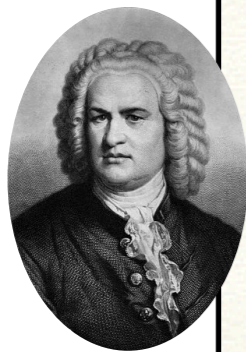
How the above question differs from ...?

# A New Learning Problem

## Automatic Concept Learning

Input

Output



A musical score for four voices: Soprano, Alto, Tenor, and Bass. The score is written in 4/4 time and consists of four staves. The Soprano staff uses a treble clef, the Alto staff uses a treble clef, the Tenor staff uses a bass clef, and the Bass staff uses a bass clef. The music is a simple melody with a final cadence.

Concepts

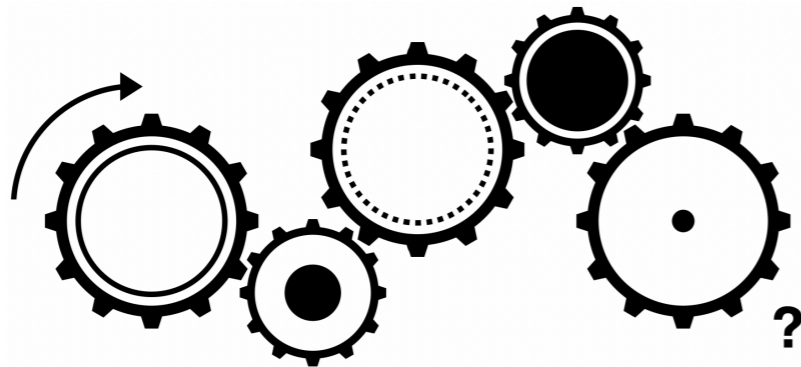
Rules

Laws

...

# Automatic Concept Learning

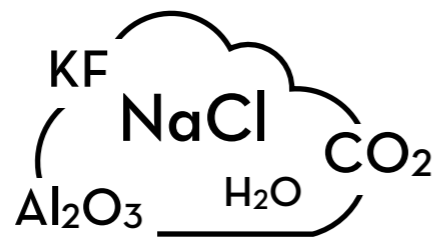
From phenomenology to theory



Parity Rule  
(clockwise iff odd)



Pythagorean Theorem  
( $a^2 + b^2 = c^2$ )



Chemical laws on compound  
formation and nomenclature

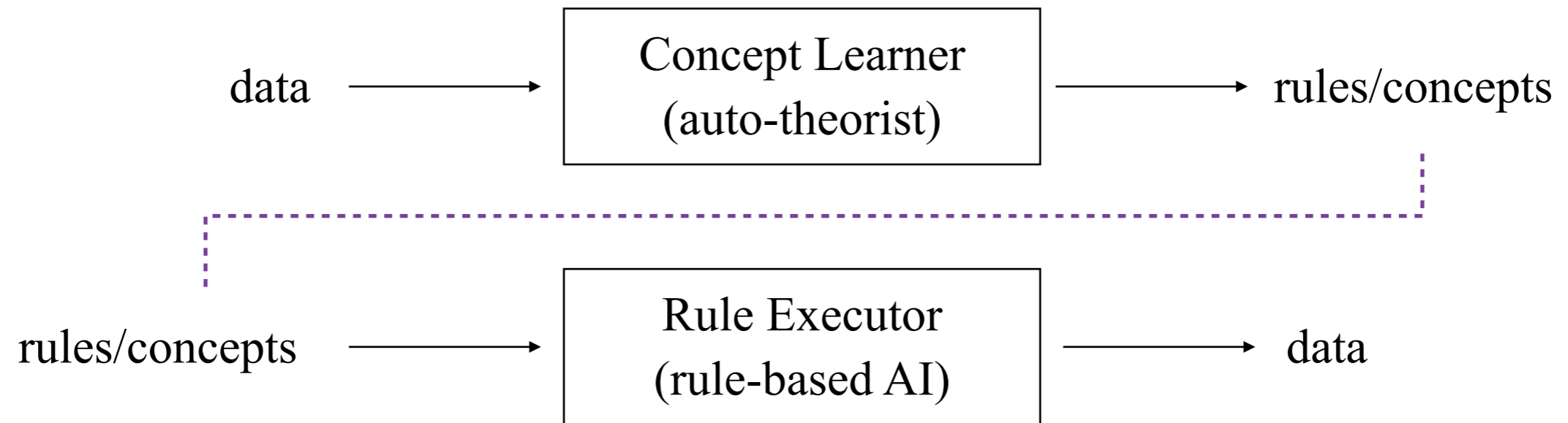
Music



Music Theory



# Concept Learning in a Nutshell



## Information Lattice Learning (ILL)



How this differs from an auto-encoder?

# A Learning Paradigm

- Representation: Information Lattice (IL)
- Algorithm: Information Lattice Learning (ILL)

# Representation: IL

# Abstraction, Concept, Rule

An **abstraction**  $\mathcal{A}$  is a partition of the data space  $X$ .

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

$$\mathcal{A} = \{\{x_1, x_6\}, \{x_3\}, \{x_2, x_4, x_5\}\}$$

cells (or less formally, clusters)



concepts

A **concept** is a partition cell.

$$X = \{\text{every vertebrate}\}$$

$$\mathcal{A} = \{\text{mammals, birds, fish, amphibians, reptiles}\}$$

# Abstraction, Concept, Rule

An **abstraction**  $\mathcal{A}$  is a partition of the data space  $X$ .

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

$$\mathcal{A} = \{\{x_1, x_6\}, \{x_3\}, \{x_2, x_4, x_5\}\}$$

cells (or less formally, clusters)

⋮  
↓  
concepts

A **concept** is a partition cell.

$$X = \{ \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array} \} \text{ (treble clef)}$$

$$\mathcal{A} = \{ \{ \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array} \}, \{ \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \end{array} \}, \{ \begin{array}{|c|} \hline \text{8} \\ \hline \end{array} \} \}$$

major

minor

diminished

# Abstraction, Concept, Rule

An **abstraction**  $\mathcal{A}$  is a partition of the data space  $X$ .

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

$$\mathcal{A} = \{\{x_1, x_6\}, \{x_3\}, \{x_2, x_4, x_5\}\}$$

cells (or less formally, clusters)

⋮  
↓  
concepts

A **concept** is a partition cell.

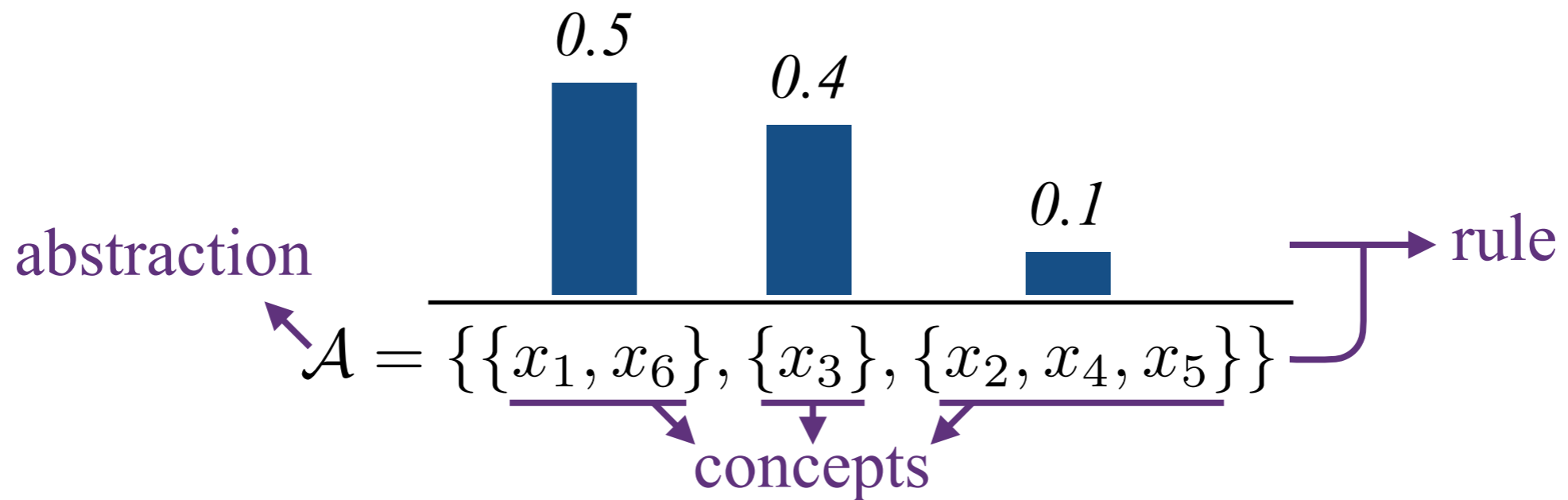
$$X = \{ \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array} \} \text{ (treble clef)}$$

$$\mathcal{A}' = \{ \{ \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array} \}, \{ \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array} \}, \{ \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{8} \\ \hline \text{8} \\ \hline \end{array} \} \}$$

root position    1st inversion    2nd inversion

# Abstraction, Concept, Rule

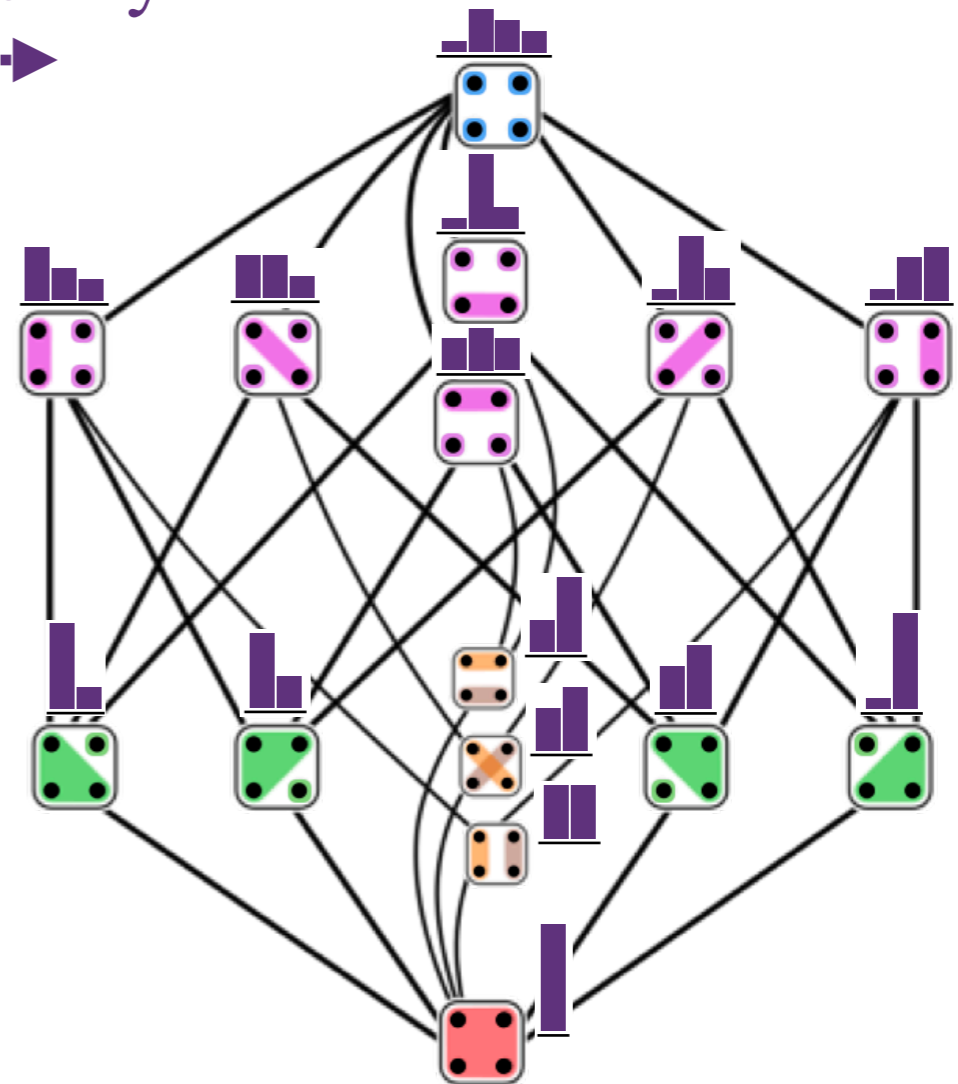
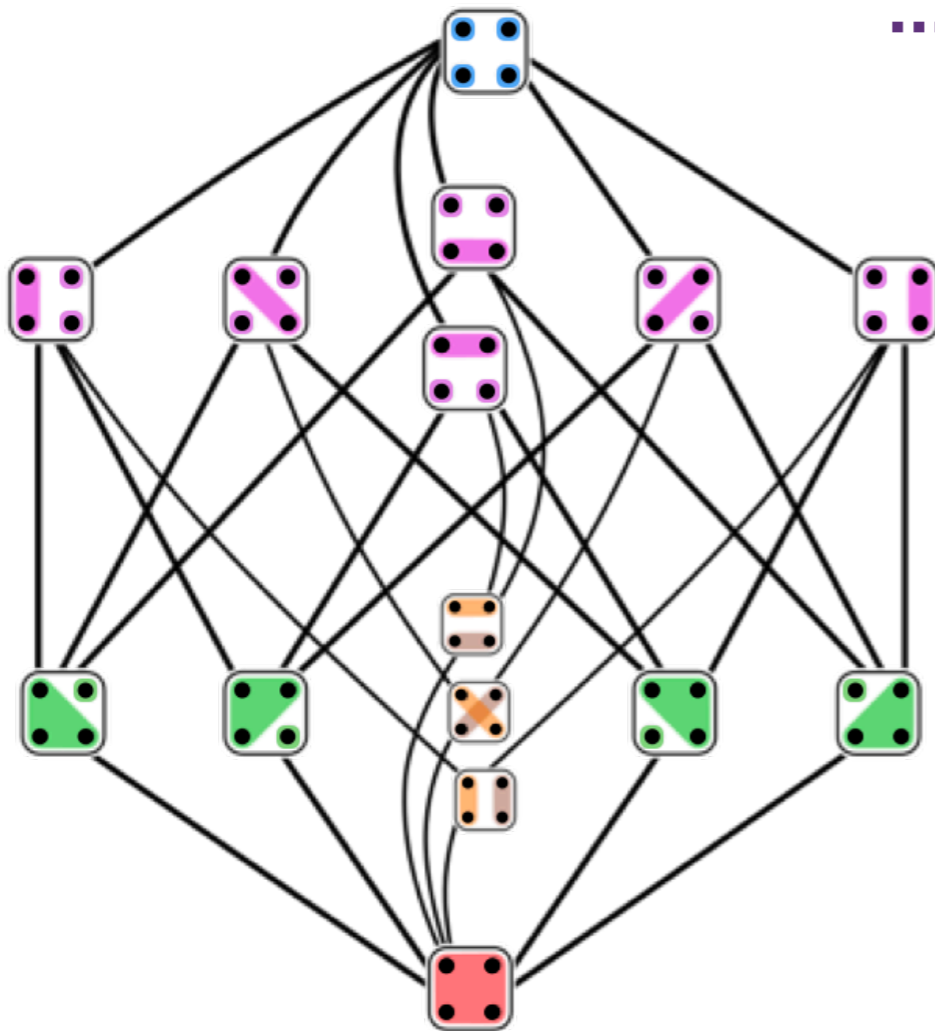
A **(probabilistic) rule** is a probability distribution projected onto a partition



# Partition Lattice

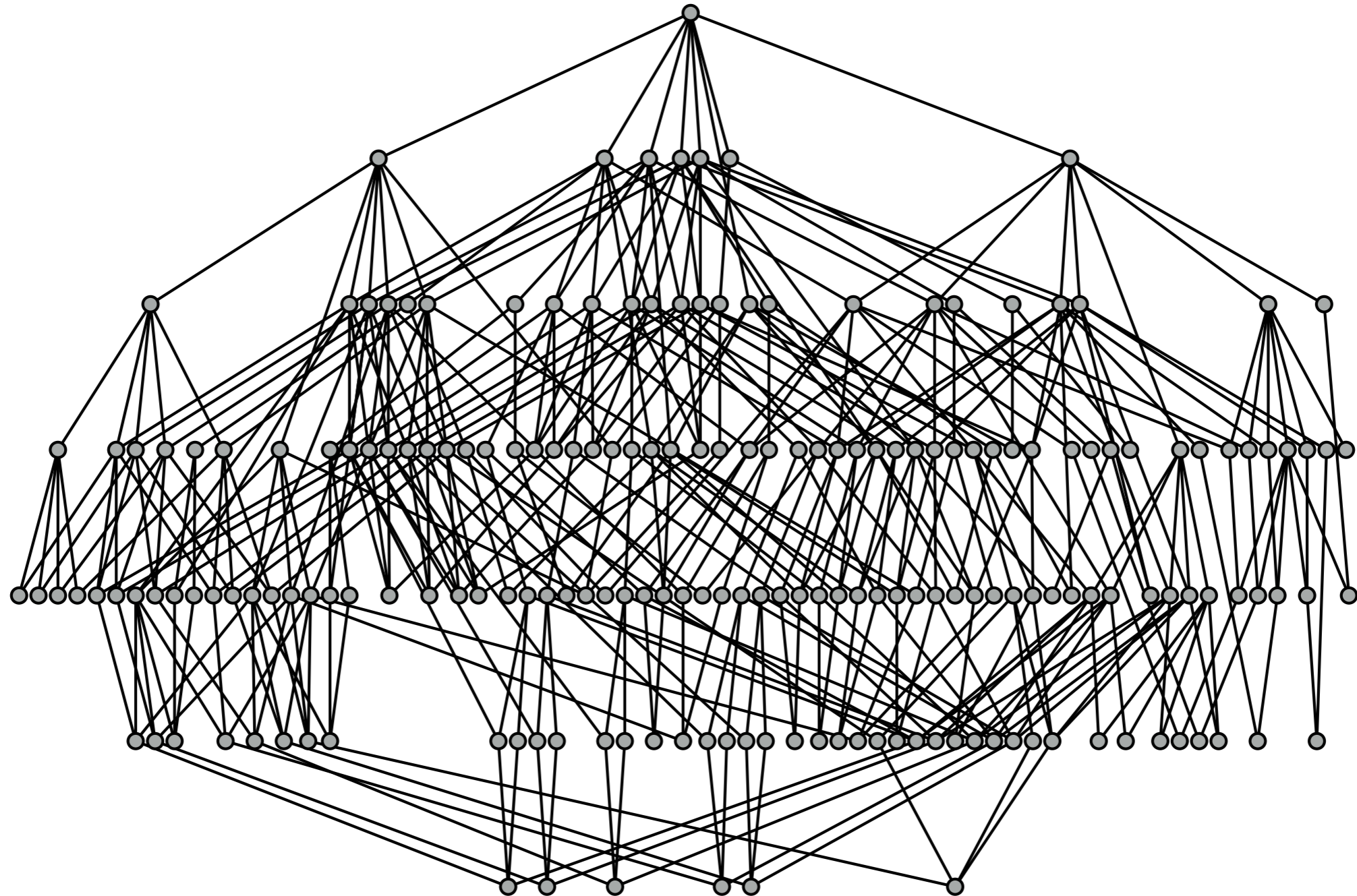
# Information Lattice

project probability

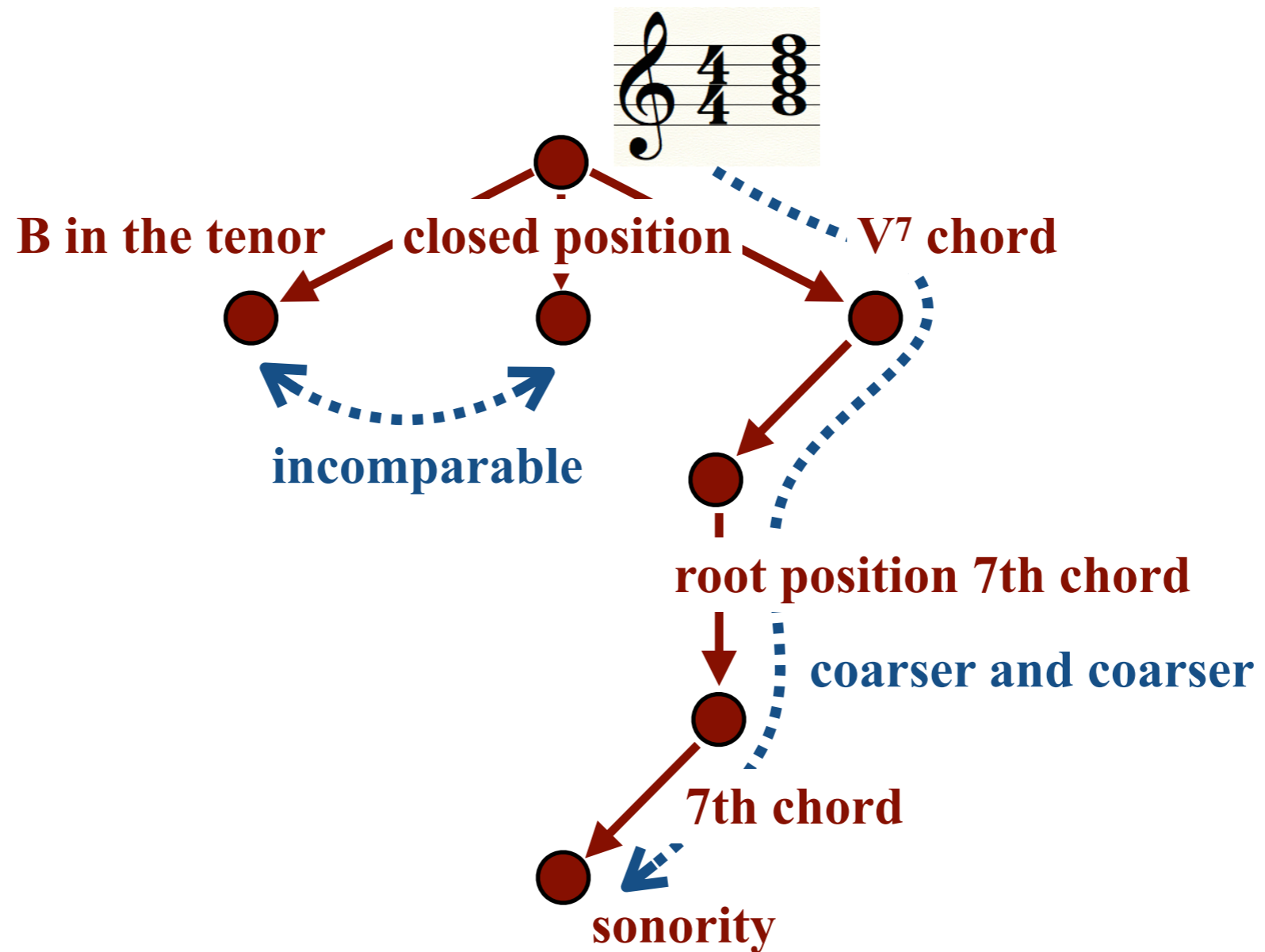




# Bach's Information Lattice



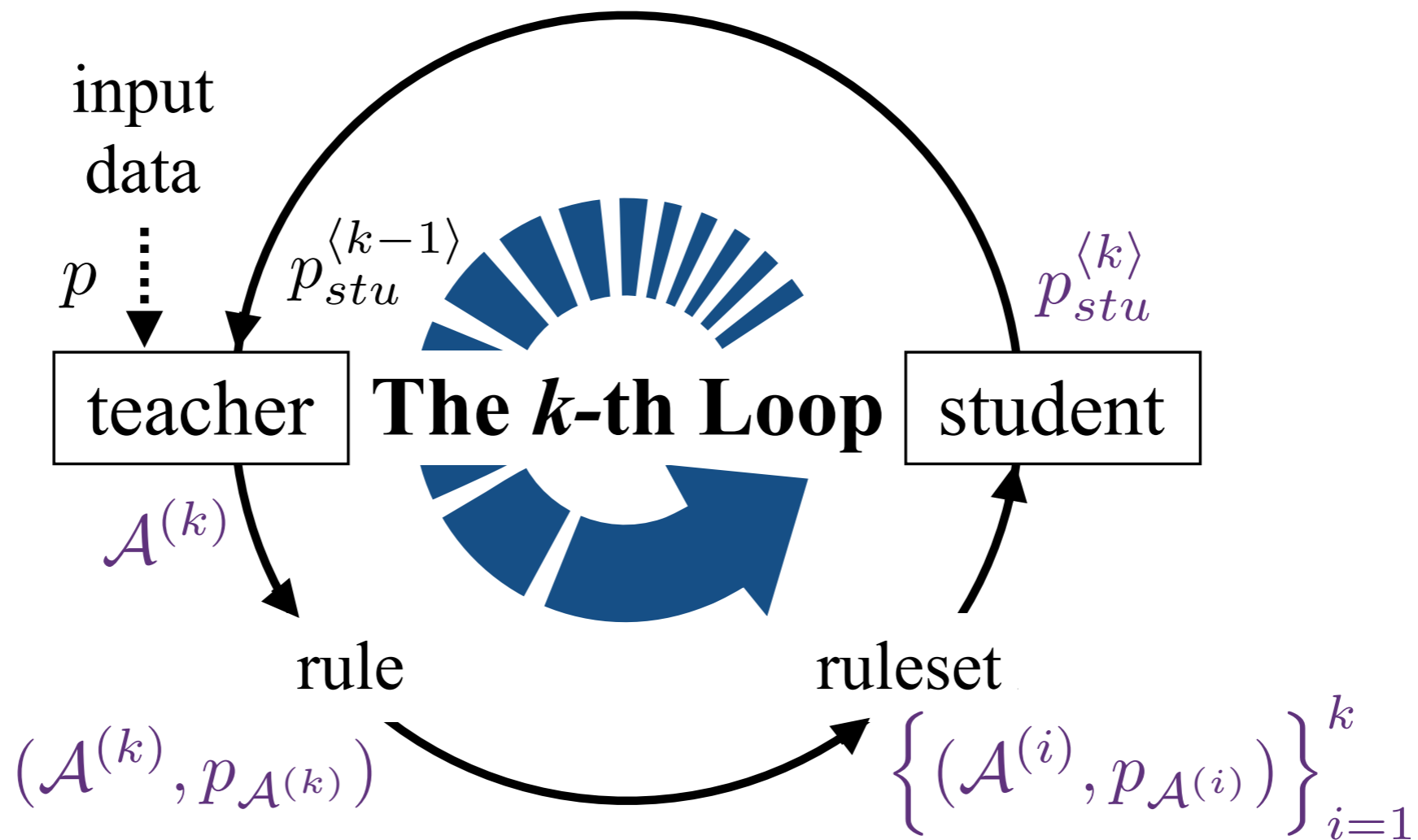
# Bach's Concept Lattice



# Algorithm: ILL

# The Self-Learning Loop

A Teacher-Student Architecture: Learning by Comparison



# Teacher: a Discriminative Model

The teacher solves an optimization problem:

$$\begin{aligned} & \underset{\mathcal{A} \in \mathfrak{P}_X}{\text{maximize}} && D_{KL} \left( p_{\mathcal{A},stu}^{\langle k-1 \rangle} \parallel p_{\mathcal{A}} \right) \\ & \text{subject to} && \mathcal{A} \notin \mathfrak{P}^{\langle k-1 \rangle} \\ & && H(p_{\mathcal{A}}) \leq \delta_k \end{aligned}$$

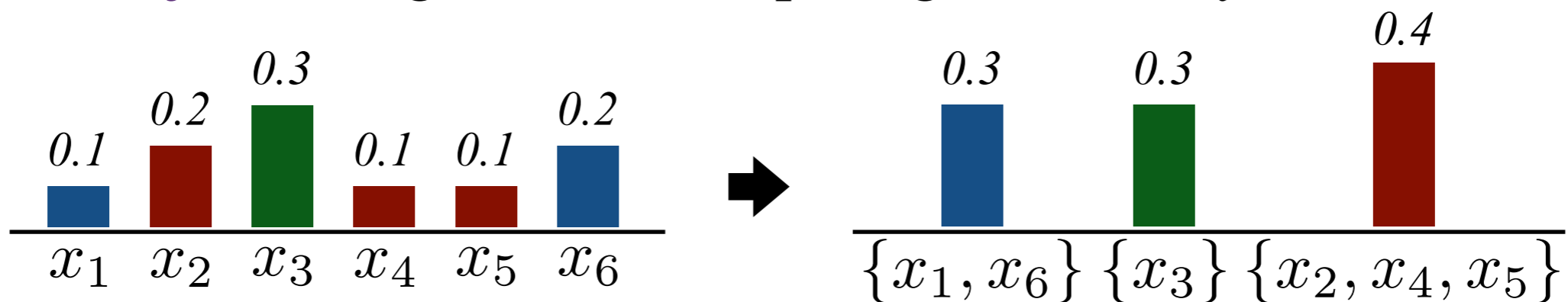
Want to find the abstraction which reveals the **largest statistical difference** between the student and the input data.

# Teacher: a Discriminative Model

The teacher solves an optimization problem:

$$\begin{aligned} & \underset{\mathcal{A} \in \mathfrak{P}_X}{\text{maximize}} && D_{KL} \left( p_{\mathcal{A},stu}^{\langle k-1 \rangle} \parallel p_{\mathcal{A}} \right) \\ & \text{subject to} && \mathcal{A} \notin \mathfrak{P}^{\langle k-1 \rangle} \\ & && H(p_{\mathcal{A}}) \leq \delta_k \end{aligned}$$

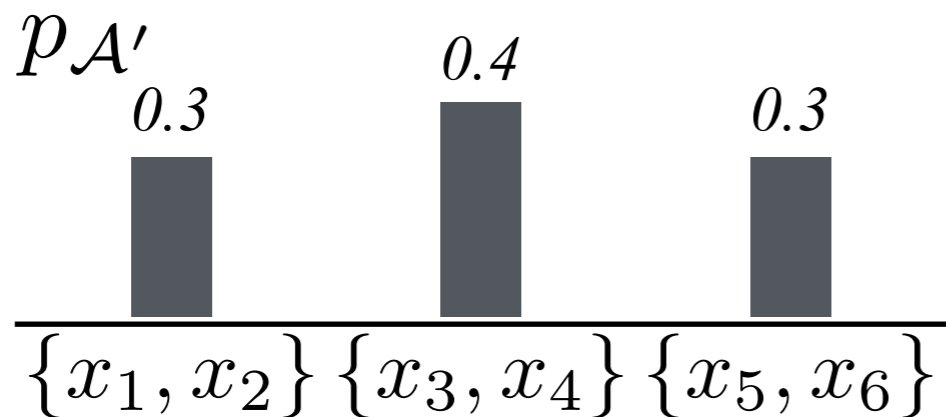
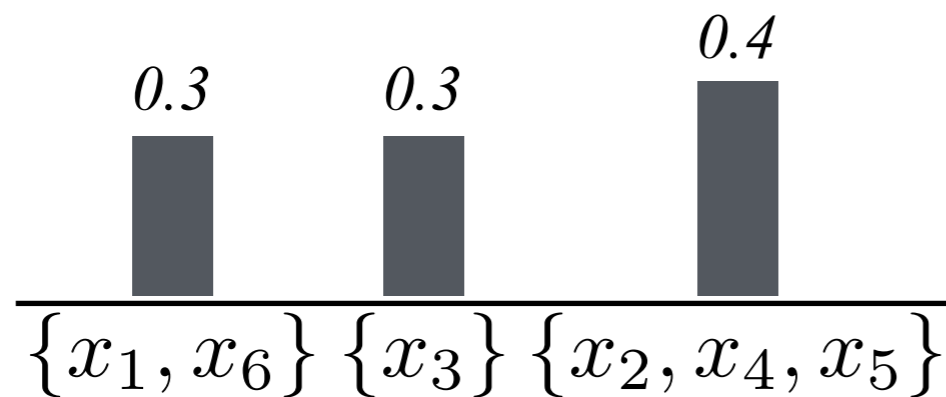
**Projection:** given  $\mathcal{P}$ , computing  $p_{\mathcal{A}}$  is easy:



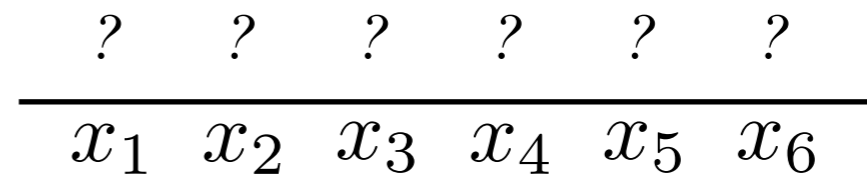
# Student: a Generative Model

Apply probabilistic rules, which is known as the **rule realization** problem.

Given  $\mathcal{P}_A$ , compute  $\mathcal{P}$ :



...



not necessarily unique  
which one do we prefer?



# Student: a Generative Model

The student solves a MaxEnt problem:

$$\begin{aligned} & \underset{p_{stu}^{\langle k \rangle} \in \Delta^{|X|}}{\text{maximize}} && S_q(p_{stu}^{\langle k \rangle}) := (q - 1)^{-1} \left( 1 - \|p_{stu}^{\langle k \rangle}\|_q^q \right) \\ & \text{subject to} && A^{(i)} p_{stu}^{\langle k \rangle} = p_{\mathcal{A}^{(i)}}, \quad i = 1, \dots, k \end{aligned}$$

Want to find the most **creative** probabilistic model which satisfies all the rules (**quality**) while at the same time enables **novelty**



How MUS-ROVER Self-Evolves?

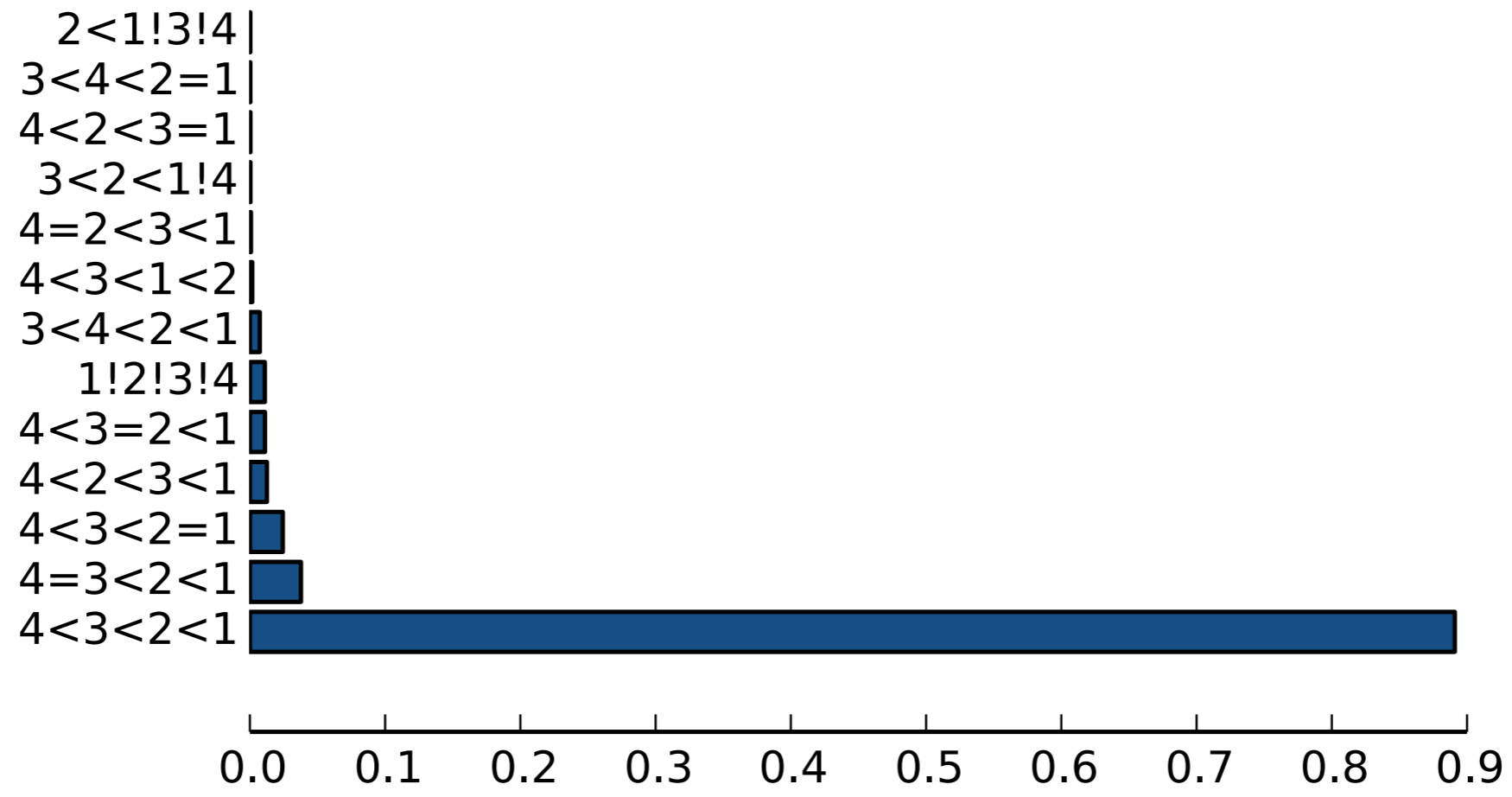
1-gram

# Student 0

A musical score for four voices: Soprano, Alto, Tenor, and Bass. The score is written in 4/4 time and features a key signature of one sharp (F#). The Soprano part is in a treble clef, while the Alto, Tenor, and Bass parts are in bass clefs. The Alto part includes a vocal line and a piano accompaniment line. The Tenor and Bass parts also include vocal lines and piano accompaniment lines. The music consists of several measures of music, with a final double bar line at the end of the piece.

1-gram

# Rule 1: $\text{order} \circ w_{1,2,3,4}$



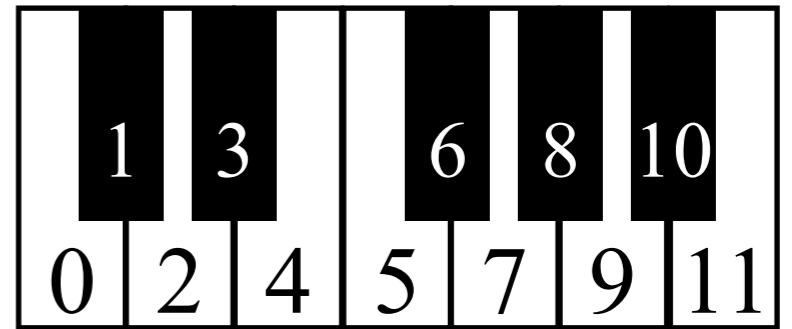
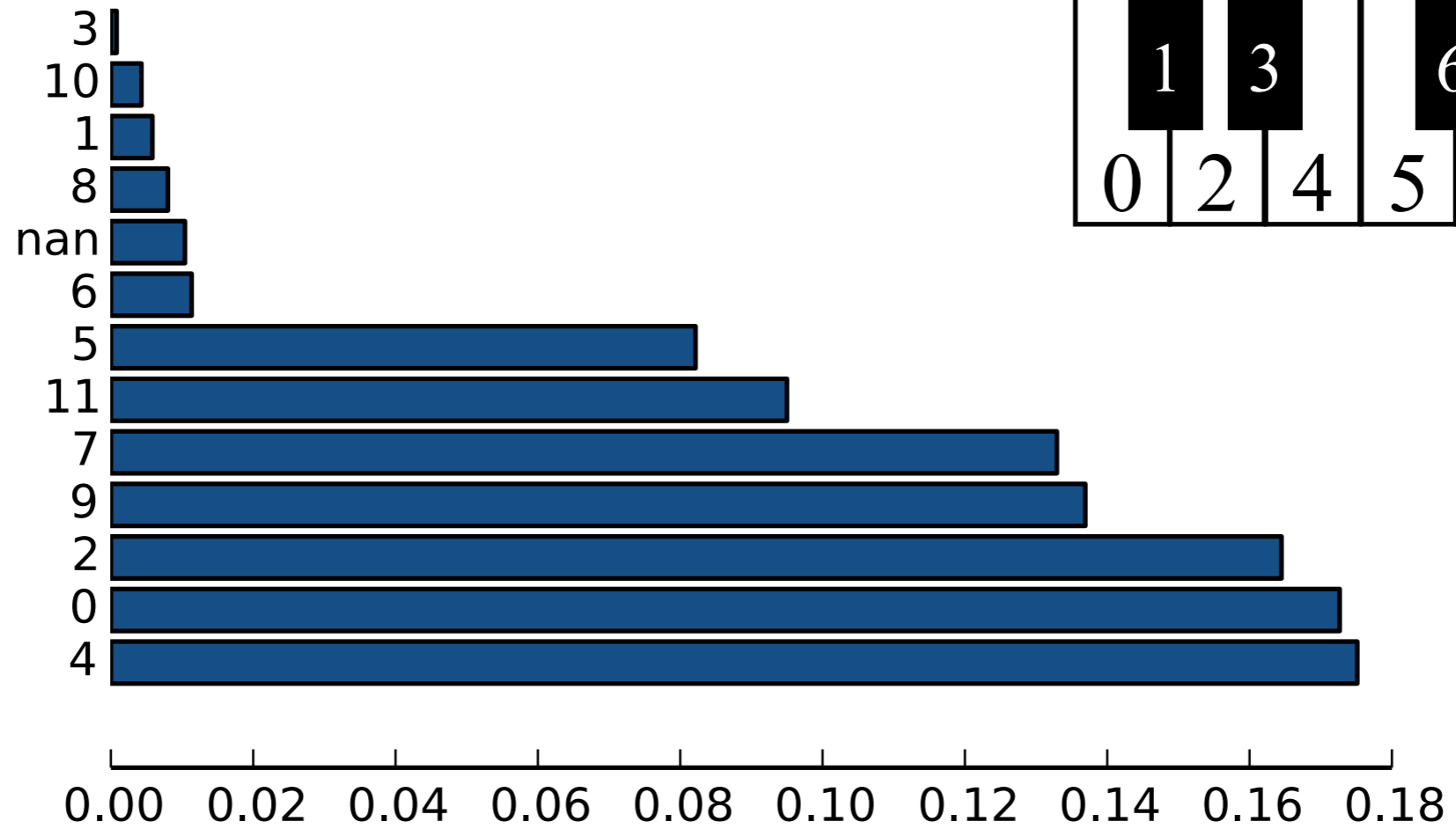
1-gram

# Student 1

The image displays a musical score for a four-part vocal ensemble, labeled 'Student 1'. The score is written in 4/4 time and consists of four staves: Soprano, Alto, Tenor, and Bass. The key signature is one sharp (F#), and the time signature is 4/4. The Soprano part begins with a treble clef and a key signature of one sharp. The Alto part also begins with a treble clef and a key signature of one sharp. The Tenor part begins with a bass clef and a key signature of one sharp. The Bass part begins with a bass clef and a key signature of one sharp. The music is a single melodic line for each voice part, with various intervals and accidentals. The Soprano part starts on a high note, the Alto on a middle note, the Tenor on a lower note, and the Bass on the lowest note. The music concludes with a double bar line at the end of each staff.

1-gram

Rule 2:  $\text{mod}_{12} \circ w_1$



1-gram

# Student 2

The image displays a musical score for four voices: Soprano, Alto, Tenor, and Bass. The score is written in 4/4 time and consists of four staves. The Soprano staff uses a treble clef and contains a melodic line of quarter notes. The Alto staff uses a treble clef and contains a melodic line with some chromaticism. The Tenor staff uses a bass clef and contains a melodic line with some chromaticism. The Bass staff uses a bass clef and contains a melodic line with some chromaticism. The key signature is one sharp (F#), and the time signature is 4/4. The score is presented on a light yellow background.

1-gram

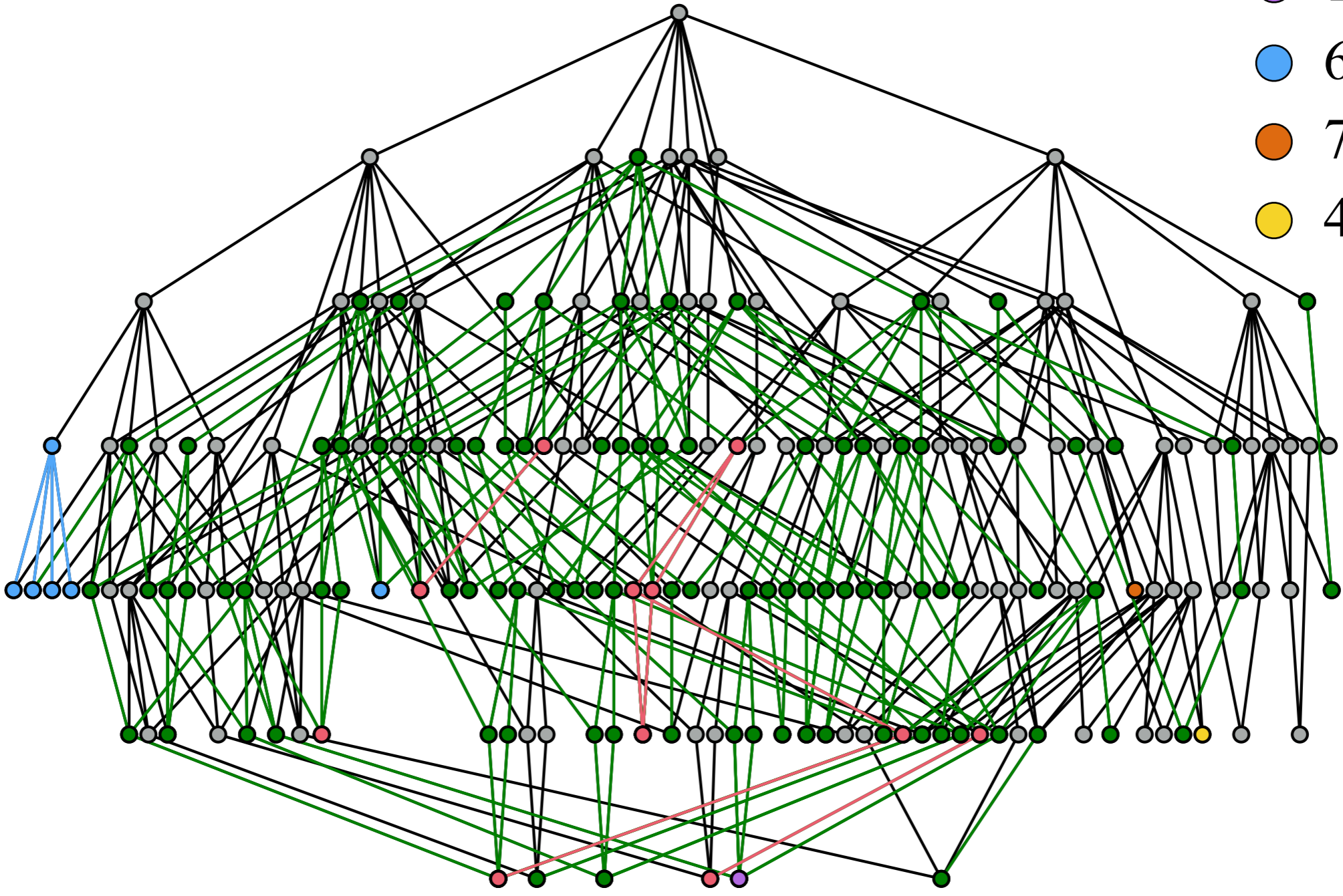
# Student 22

The image displays a musical score for a four-part vocal ensemble, labeled 'Student 22'. The score is written in 4/4 time and consists of four staves: Soprano, Alto, Tenor, and Bass. Each staff begins with a treble clef for Soprano and Alto, and a bass clef for Tenor and Bass. The key signature is one sharp (F#), and the time signature is 4/4. The melody for all parts is a simple, stepwise line. The Soprano part starts on G4 and ends on G4. The Alto part starts on E4 and ends on E4. The Tenor part starts on C3 and ends on C3. The Bass part starts on G2 and ends on G2. The notes are: Soprano (G4, A4, B4, C5, B4, A4, G4); Alto (E4, F4, G4, A4, B4, C5, B4, A4, G4); Tenor (C3, D3, E3, F3, G3, A3, B3, C4); Bass (G2, A2, B2, C3, D3, E3, F3, G3). There is a sharp sign (#) on the final note of the Alto and Bass staves, indicating a sharp (F#).

n-gram

# End of Loop 10

- unlearned
- 1-gram
- 3-gram
- 10-gram
- 6-gram
- 7-gram
- 4-gram

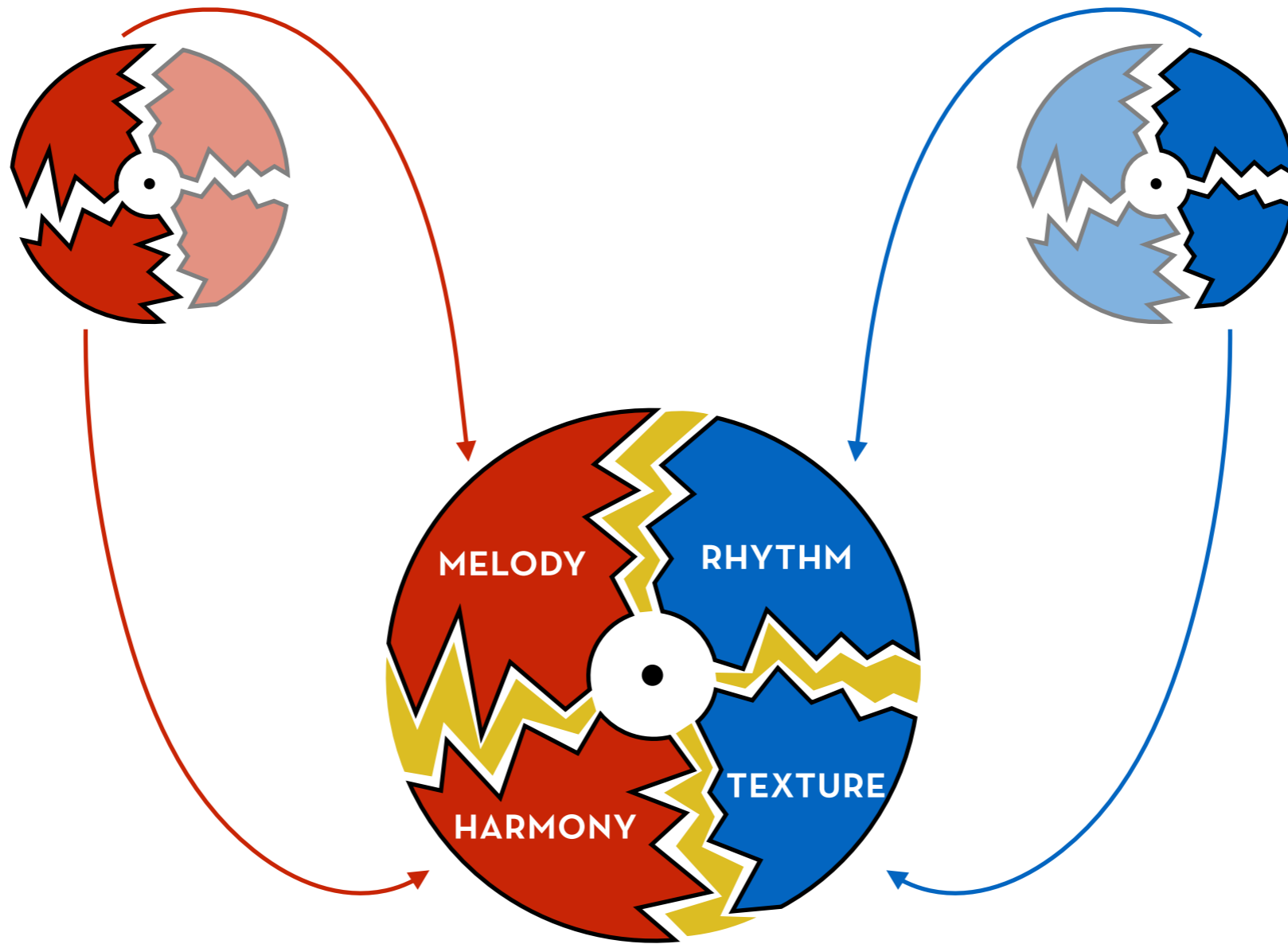




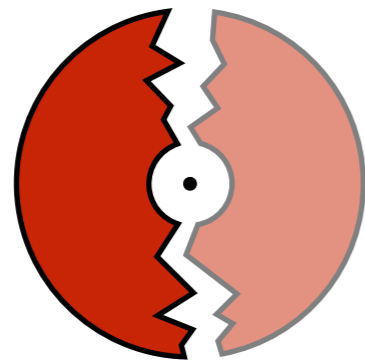
# Music Mosaic

Happy  
Birthday

Mozart's  
K545



Happy  
Birthday



Beethoven's  
5th Symphony

