

Deep Learning Techniques for Music Generation

3. Generation by Feedforward Architectures

Jean-Pierre Briot

`Jean-Pierre.Briot@lip6.fr`

Laboratoire d'Informatique de Paris 6 (LIP6)
Sorbonne Université – CNRS

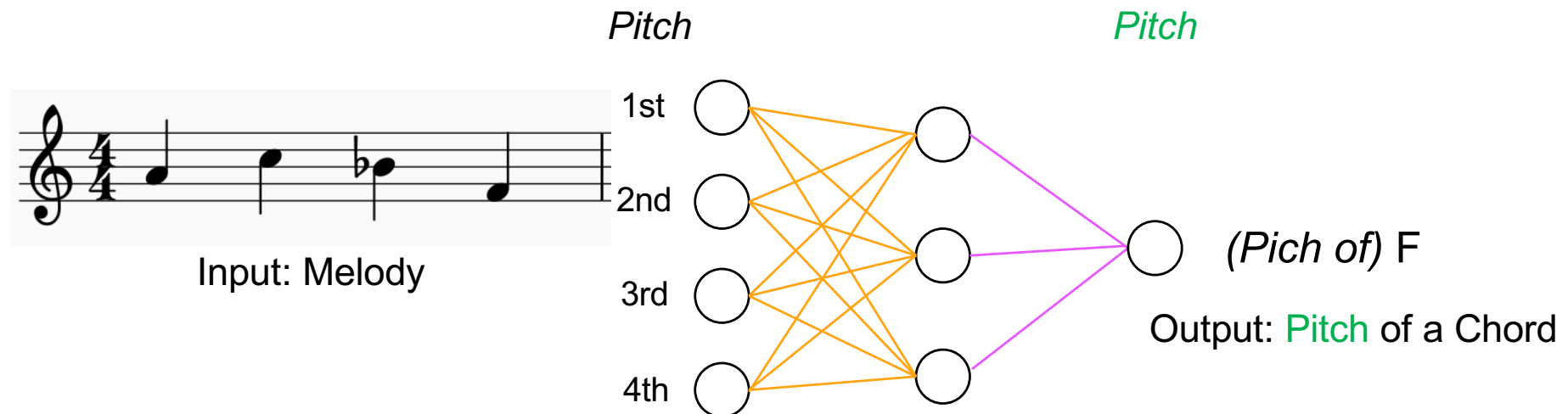


Programa de Pós-Graduação em Informática (PPGI)
UNIRIO



Direct Use – Feedforward – Ex 1

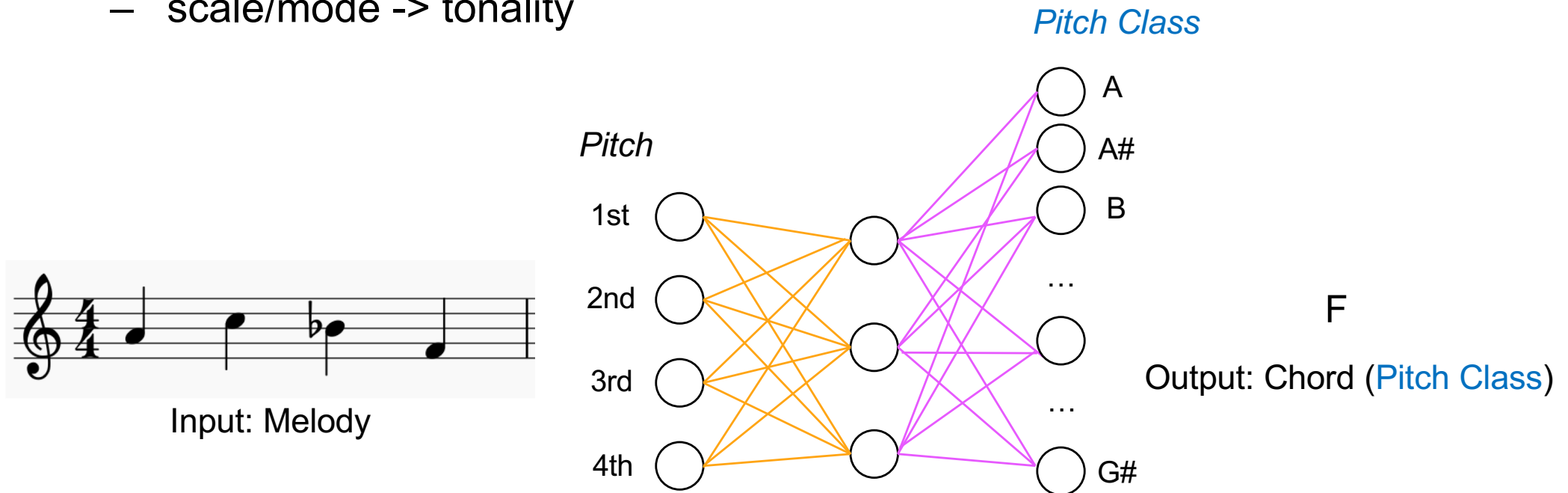
- Feedforward Architecture
- Prediction Task
- Ex1: Predicting a chord associated to a melody segment
 - scale/mode -> tonality



- Training on a corpus/dataset <melody, chord>
- Production (Prediction)

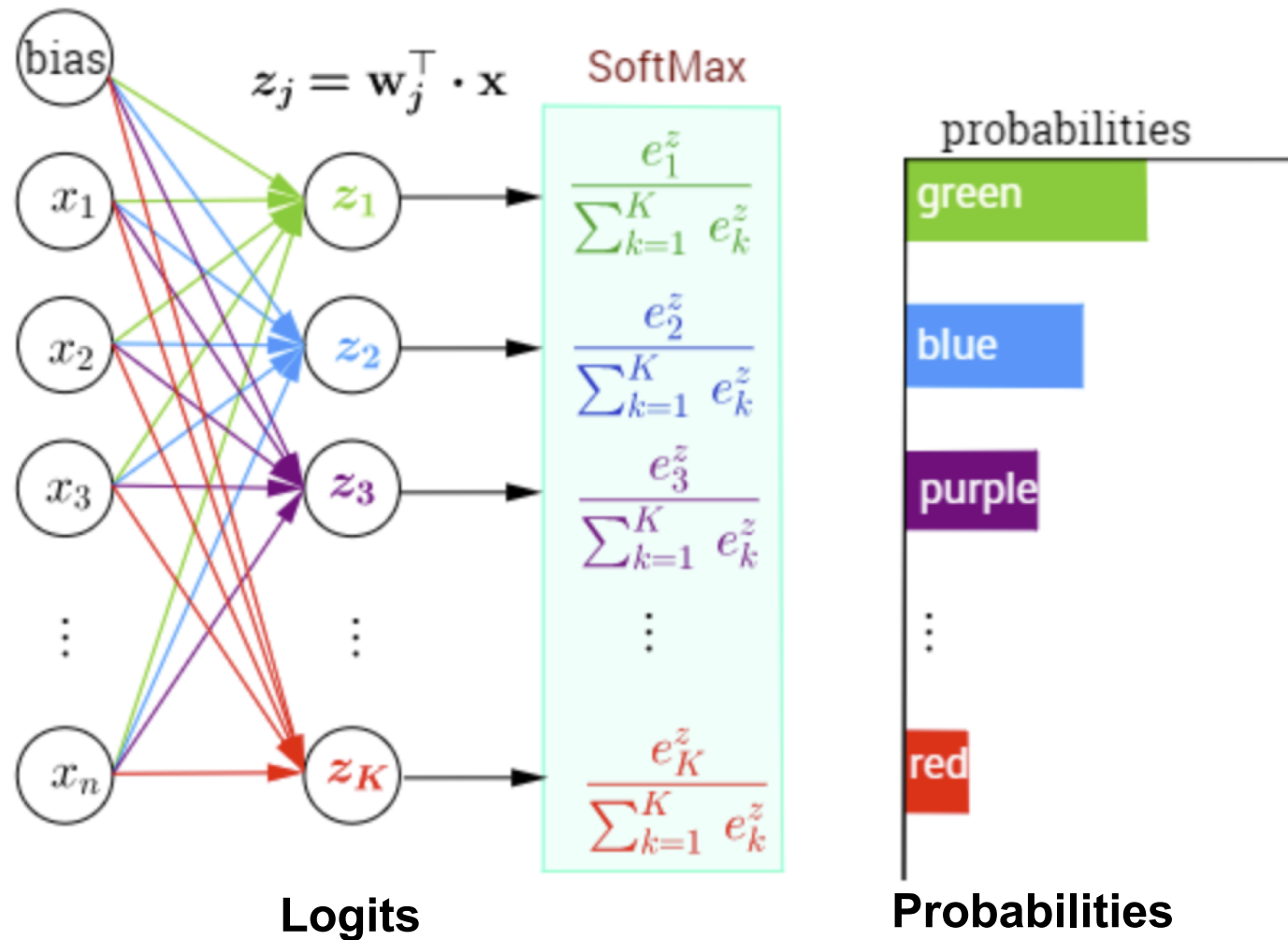
Direct Use – Feedforward – Ex 1

- Feedforward Architecture
- **Classification** Task
- Ex1: Predicting a chord associated to a melody segment
 - scale/mode -> tonality



- Training on a corpus/dataset <melody, chord>
- Production (Classification)

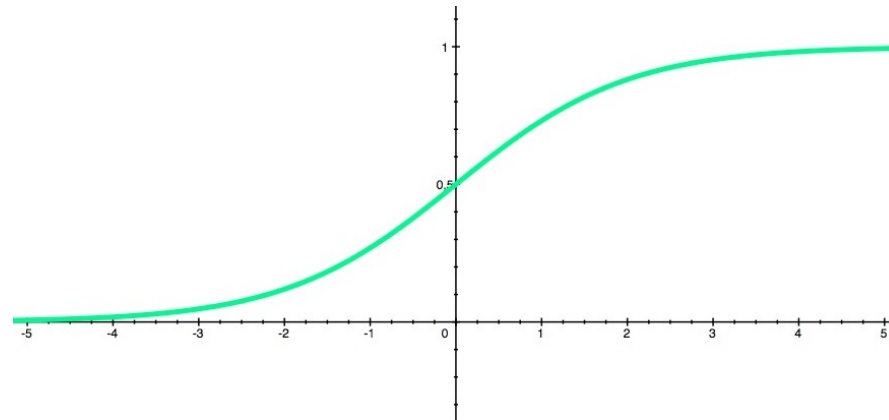
Softmax



Softmax and Sigmoid

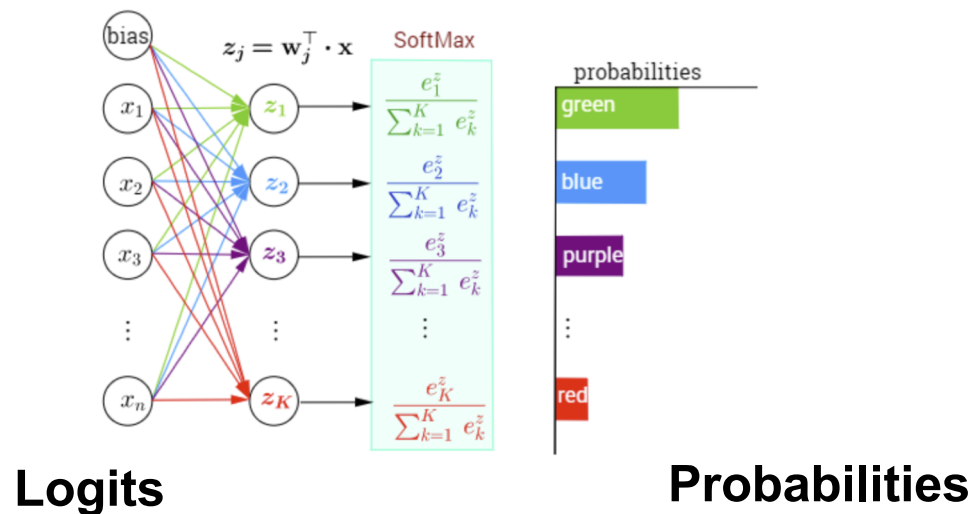
- Softmax is the generalization of Sigmoid
- From Binary classification to Categorical (Multiclass) classification

Sigmoid



Probability $\in [0, 1]$

Softmax



$\sum \text{Probabilities} = 1$

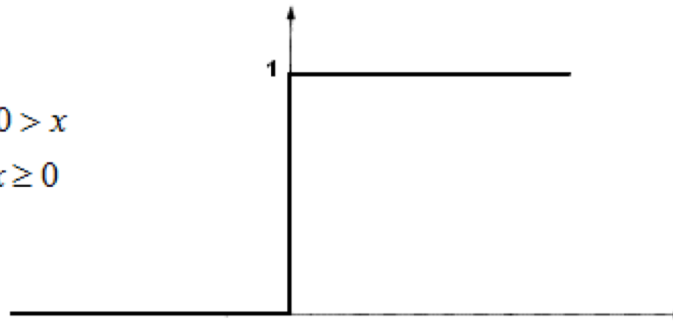
Softmax and Sigmoid

- Step function and Argmax are NOT differentiable
- No gradient -> No possibility of back propagation

Step function
(Perceptron)

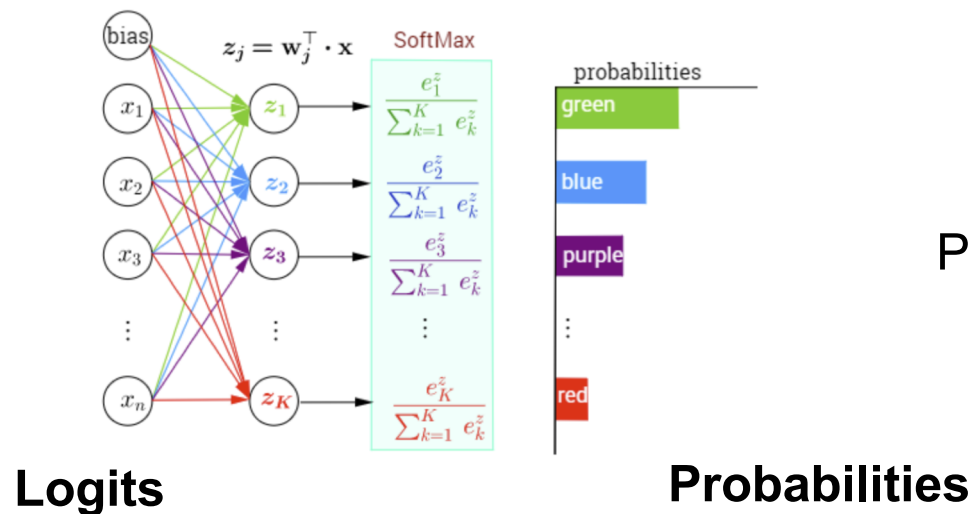
$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

Unit step (threshold)



Probability $\in \{0, 1\}$

Argmax

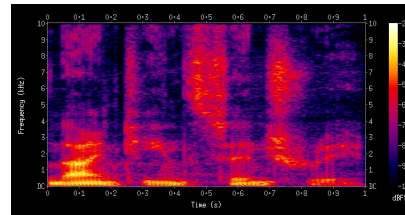


Probability(Argmax) = 1

Representation

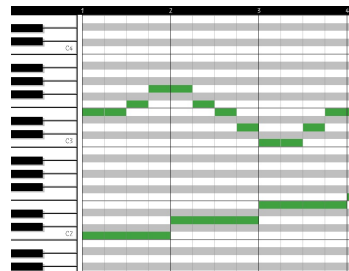
- Audio

- Waveform
- Spectrogram (Fourier Transform)
- Other (ex: MFCC)



- Symbolic

- Note
- Rest
- Note hold
- Duration
- Chord
- Rhythm
- Piano Roll
- MIDI
- ABC, XML...



Representation



Score

C				
B				
A#				
A				
G#				
G				

Piano Roll

B ₄	0	B ₄	0	B ₄	0	B ₄	0
A# ₄	0	A# ₄	0	A# ₄	0	A# ₄	0
A ₄	0	A ₄	1	A ₄	0	A ₄	0
G# ₄	0	G# ₄	0	G# ₄	0	G# ₄	0
G ₄	0	G ₄	0	G ₄	0	G ₄	0
F# ₄	0	F# ₄	0	F# ₄	0	F# ₄	0
F ₄	0	F ₄	0	F ₄	0	F ₄	0
E ₄	0	E ₄	0	E ₄	0	E ₄	0
D# ₄	0	D# ₄	0	D# ₄	0	D# ₄	0
D ₄	0	D ₄	0	D ₄	0	D ₄	0
C# ₄	0	C# ₄	0	C# ₄	0	C# ₄	0
C ₄	0	C ₄	0	C ₄	1	C ₄	1
one-hot		one-hot		one-hot		one-hot	

One hot Encoding

Encoding of Features (ex : Note Pitch)

- Value
 - *Analogic*
- One-Hot
 - *Digital*
- Embedding
 - *Constructed*

A₄ 440.0
scalar continuous

A₄ 69
scalar discrete

B ₄	0
A# ₄	0
A ₄	1
G# ₄	0
G ₄	0
F# ₄	0
F ₄	0
E ₄	0
D# ₄	0
D ₄	0
C# ₄	0
C ₄	0

one-hot

	0
	0
A ₄	1
	0
	0
	0
F ₄	1
	0
	0
D ₄	1
	0
	0

many-hot

	0
	0
A ₄	1
	0
	0
	0
	0
	0
	0
	0
	0
	0

	0
	0
A ₄	1
	0
	0
	0
F ₄	1
	0
	0
D ₄	1
	0
	0

	0
	0
	0
D ₃	1
	0
	0
	0

multi-one-hot

	0
	0
	0
	0
	0
	0
C ₃	1

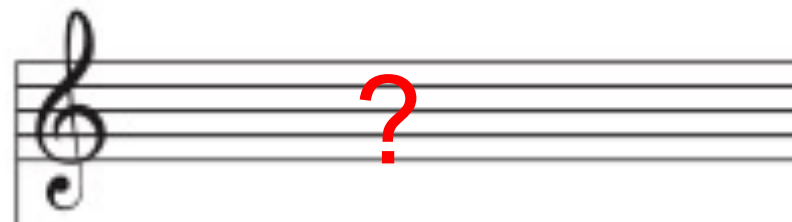
multi-many-hot

Encoding

- Rest
 - Zero-hot
 - » But ambiguity with low probability notes
 - One more one-hot element
 - ...
- Hold
 - One more one-hot element
 - » But only for monophonic melodies
 - Replay matrix
 - ...

Representation

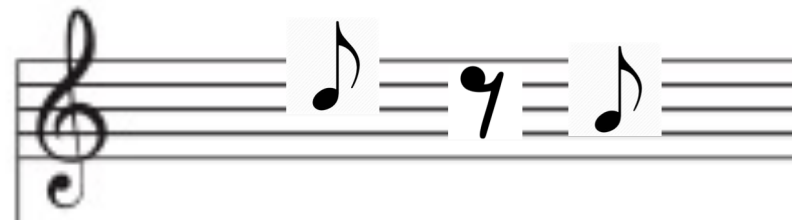
hold	0	1	0	1	0	1
rest	0	0	1	0	0	0
	0	0	0	0	0	0
	0	0	0	0	0	0
A					1	
C	1					
	0	0	0	0	0	0
	0	0	0	0	0	0



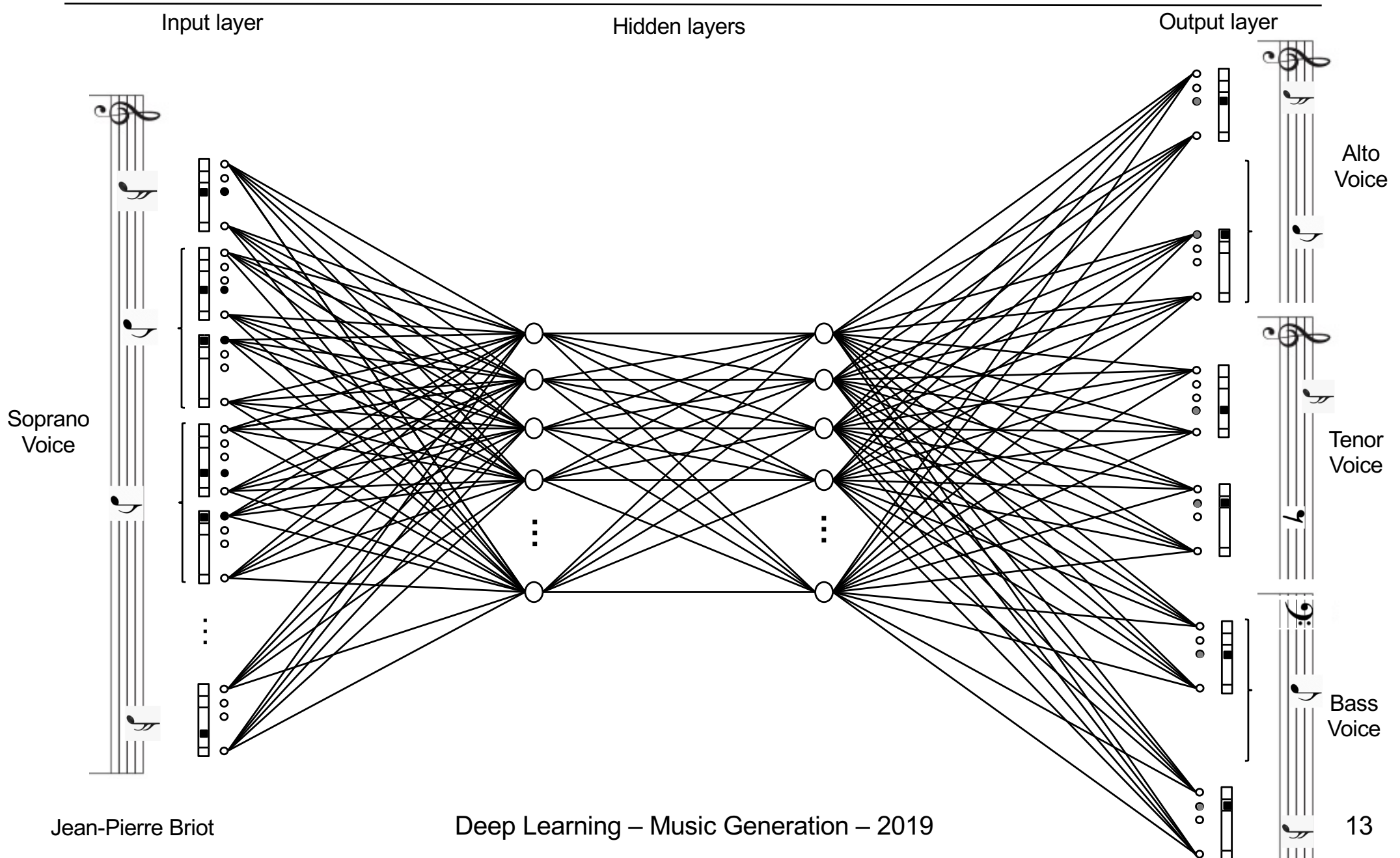
Representation

hold	0	1	0	1	0	1
rest	0	0	1	0	0	0
	0	0	0	0	0	0
	0	0	0	0	0	0
A					1	
C	1					
	0	0	0	0	0	0
	0	0	0	0	0	0

If time slice = sixteenth



Music / Representation / Network

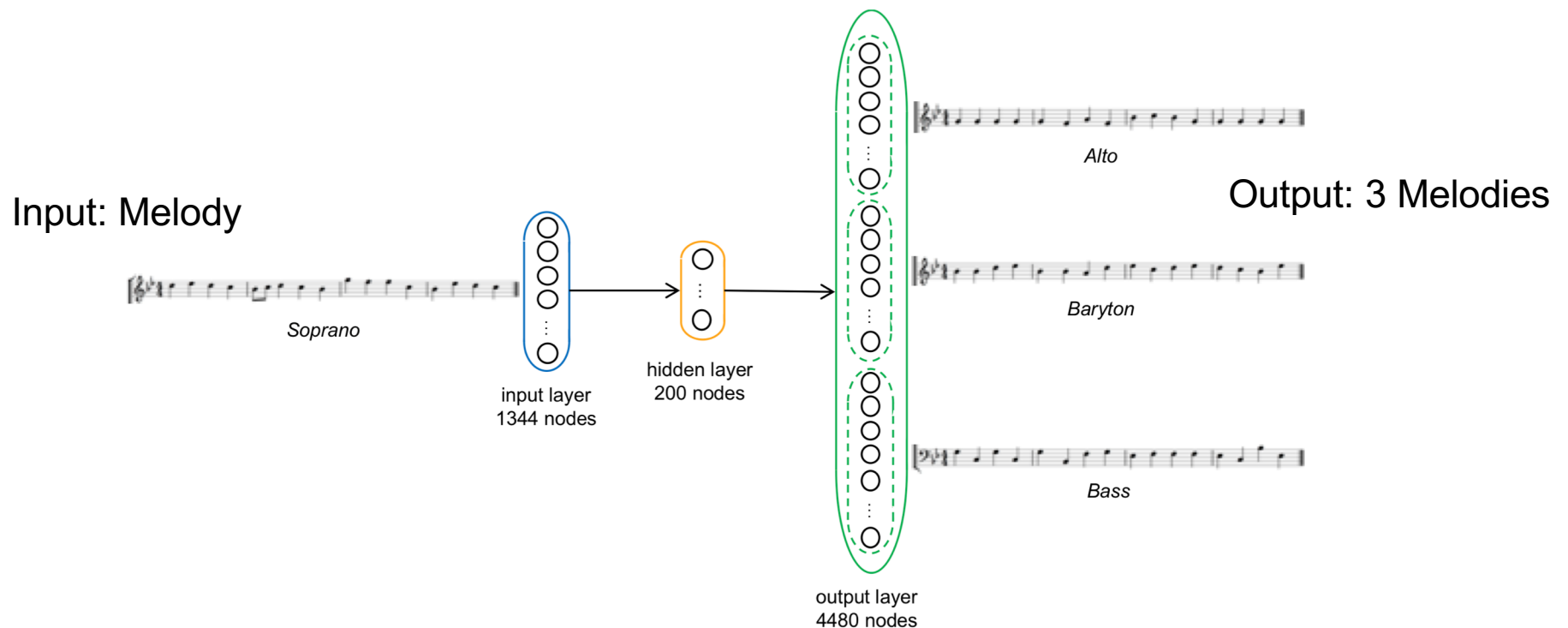


Code

- Python (3) 
- Keras 
- Theano  or TensorFlow 
- Music21 

Direct Use – Feedforward – Ex 2: ForwardBach

- Feedforward Architecture
- Prediction Task
- Ex2: Counterpoint (Chorale) generation
- Training on the set of (389) J.S. Bach Chorales (Choral Gesang)



ForwardBach



Bach BWV 344 Chorale
(Training Example)

Original



Regenerated



ForwardBach



Bach BWV 423 Chorale
(Test Example)

Original

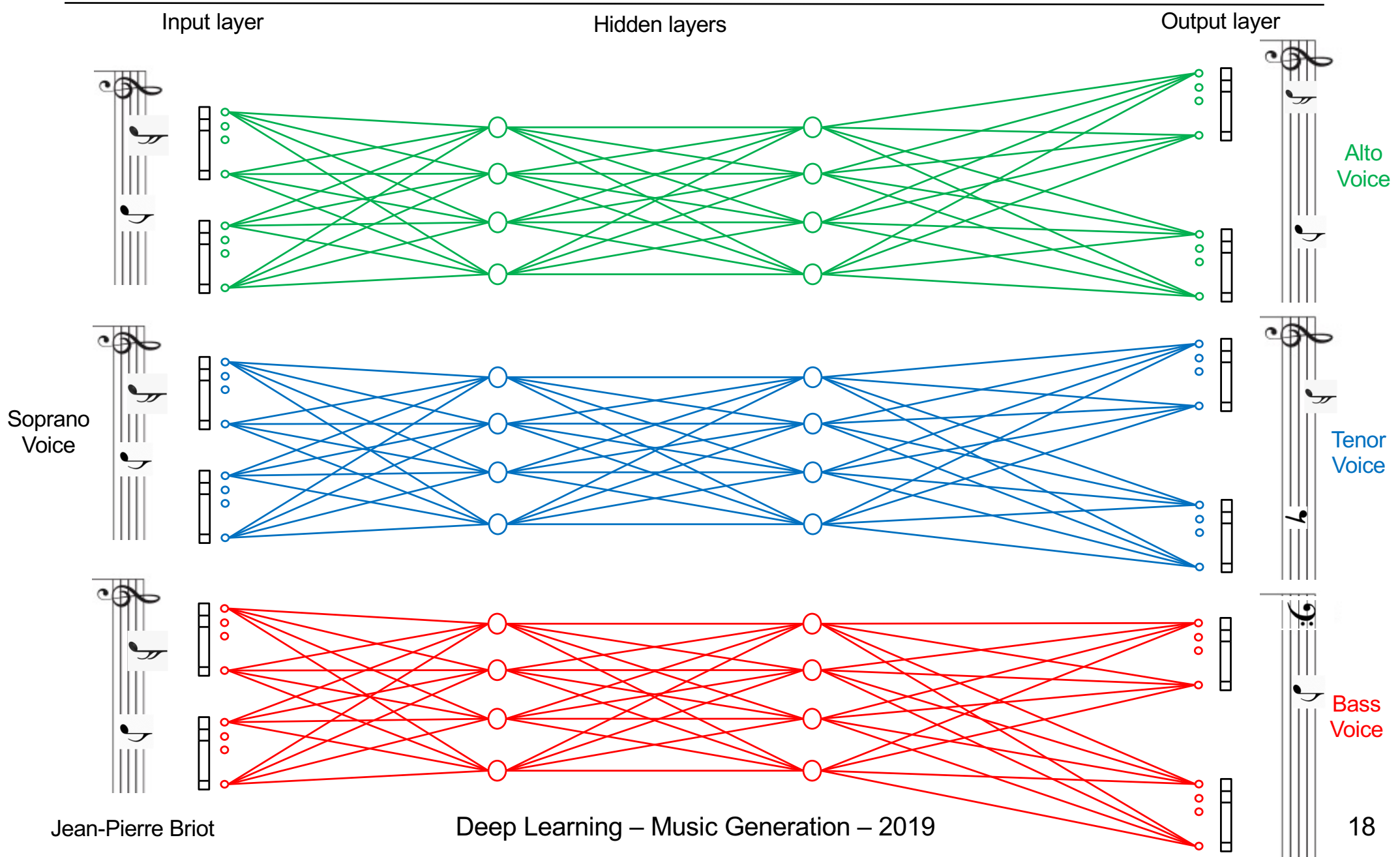


Regenerated



Music / Representation / Network

Alternative 3 Models Architecture [Cotrim & Briot, 2019]



Forward3Bach [Cotrim & Briot, 2019]



Bach BWV 423 Chorale
(Test Example)

Original



Single Architecture
Regenerated



Triple Architecture
Regenerated



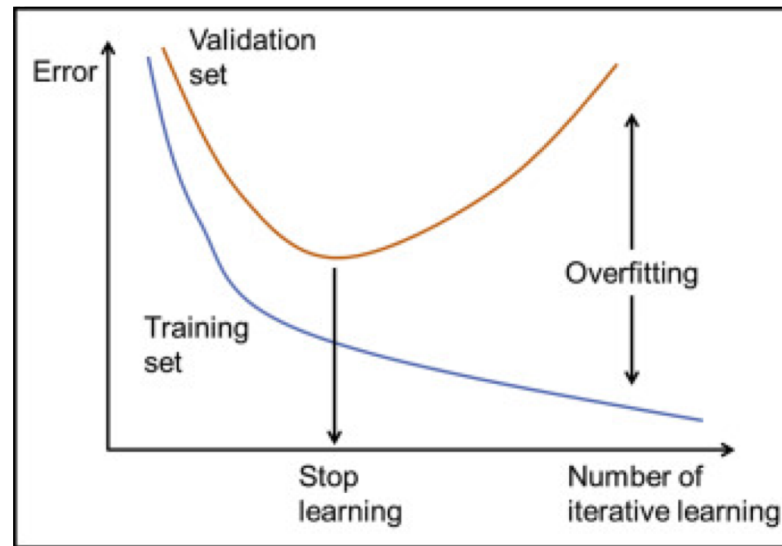
Comparison ? What happened ?

Overfitness Limitations

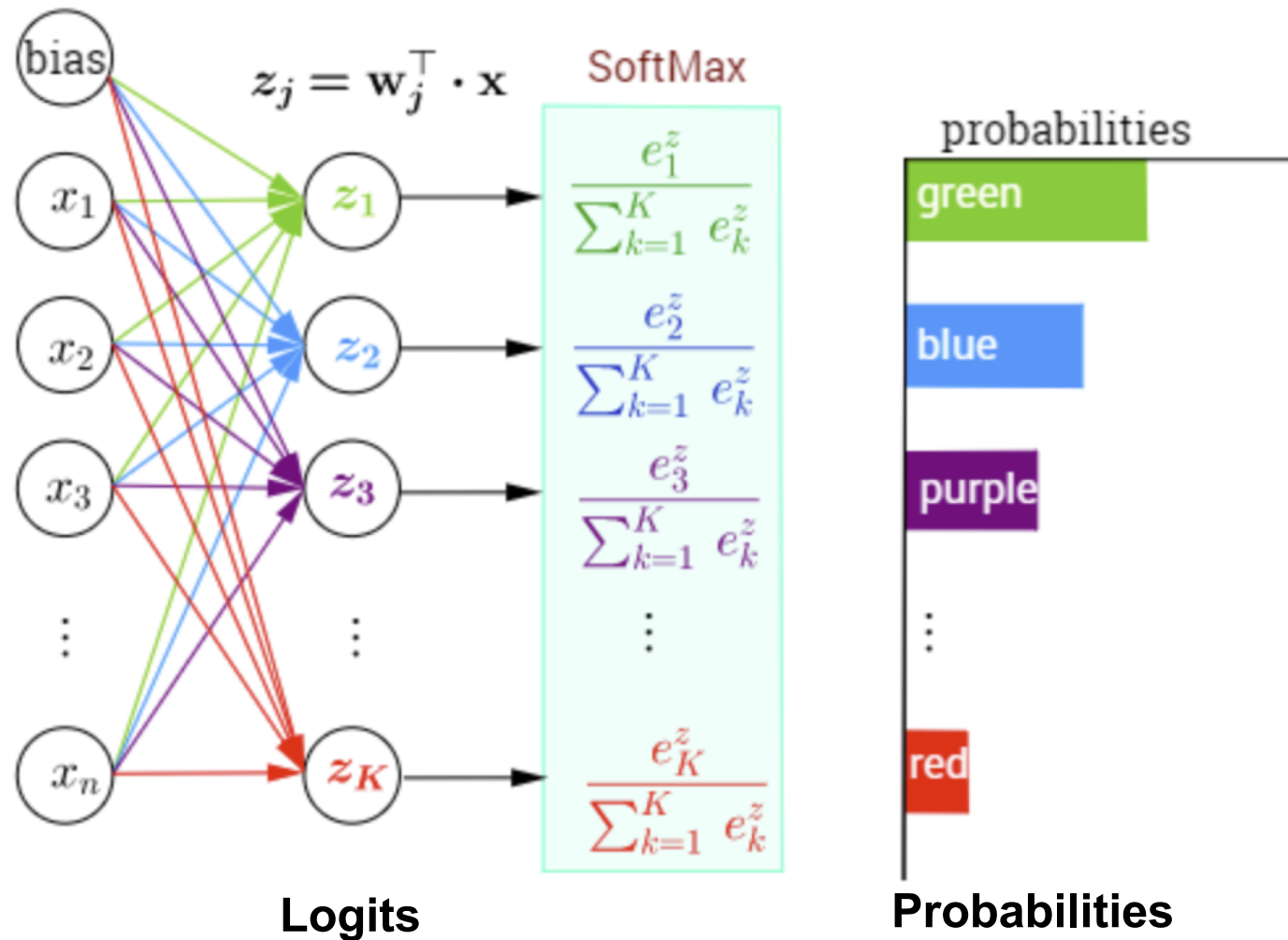
- Musical accuracy is not that good (yet)
- Regeneration of training example is better than Regeneration of test/validation example
- Case of Overfitness

Techniques

- Limit Accuracy and Control Overfitting
- More Examples (Augment the Corpus)
 - Keeping a Good Style Representation, Coverture and Consistency
 - More Consistency and Coverture
 - Transpose (**Align**) All Chorales to **Only One** Key (ex: **C**)
- More Synthetic Examples
 - More Coverture
 - Transpose All Chorales in **All** Keys (**12**)
- Regularization
 - Weight-based
 - » L1, L2
 - Connexion-based
 - » Dropout
 - Epochs-based
 - » Early-Stop
 - Analysis of Learning Curves

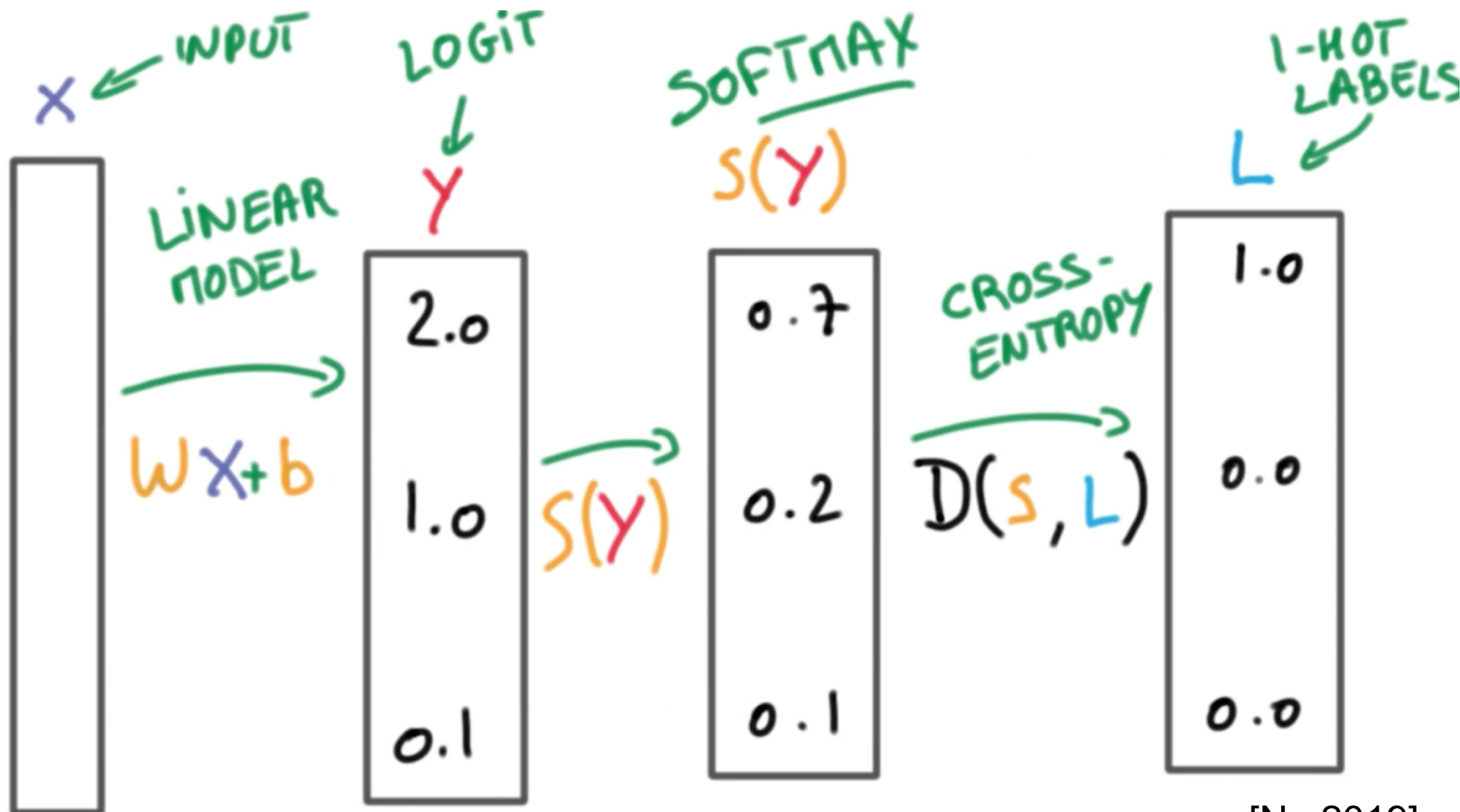


Softmax



Softmax + Cross-Entropy

- Cross-Entropy measures dissimilarity between two probability distributions (prediction and target/true value)



Output Activation and Cost/Loss Functions

<i>Task</i>	<i>Type of the output (\hat{y})</i>	<i>Encoding of the target (y)</i>	<i>Output activation function</i>	<i>Cost (loss)</i>	<i>Application</i>
Regression	Real	IR	Identity (Linear)	Mean squared error	
Classification	Binary	{0, 1}	Sigmoid	Binary cross-entropy	
Classification	Multiclass single label	One-hot	Softmax	Categorical cross-entropy	Monophony
Classification	Multiclass multilabel	Many-hot	Sigmoid	Binary cross-entropy	Polyphony
Multiple Classification	Multi Multiclass single label	Multi One-hot	Sigmoid	Binary cross-entropy	Multivoice
			Multi Softmax	Multi Categorical cross-entropy	

Ex. multiclass single label: Classification among a set of possible notes for a monophonic melody, with only one single possible note choice (single label)

Ex. multiclass multilabel: Classification among a set of possible notes for a single-voice polyphonic melody, therefore with several possible note choices (several labels)

Ex. multi multiclass single label: Multiple classification among a set of possible notes for multivoice monophonic melodies, therefore with only one single possible note choice for each voice; Multiple classification among a set of possible notes for a set of time slices for a monophonic melody, therefore for each time slice with only one single possible note choice

Output Activation and Cost/Loss Functions

<i>Task</i>	<i>Type of the output (\hat{y})</i>	<i>Encoding of the target (y)</i>	<i>Output activation function</i>	<i>Cost (loss)</i>	<i>Interpretation</i>
Regression	Real	\mathbb{R}	Identity (Linear)	Mean squared error	none
Classification	Binary	$\{0, 1\}$	Sigmoid	Binary cross-entropy	none
Classification	Multiclass single label	One-hot	Softmax	Categorical cross-entropy	argmax or sampling
Classification	Multiclass multilabel	Many-hot	Sigmoid	Binary cross-entropy	argsort and > threshold & max-notes
Multiple Classification	Multi Multiclass single label	Multi One-hot	Sigmoid	Binary cross-entropy	p argmax or sampling
			Multi Softmax	Multi Categorical cross-entropy	

Other cost functions:

Mean absolute error, Kullback-Leibler (KL) divergence...

Output Activation and Cost/Loss Functions

<i>Output type</i>	<i>Output activation function</i>	<i>Output value (\hat{y})</i>	<i>Cost</i>	<i>Target (true) value (y)</i>	<i>Interpretation</i>	<i>Meaning</i>
Real	Identity	<div>439.7</div>	<div><div></div><div></div><div>Mean squared error</div><div></div></div>	<div>440</div>		A_4
Binary	Sigmoid	<div>0.96</div>		<div>1</div>	> 0.5	True

Output Activation and Cost/Loss Functions

Output type	Output activation function	Output value (\hat{y})	Cost	Target (true) value (y)	Interpretation	Meaning
Multiclass single label	Softmax	0.03	Categorical cross-entropy	0	Argmax	A_4
		0.02		0		
		0.52		1		
		0.03		0		
		0.04		0		
		0.13		0		
		0.05		0		
		0.02		0		
		0.03		0		
		0.07		0		
		0.04		0		
		0.02		0		

Output Activation and Cost/Loss Functions

Output type	Output activation function	Output value (\hat{Y})	Cost	Target (true) value (Y)	Interpretation	Meaning
-------------	----------------------------	----------------------------	------	-----------------------------	----------------	---------

Multiclass multilabel

Sigmoid

0.07	↔	0
0.03	↔	0
0.54	↔	1
0.03	↔	0
0.07	↔	0
0.02	↔	0
0.45	↔	1
0.02	↔	0
0.09	↔	0
0.63	↔	1
0.12	↔	0
0.04	↔	0

Binary cross-entropy

Argsort
& > *threshold*
& < *max number notes*

A_4

F_4

D_4

Output Activation and Cost/Loss Functions

Output type	Output activation function	Output value (\hat{Y})	Cost	Target (true) value (Y)	Interpretation	Meaning
Multi multiclass single label	Sigmoid	0.03	Binary cross-entropy	0	Argmax	A_4
		0.02		0		
		0.52		1		
		0.03		0		
		0.04		0		
		0.13		0		
		0.05		0		
		0.02		0		
		0.03		0		
		0.07		0		
		0.04		0		
		0.02		0		
		0.05	Binary cross-entropy	0	Argmax	F_4
		0.02		0		
		0.04		0		
		0.03		0		
		0.02		0		
		0.05		0		
		0.53		1		
		0.06		0		
		0.03		0		
		0.09		0		
		0.03		0		
		0.05		0		

(Summary of) Principles of Loss Functions

- Probability theory + Information theory

See also Maximum likelihood principle

- Intuition:

- Information content (Likely Event) : Low
- Information content (Unlikely Event) : High

- Self-information: $I(x) = \log(1/P(x)) = -\log P(x)$

- Ex: $I(\text{note}=\text{B}) = -\log P(\text{note}=\text{B})$

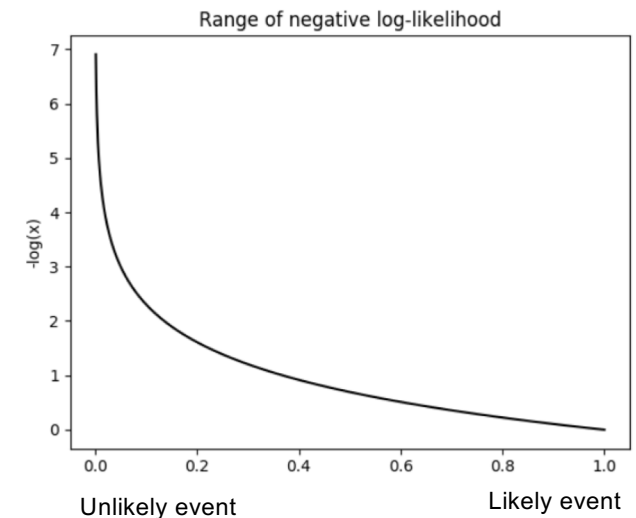
- Entropy of Probability distribution : $\sum_i I(\text{note}=\text{Note}_i)$, weighted by $P(\text{note}=\text{Note}_i)$

- $H(\text{note}) = \sum_i P(\text{note}=\text{Note}_i) I(\text{note}=\text{Note}_i)$

- Expectation-based alternative definition:

- Expectation: Mean value of $f(x)$ when $x \sim P$: $E_{x \sim P} [f(x)] = \sum_x P(x) f(x)$

- $H(\text{note}) = E_{\text{note} \sim P} I(x) = E_{\text{note} \sim P} [-\log P(\text{note})] = -E_{\text{note} \sim P} [\log P(\text{note})]$



KL-Divergence and Cross-Entropy

- Measures of Differences between Distributions (over a same variable: note)
 - **Assymmetric** $D_{KB}(P||Q) \neq D_{KB}(Q||P)$ $H(P,Q) \neq H(Q,P)$
- Kullback-Leibler Divergence (KL- Divergence):
- $D_{KB}(P||Q) = E_{\text{note} \sim P} [\log P(\text{note})/Q(\text{note})] = E_{\text{note} \sim P} [\log P(\text{note}) - \log Q(\text{note})]$
- Categorical Cross-Entropy:
- $H(P,Q) = E_{\text{note} \sim P} [-\log Q(\text{note})] = - E_{\text{note} \sim P} [\log Q(\text{note})]$
- Difference with KL-Divergence: $\log P(\text{note})$ term, constant with respect to Q
- $D_{KB}(y||\hat{y}) = E_{\text{note} \sim P} [\log y - \log \hat{y}] = \sum_i y_i (\log y_i - \log \hat{y}_i)$
- $H(y, \hat{y}) = - E_{\text{note} \sim P} [\log \hat{y}] = - \sum_i y_i \log \hat{y}_i$
- Binary Cross-Entropy:
- $H_B(y, \hat{y}) = - (y_0 \log \hat{y}_0 + y_1 \log \hat{y}_1) = - (y \log \hat{y} + (1-y) \log (1-\hat{y}))$

ForwardBach Brazilian Hymn Counterpoint



The image displays a musical score for a piece titled "ForwardBach Brazilian Hymn Counterpoint". The score is written for four staves, arranged in two systems of two staves each. The top system consists of a treble staff and a bass staff, both in 2/4 time. The bottom system also consists of a treble staff and a bass staff, both in 2/4 time. The key signature is one flat (B-flat). The music features a complex counterpoint between the two systems, with various rhythmic patterns and melodic lines. The notation includes eighth notes, sixteenth notes, and quarter notes, with some rests and accidentals. A speaker icon is visible on the right side of the image, indicating that the music is playable.

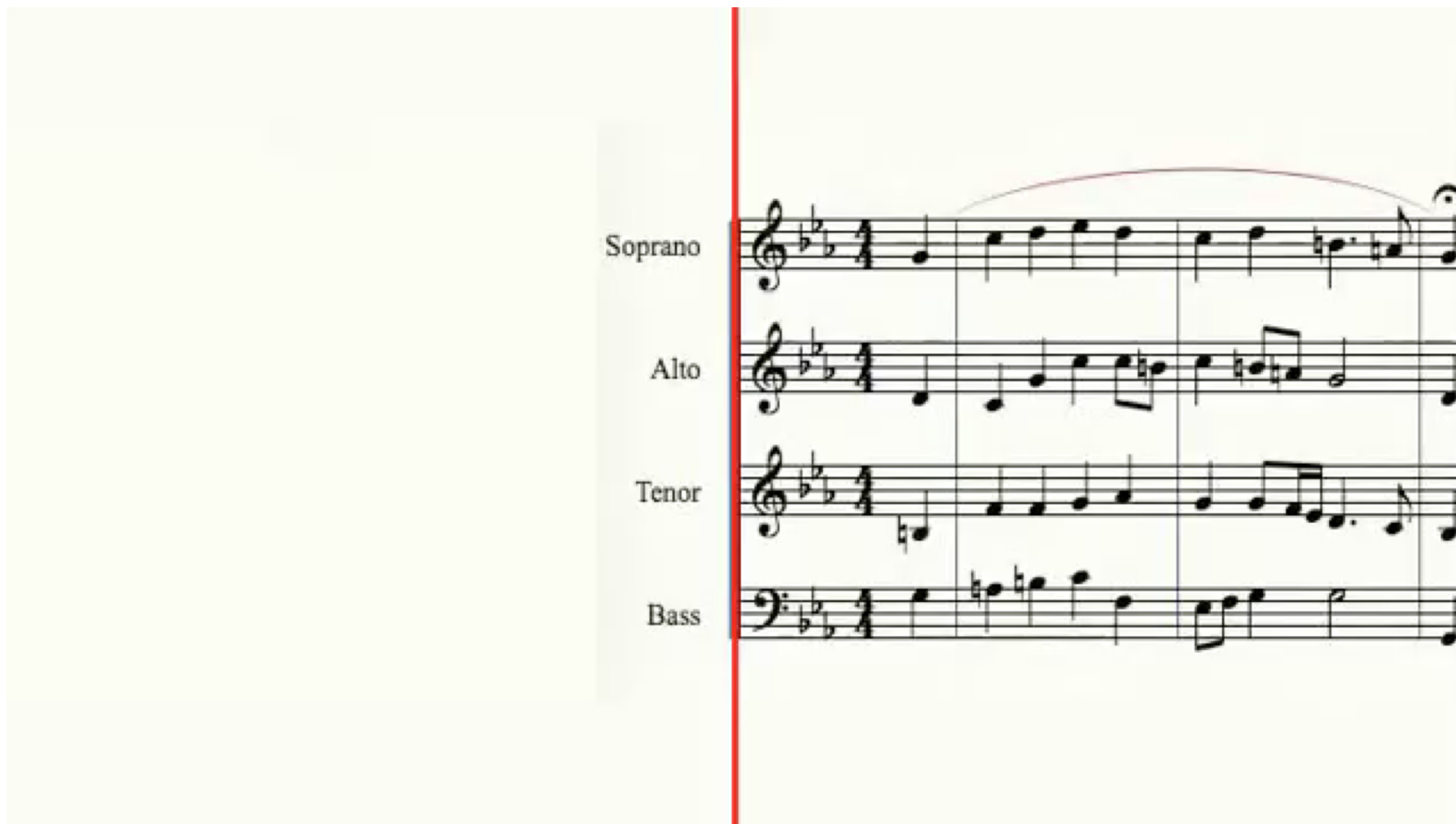
ForwardBach Brazilian Hymn Counterpoint (2 times slower and removing the intro)



The image displays a musical score for a piece titled "ForwardBach Brazilian Hymn Counterpoint". The score is presented in two systems, each containing four staves. The first system consists of two treble staves and two bass staves, all in 2/4 time. The melody is written in the first treble staff, featuring a key signature of one flat (B-flat) and a tempo marking of "2 times slower". The counterpoint is provided in the second treble staff and both bass staves. The second system follows a similar layout, with the melody in the first treble staff and counterpoint in the other three staves. The score includes various musical notations such as notes, rests, and accidentals, indicating a complex harmonic structure.



DeepBach – Demo [Hadjeres, 2017]



A musical score for four voices: Soprano, Alto, Tenor, and Bass. The score is written in 4/4 time with a key signature of two flats (B-flat and E-flat). A vertical red line is positioned between the Soprano and Alto staves. A red slur is placed over the Soprano staff, spanning from the first measure to the end of the fourth measure. The music is presented on four staves, each with its respective voice label to the left.

<https://www.youtube.com/watch?v=QiBM7-5hA6o>

Reorchestration of God Save the Queen by DeepBach [Hadjeres, 2018]

https://www.youtube.com/watch?time_continue=1&v=x-W0ixD9Cpg