

# Identificador de Sistema Dinâmico e Previsor Futebolístico

Práticas em reconhecimento de padrões, modelagem e neurocomputação



Gabriel de Bodt Sivieri  
Leonardo Vergueiro  
Maria Eduarda Barroso Waisman

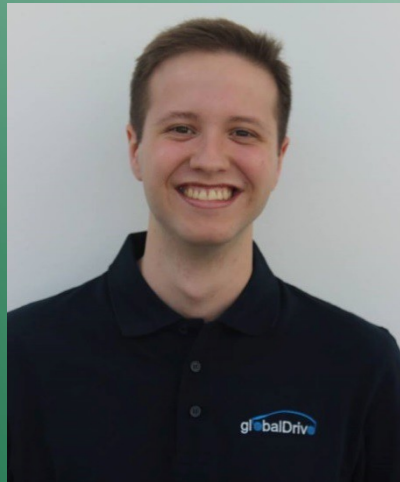
Professor: Emilio Del Moral Hernandez

Escola Politécnica, Universidade de São Paulo,  
Av. Prof. Luciano Gualberto 158, 05508900 São Paulo, Brasil

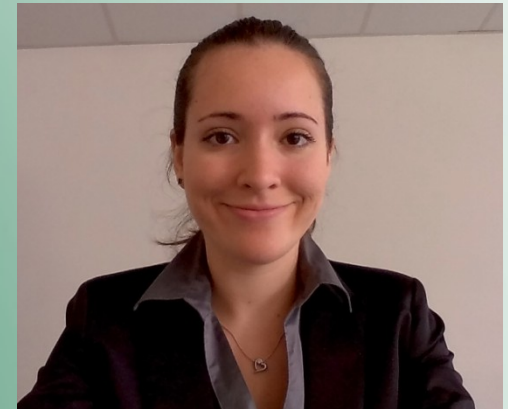
# Integrantes



Gabriel Sivieri



Leonardo Vergueiro



Maria Waisman

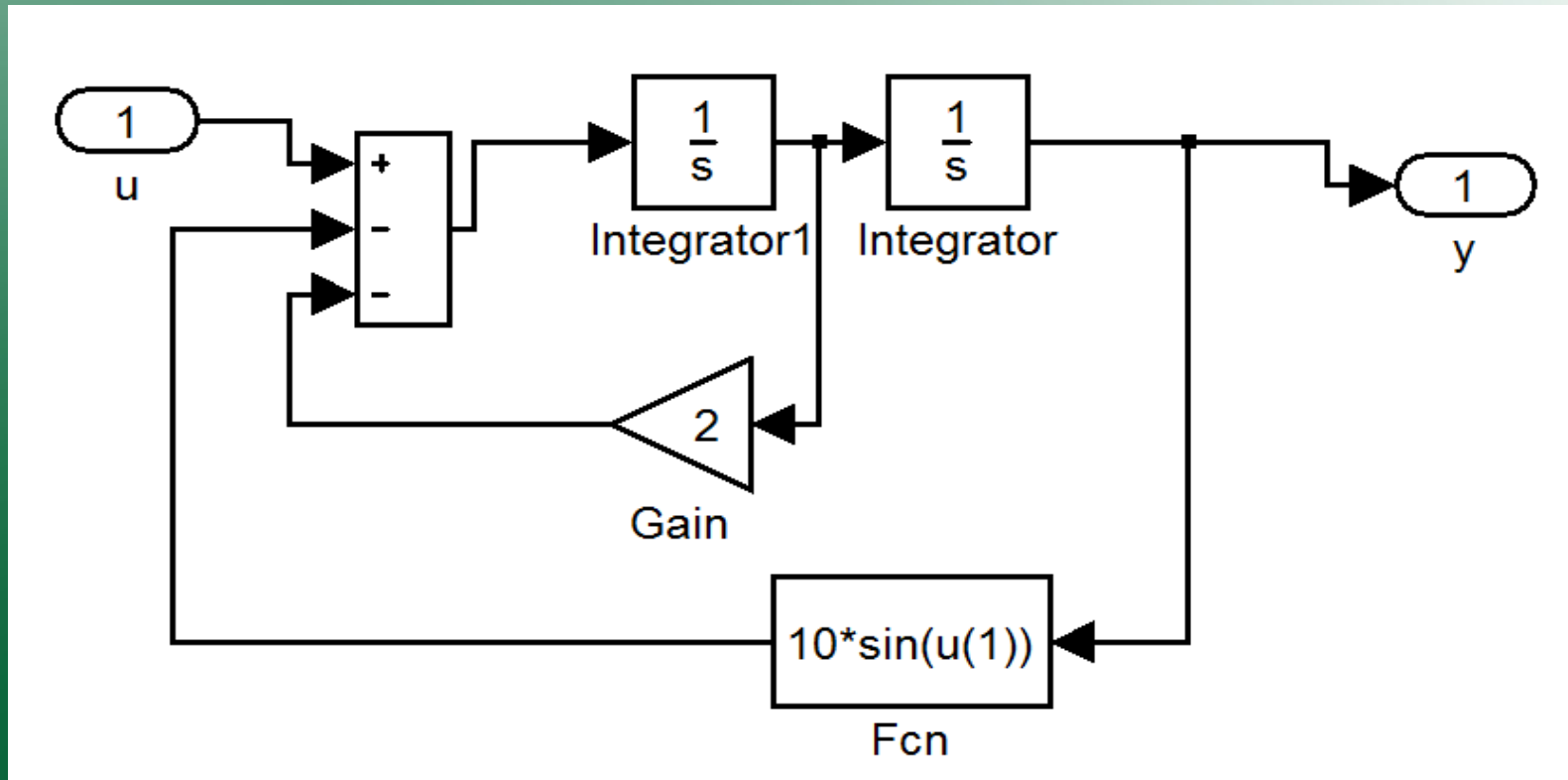
# Sumário

- Identificador de sistema dinâmico
- Aquisição de dados
- Treinamento da rede
- Medida de qualidade
- Previsor futebolístico

# Identificador de Sistema Dinâmico

- Modelagem de um sistema dinâmico qualquer
- Caixa Preta
- Rede prevê o seu funcionamento
- Caso de estudo: Braço mecânico

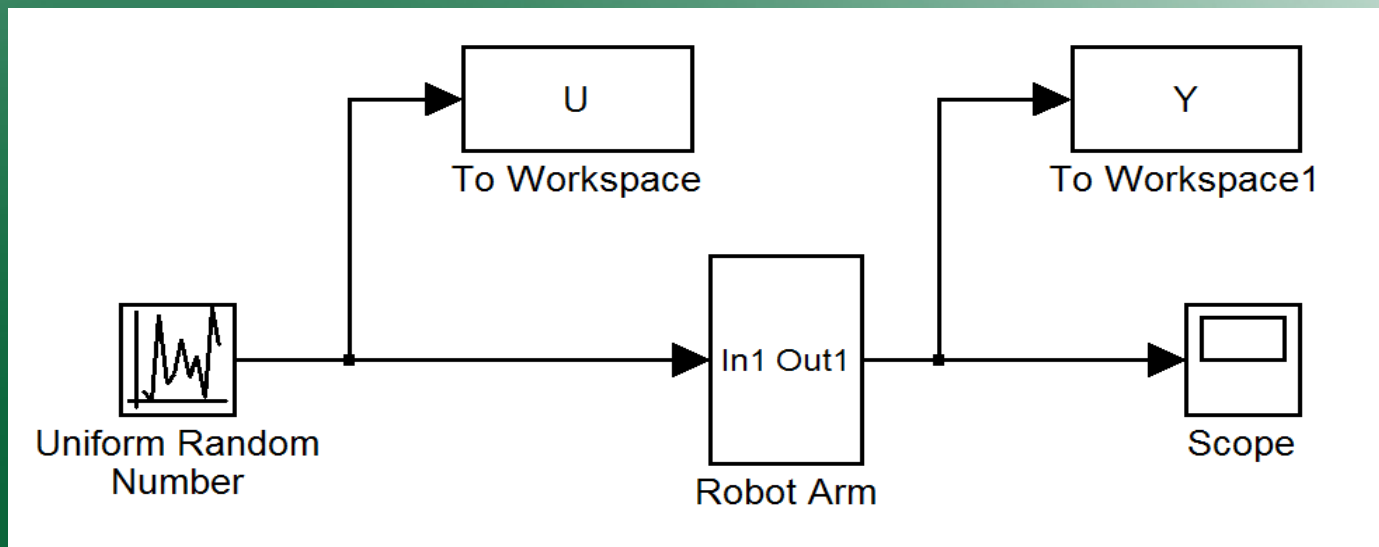
# Aquisição de dados



Modelo do braço robótico simples, fonte: Mathworks

# Aquisição de dados

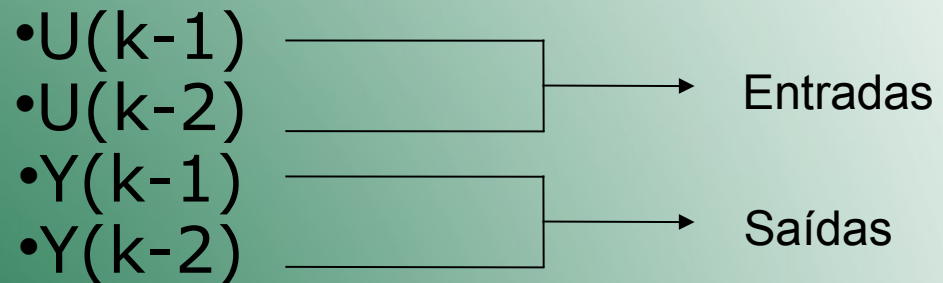
- Sistema não-linear devido à função seno
- Dados de treino e teste obtidos com o uso de gerador aleatório na entrada



Mathworks.

# Treinamento

- Features de entrada escolhidas



- Rede PCM (Perceptron multi-camadas) de entradas atrasadas no tempo
- 500 amostras de entrada e saída com  $k$  variando de 3 a 500

# Treinamento da rede (4-10-1)

- 400 dados para treino e 100 para teste
- Topologia da RNA:
  - 4 dados de entrada
  - 10 neurônios de camada escondida
  - 1 neurônio de camada de saída



# Medida de Qualidade (4-10-1)

- $W$ 's iniciais entre -1 e 1
  - Número de iterações: 200000
  - Erro quadrático médio do treino: 1.25%
  - Erro quadrático médio do teste: 1.40%
- 
- $W$ 's iniciais entre -0.5 e 0.5
  - Número de iterações: 200000
  - Erro quadrático médio do treino: 1.37%
  - Erro quadrático médio do teste: 1.69%

# Medida de Qualidade (4-10-1)

- $W$ 's iniciais entre -1 e 1
- Número de iterações: 700000
- Erro quadrático médio do treino: 1.14%
- Erro quadrático médio do teste: 1.35%

# Treinamento da rede (4-50-1)

- Topologia da RNA:
  - 4 dados de entrada
  - 50 neurônios de camada escondida
  - 1 neurônio de camada de saída

# Medida de Qualidade (4-10-1)

- $W$ 's iniciais entre -1 e 1
- Número de iterações: 200000
- Erro quadrático médio do treino: 1.25%
- Erro quadrático médio do teste: 1.37%
- Maior tempo em cada iteração em relação à topologia anterior

# Treinamento da rede (4-7-7-1)

- Topologia da RNA:
  - 4 dados de entrada
  - 7 neurônios na primeira camada escondida
  - 7 neurônios na segunda camada escondida
    - 1 neurônio de camada de saída

# Medida de Qualidade (4-7-7-1)

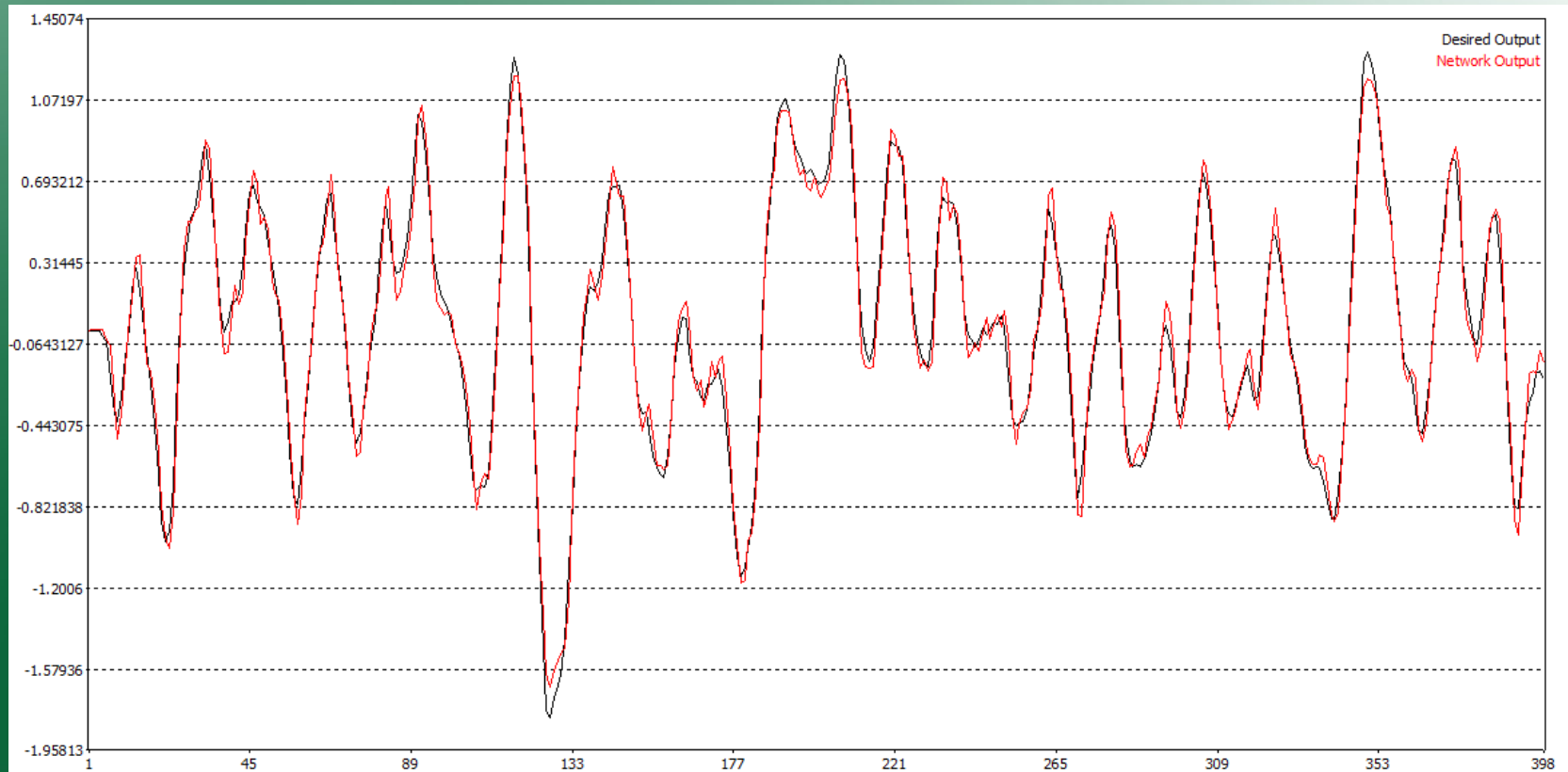
- W's iniciais entre -1 e 1
- Número de iterações: 200000
- Erro quadrático médio do treino: 1.20%
- Erro quadrático médio do teste: 1.30%

# Treinamento da rede (4-7-7-1)

- Topologia da RNA:
  - 4 dados de entrada
  - 7 neurônios na primeira camada escondida
  - 7 neurônios na segunda camada escondida
    - 1 neurônio de camada de saída

# Gráficos

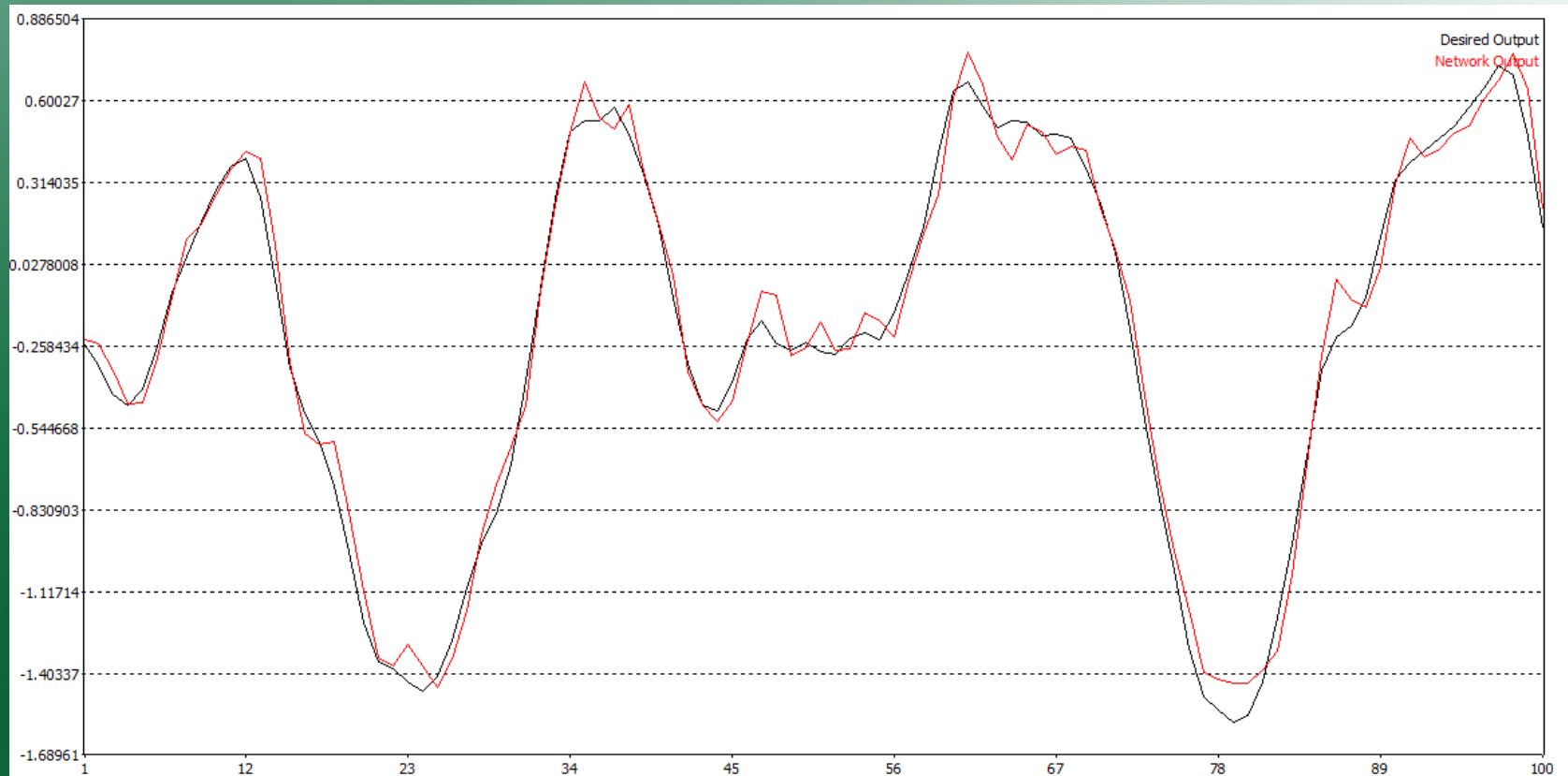
## Saída desejada e saída da RNA na fase de treino





# Gráficos

## Saída desejada e saída da RNA na fase de teste



# Previsor Futebolístico

## CONTEXTO:

- Crescimento da base de dados e análise estatística em esportes.  
(Prozone, Optasports, Amisco, Infostrada...)
- Final da Champions de 2010 – 2842 eventos coletados (Optasports 2010).

## DESAFIO:

Traduzir os dados em vantagem no campo.

- Softwares de análise: Wyscout, Mycouch.

# Justificativa

## TENTATIVAS:

- RNA, Fuzzy Logic, Machine Learning -> IA.
- Robocup – Robot Soccer Simulation. (Aplicações de IA em Software).

META: “Em meados do Século 21, um time de robôs autônomos ganhe uma partida de futebol, nas regras da FIFA, contra o último campeão do mundo” (Robocup 1998).

NOSSA IDÉIA: Aplicação global de uma RNA.

# Aquisição de Dados

- Determinação da Vitória, Derrota ou Empate de um time.
- Aquisição de dados com o site Footstats.
- 187 jogos do primeiro turno do campeonato brasileiro
  - 86 vitórias do time da casa
    - 57 empates
  - 44 derrotas do time da casa
- 150 amostras de treino
- 37 amostras de teste

# Tentativas de Treinamento

- Tentativa 1:
  - 36 Features (18 de cada time).
  - 3 Saídas(Vitória, Empate, Derrota).
  - 150 amostras de treino.
  - 37 amostras de teste.
- Tentativa 2:
  - 20 Features (10 de cada time).
  - 3 Saídas(Vitória, Empate, Derrota).
  - 150 amostras de treino.
  - 37 amostras de teste.

# Medida de qualidade

- Taxa de acerto da RNA
  - Porcentagem do número de acertos da RNA no conjunto de teste / total de amostras de teste
- Decisor após o uso da RNA:
  - Valores acima de 0.8 ativam a saída respectiva
    - Torna-se o valor binário
    - Compara-se com o esperado
    - Detecta-se acerto ou erro

# Treinamento 1

- 36 features (18 de cada time):
  - Finalizações certas e erradas, passes certos e errados, cruzamentos certos e errados, lançamentos certos e errados, dribles certos e errados, faltas cometidas e recebidas, penaltis recebidos, impedimentos, escanteios certos e errados
- Pré-tratamento com normalização dos dados com o desvio padrão da média de cada feature

# Primeira iteração

- W's iniciais entre -1 e 1
- Número de iterações: 300
- Taxa de aprendizado: 0.7
- Erro quadrático médio do teste: 1.69%
- Topologia: 36-7-3

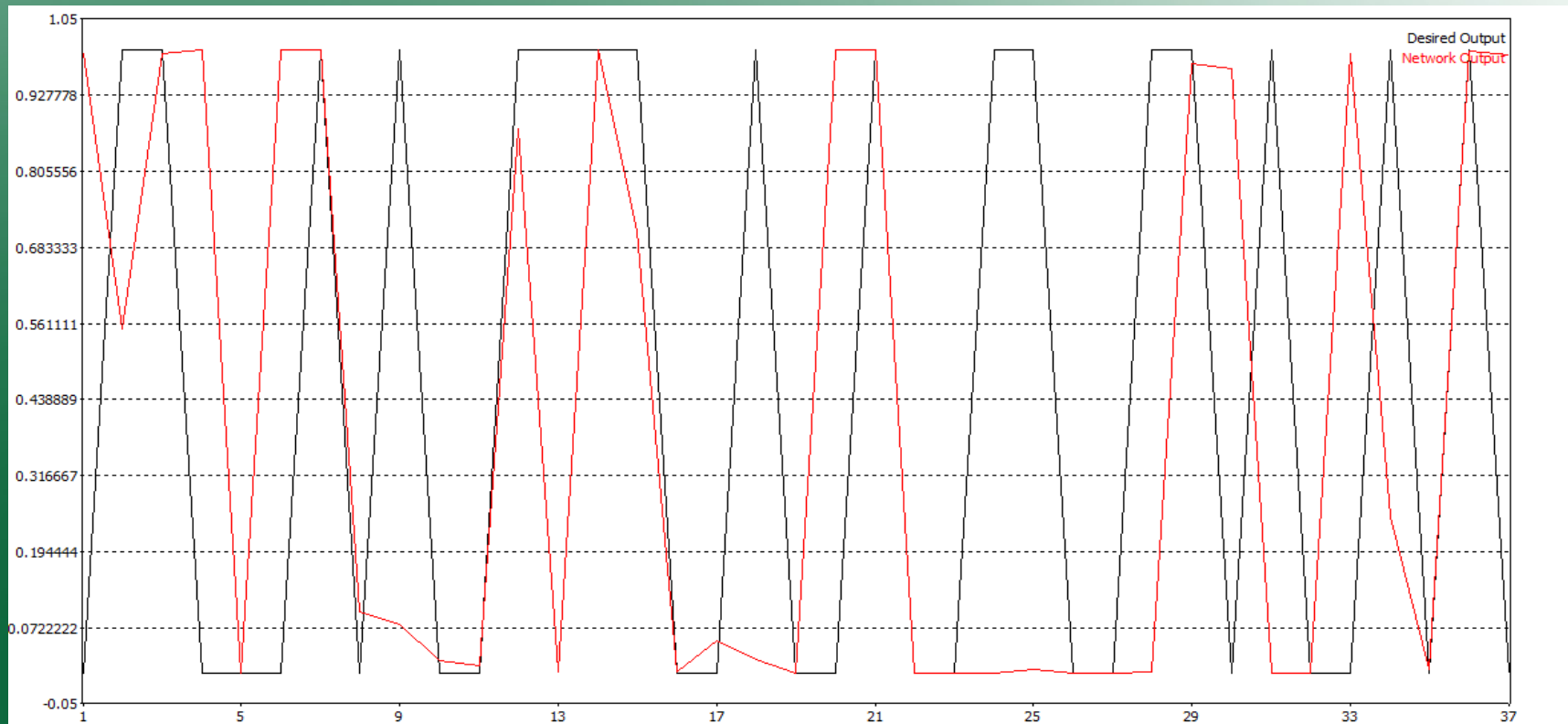


Gráfico RMS



# Primeira iteração

- Taxa de acerto: 27%



# Terceira iteração

- $W$ 's iniciais entre -0.3 e 0.3
- Taxa de acerto: 38%



Gráfico RMS

# Resumo Tentativa 1

Pós maiores iterações descobriu-se que:

- Baixa taxa de acerto
- Aumento do número de nós não melhorava a taxa de acerto.
- Muitas features complementares
- Número de features aumentava tempo de treinamento.

# Tentativa 2

Tentativa baseou-se nas seguintes premissas:

- Diminuição e normalização das Features.
- Avaliação do efeito do número de nós na camada oculta.
- Novas abordagens de taxa de acerto com a Matriz de confusão.

Abordagens:

- Acerto Global: Contém previsões não feitas.
- Acerto Restrito: Sem previsões não feitas.

# Treinamento 2

- 20 features (10 de cada time):
  - % de acerto Finalizações, % de acerto Passes, % de acerto Cruzamentos, % de acerto Lançamentos, % de acerto Dribles, Faltas recebidas, Penaltis recebidos, impedimentos, % de acerto Escanteios.
- Pré-tratamento com normalização dos dados não percentuais com o desvio padrão da média de cada Feature.

# Ressorteios dos Ws

- Topologia: 20-40-3
- W's de -1 à 1:  
Taxa Global:40.54%  
Taxa Restrita:50%
- W's de -0.5 à 0.5:  
Taxa Global:48.65%  
Taxa Restrita:52.94%
- W's de -0.3 à 0.3:  
Taxa Global:37.84%  
Taxa Restrita:45.16%
- W's de -0.8 à 0.8:  
Taxa Global:40.54%  
Taxa Restrita:57.69%
- W's de -0.9 à 0.9:  
Taxa Global:45.95%  
Taxa Restrita:60.71%

# Análise dos Ws

Matriz de Confusão Global					Matriz de Confusão Restrita				
Taxa Global	45.95%	Classificação			Taxa Restrita	60.71%	Classificação		
Verdade	Vitória	76.47%	5.88%	5.88%	Verdade	Vitória	41.18%	11.76%	17.65%
	Empate	45.45%	18.18%	18.18%		Empate	0.00%	18.18%	9.09%
	Derrota	22.22%	0.00%	22.22%		Derrota	0.00%	0.00%	40.00%
Matriz de Confusão Global					Matriz de Confusão Restrita				
Taxa Global	45.95%	Classificação			Taxa Restrita	60.71%	Classificação		
Verdade	Vitória	13	1	1	Verdade	Vitória	13	1	1
		17	17	17			17	17	17
	Empate	5	2	2		Empate	5	2	2
		11	11	11			7	7	7
	Derrota	2	0	2		Derrota	2	0	2
		9	9	9			5	5	5

Taxas e Matrizes com Ws entre -0.9 e 0.9

# Ensaaios de Epochs

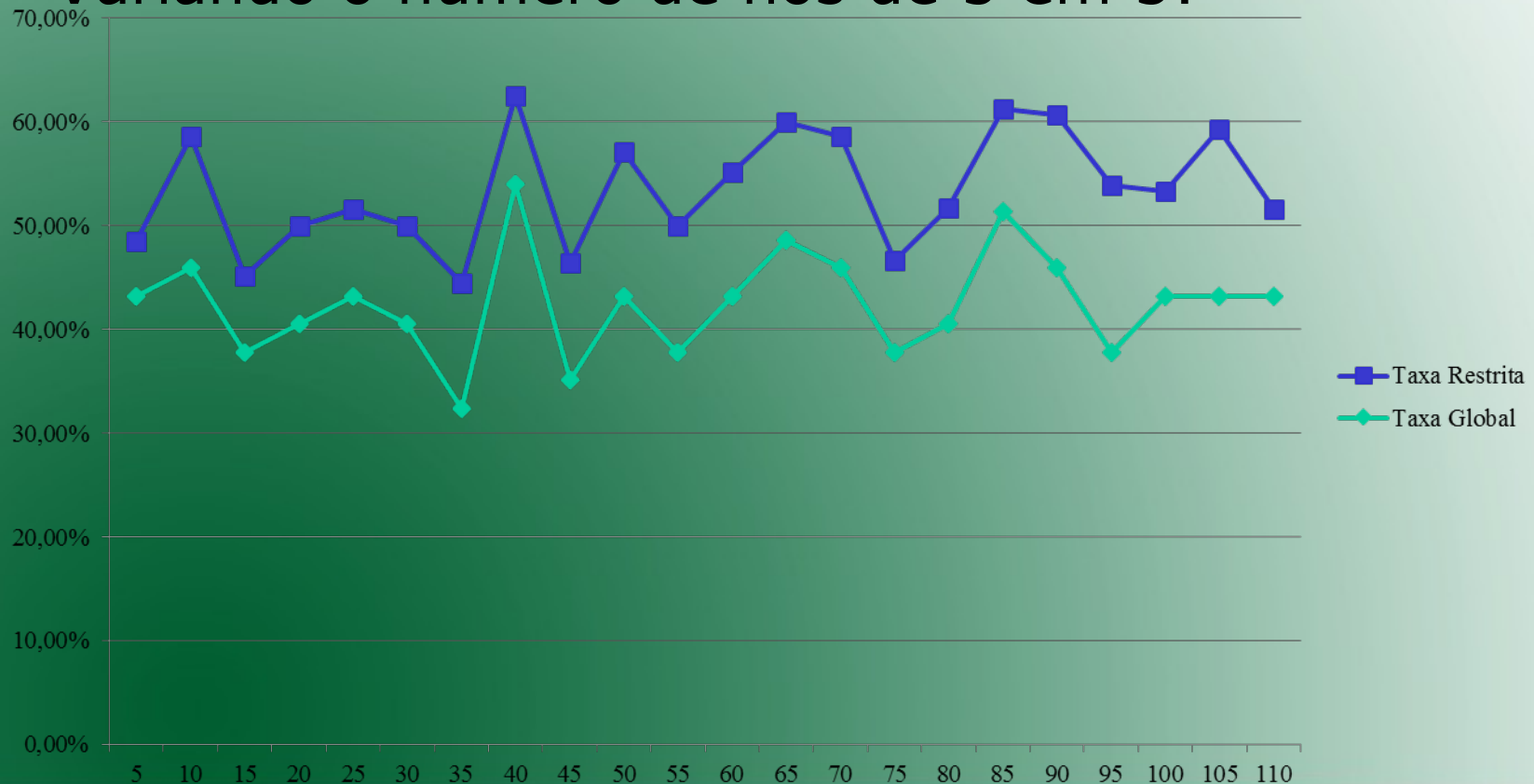
- Topologia: 20-40-3
- Epochs 500:  
Taxa Global:43.24%  
Taxa Restrita:59.26%
- Epochs 300:  
Taxa Global:37.84%  
Taxa Restrita:56.00%
- Epochs 1000:  
Taxa Global:43.24%  
Taxa Restrita:57.14%
- Epochs 2500\*:  
Taxa Global:45.95%  
Taxa Restrita:58.62%  
\*Ws anteriores.
- Epochs 3000:  
Taxa Global:43.24%  
Taxa Restrita:53.33%

Decisão: 500 Epochs



# Ensaio de Nós

- Variando o número de nós de 5 em 5:



# Ensaio de Nós

Topologia: 20-40-3 Taxa Global: 54,05% Taxa Restrita: 62,50%

Matriz de Confusão Global					Matriz de Confusão Restrita				
Taxa Global	54.05%	Classificação			Taxa Restrita	62.50%	Classificação		
		Vitória	Empate	Derrota			Vitória	Empate	Derrota
Verdade	Vitória	76.47%	0.00%	11.76%	Verdade	Vitória	81.25%	23.53%	17.65%
	Empate	36.36%	27.27%	18.18%		Empate	0.00%	27.27%	27.27%
	Derrota	33.33%	11.11%	44.44%		Derrota	0.00%	12.50%	50.00%
Matriz de Confusão Global					Matriz de Confusão Restrita				
Taxa Global	54.05%	Classificação			Taxa Restrita	62.50%	Classificação		
		Vitória	Empate	Derrota			Vitória	Empate	Derrota
Verdade	Vitória	13	0	2	Verdade	Vitória	13	0	2
		17	17	17			16	16	16
	Empate	4	3	2		Empate	4	3	2
		11	11	11			8	8	8
	Derrota	3	1	4		Derrota	3	1	4
		9	9	9			8	8	8

# Considerações Finais

- Avaliação de Vitórias do time da casa.

Dados de ocorrência do Campeonato:

Vitória em casa:48%

Empates:28%

Derrotas:23%

Com a RNA obteve-se chance de acerto:

Matriz de Confusão Restrita				
Taxa Restrita	62.50%	Classificação		
		Vitória	Empate	Derrota
Verdade	Vitória	13	0	2
		32	32	32
	Empate	4	3	2
		32	32	32
	Derrota	3	1	4
		32	32	32

Matriz de Confusão Restrita				
Taxa Restrita	62.50%	Classificação		
		Vitória	Empate	Derrota
Verdade	Vitória	40.63%	0.00%	6.25%
	Empate	12.50%	9.38%	6.25%
	Derrota	9.38%	3.13%	12.50%

# Considerações Finais

Distância da realidade:

Vitória: 7.38% Empate: 18.63% Derrota: 10.50%

# Perguntas?



[decom.ufop.br](http://decom.ufop.br)

## Obrigada pela atenção!