

Cascade-correlation Architecture (Cascor)

Germano Crispim Vasconcelos
Centro de Informática - UFPE

Motivações



- Redes Feedforward muito usadas em Reconhecimento de Padrões
- Problemas com redes MLP e RBF
- Redes Construtivas vs Arquiteturas Fixas
- Redes construtivas ainda menos investigadas na teoria e na prática



Problemas das Redes de Topologia Fixa



- Dilema de definição a priori do número de camadas e unidades intermediárias para solução do problema
- Lentidão de processo de treinamento
- Problema da mudança constante de alvo



Problemas na Modelagem com MLP tradicional:



Lentidão na aprendizagem:

“Quando uma rede é treinada com back-propagation, todos os pesos das unidades escondidas precisam ser reajustadas para minimizar o erro residual”.

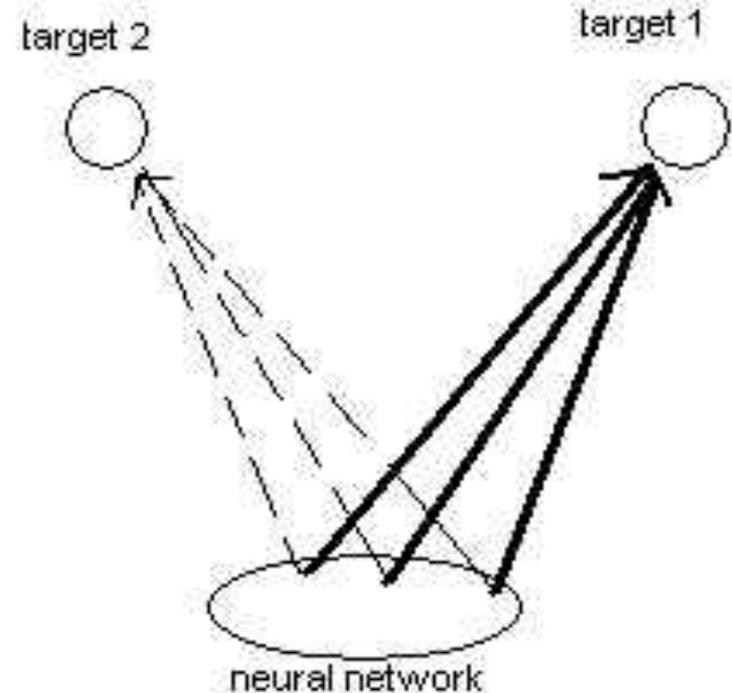


Problemas na Modelagem com MLP tradicional:

Problema da mudança de alvo:

Cada unidade dentro da rede está tentando se transformar em um detector de características mas os problemas de entrada estão mudando constantemente

Isto faz com que as unidades escondidas fiquem em um estado caótico, e leva algum tempo para elas estabilizarem.



Problemas na Modelagem com MLP tradicional:



Problema da mudança de alvo:

- Suponha 2 unidades escondidas para resolver duas tarefas. Cada unidade não se comunica com a outra, então decidem independentemente que tarefa resolver
- Se uma tarefa produz sinal de erro maior, todas as unidades tendem a resolver esta tarefa e ignorar a outra
- Uma vez resolvida, todas as unidades se movem para a segunda tarefa, mas o primeiro problema reaparece



Redes Neurais Construtivas



- Constroem gradualmente suas próprias estruturas até atingirem uma solução satisfatória
- Configuração inicial mínima
- Adotam estratégias de inserção e treinamento distintas
- Treinam, em geral, mais rápido



Cascade Correlation Architecture



- Proposta por Fahlman em 1988
- Rede construída a partir de uma arquitetura inicial mínima, apenas com camadas de entrada e saída
- Novas unidades são adicionadas uma a uma à medida que necessárias, criando estrutura em múltiplas camadas

Cascade Correlation (CC)



Fahlman & Lebiere

-> Combina duas idéias

- Arquitetura em cascata:

unidades escondidas são adicionadas uma a uma e congeladas

- Algoritmo de aprendizagem:

Inicia com estrutura mínima, treina e instala novas unidades intermediárias



Cascade Correlation CASCOR



- Cada nova unidade maximiza a correlação entre suas respostas e o erro residual
- Unidade de maior correlação com erro residual é selecionada entre “candidatas”
- Após candidata é selecionada, seus pesos de entrada são congelados, e apenas os pesos de saída são treinados

Características da Cascade



Correlation (Cascor):

- Algoritmo de treinamento supervisionado
- Rede determina seu tamanho e topologia
 - inicia-se com uma única camada de unidades
 - constrói uma rede de múltiplas camadas de estrutura mínima, criando camadas escondidas uma por uma

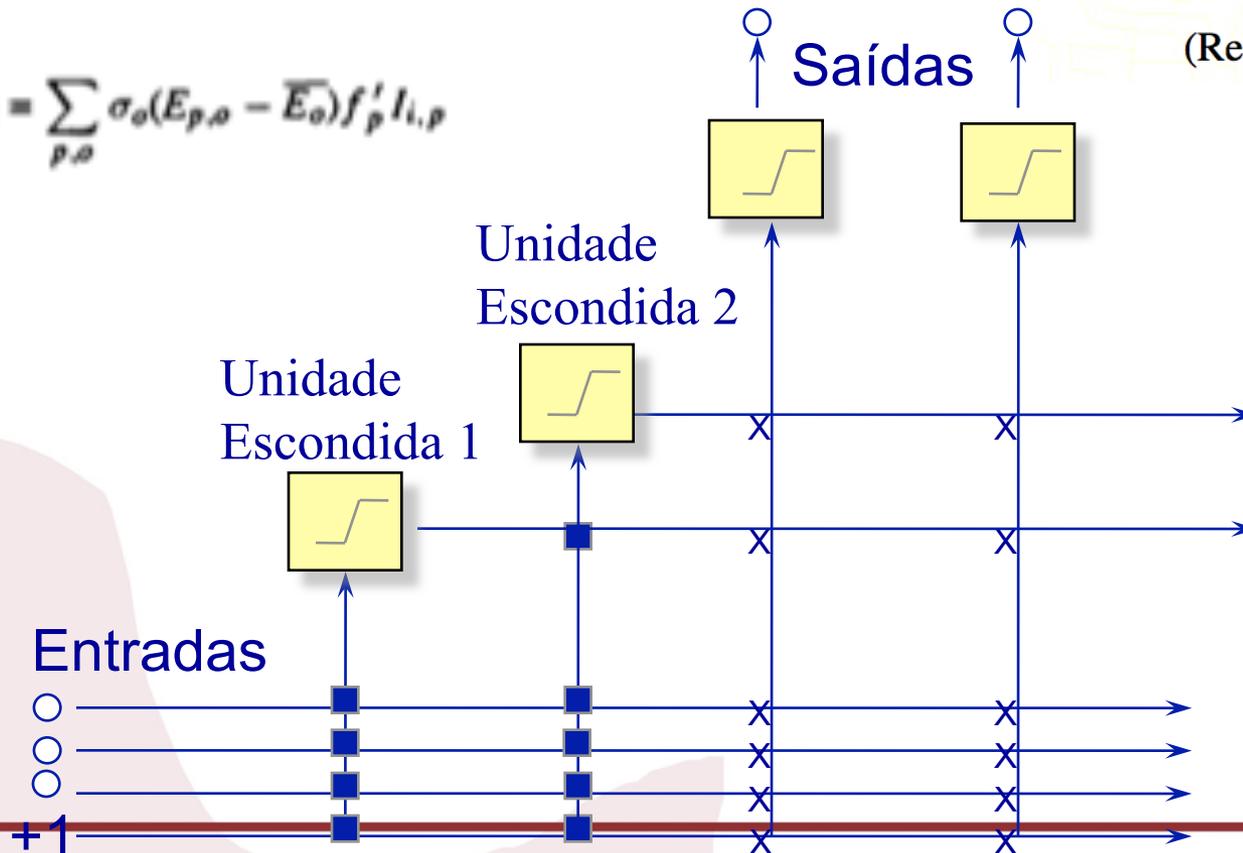
Cascade-Correlation (CASCOR)

$$S \text{ (covariância)} = \sum_o |\sum_p (y_{po} - \bar{y}_o)(e_{po} - \bar{e}_o)|$$

$$W(t+1) = W(t) + \eta e X(t)$$

(Regra de Widrow-Hoff)

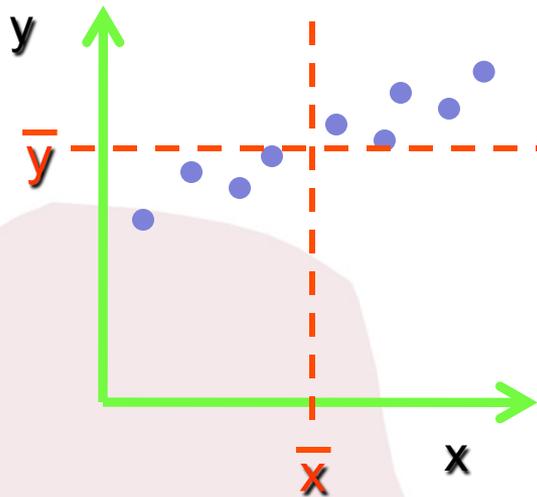
$$\frac{\partial S}{\partial w_{li}} = \sum_{p,o} \sigma_o (E_{p,o} - \bar{E}_o) f'_p l_{i,p}$$



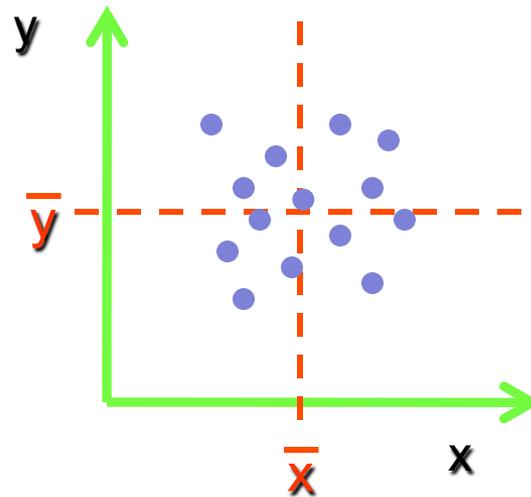
Covariância e Correlação

- Para um conjunto (x_i, y_i) , $i = 1, \dots, n$:

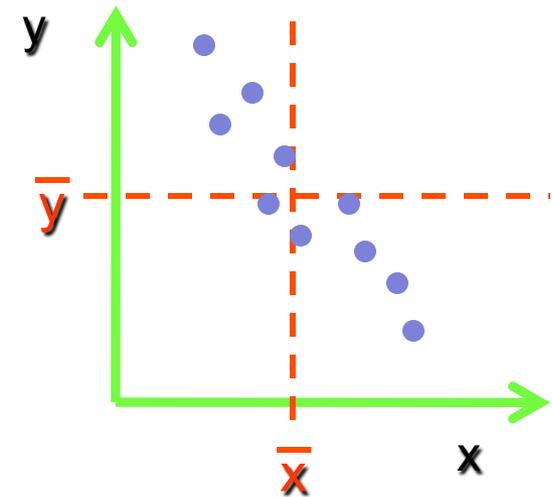
$$\text{cov}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$



$\text{cov}(x, y) > 0$



$\text{cov}(x, y) \approx 0$



$\text{cov}(x, y) < 0$

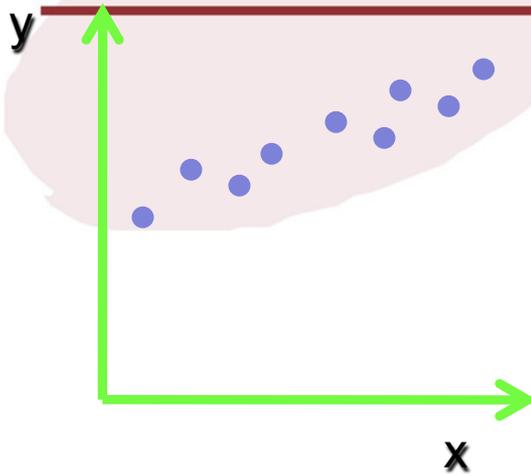
Covariância e Correlação



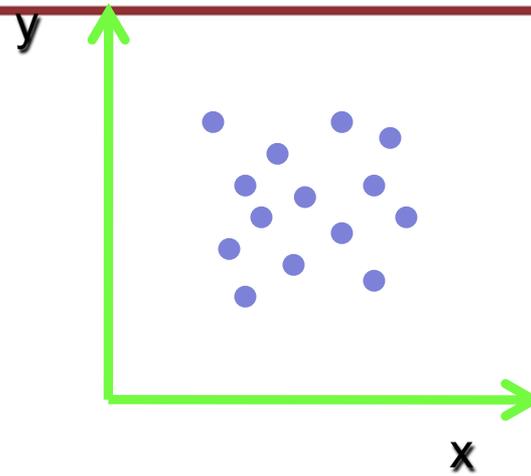
- Indica força e direção, diretamente ou inversamente proporcional, da relação linear entre x e y
- Em muitas aplicações, normaliza-se entre -1 e 1.
- Neste caso (r) para um conjunto (x_i, y_i) , $i = 1, \dots, n$:

$$r = \text{corr}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=0}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=0}^n (y_i - \bar{y})^2}}$$

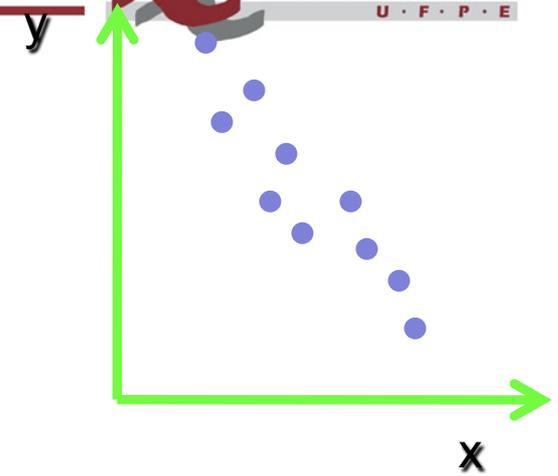
Covariância e Correlação



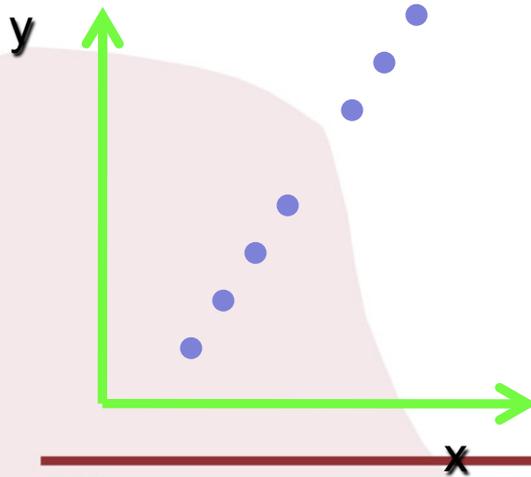
$$0 < r < 1$$



$$r \approx 0$$

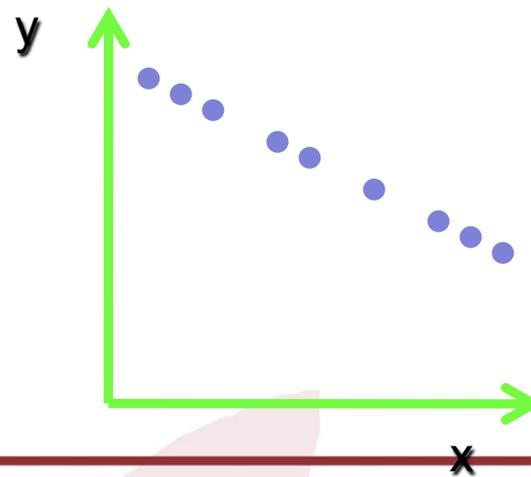


$$-1 < r < 0$$

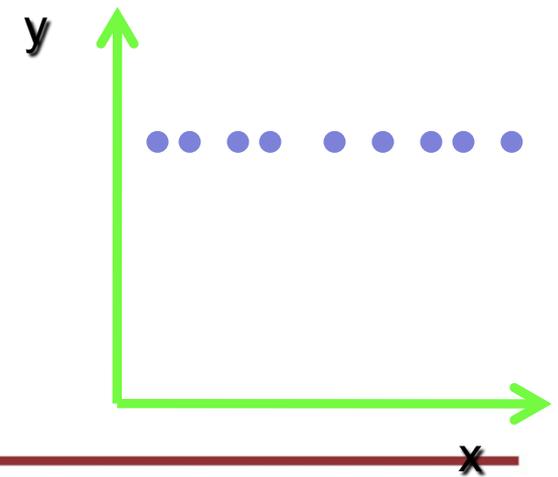


$$r = 1$$

UNIVERSIDADE FEDERAL
DE PERNAMBUCO



$$r = -1$$



$$r \text{ indefinido}$$

Características da Cascor



- Rede razoavelmente pequena é construída
 - Aprendizagem rápida (problema fixo vs dança complexa)
 - Detectores de características de alta ordem (redes profundas)
 - Potencialmente útil em aprendizagem incremental
-



Pseudo-Algoritmo (Cascor):

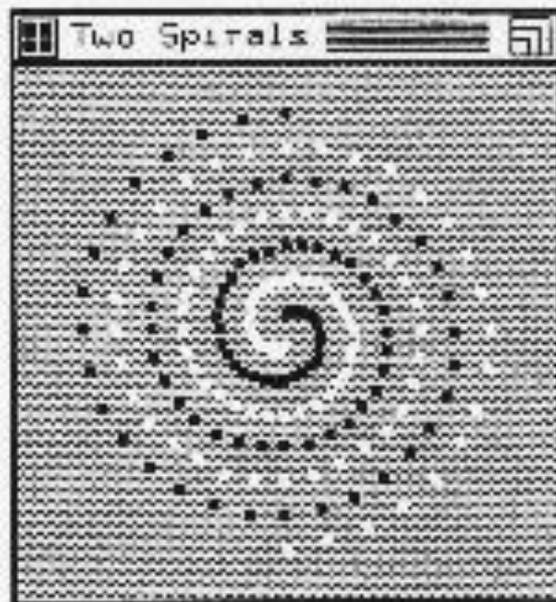


- Estrutura inicial mínima é treinada para minimizar o erro na camada de saída até o mínimo possível
- Novas unidades candidatas são adicionadas para resolver novas tarefas de acordo com o erro de aproximação residual (“Cascade”). Pesos de entrada para novas unidades candidatas são treinadas para maximizar a “Correlação” (“Correlation”) entre as saídas das unidades candidatas e o erro residual atual
- A unidade candidata com maior correlação é escolhida e se transforma em unidade escondida. Pesos de entrada da unidade são congelados e a rede é re-treinada para minimizar o erro na camada de saída
- Em seguida, os pesos de entrada para esta unidade são descongelados e treinados para minimizar o erro na camada de saída

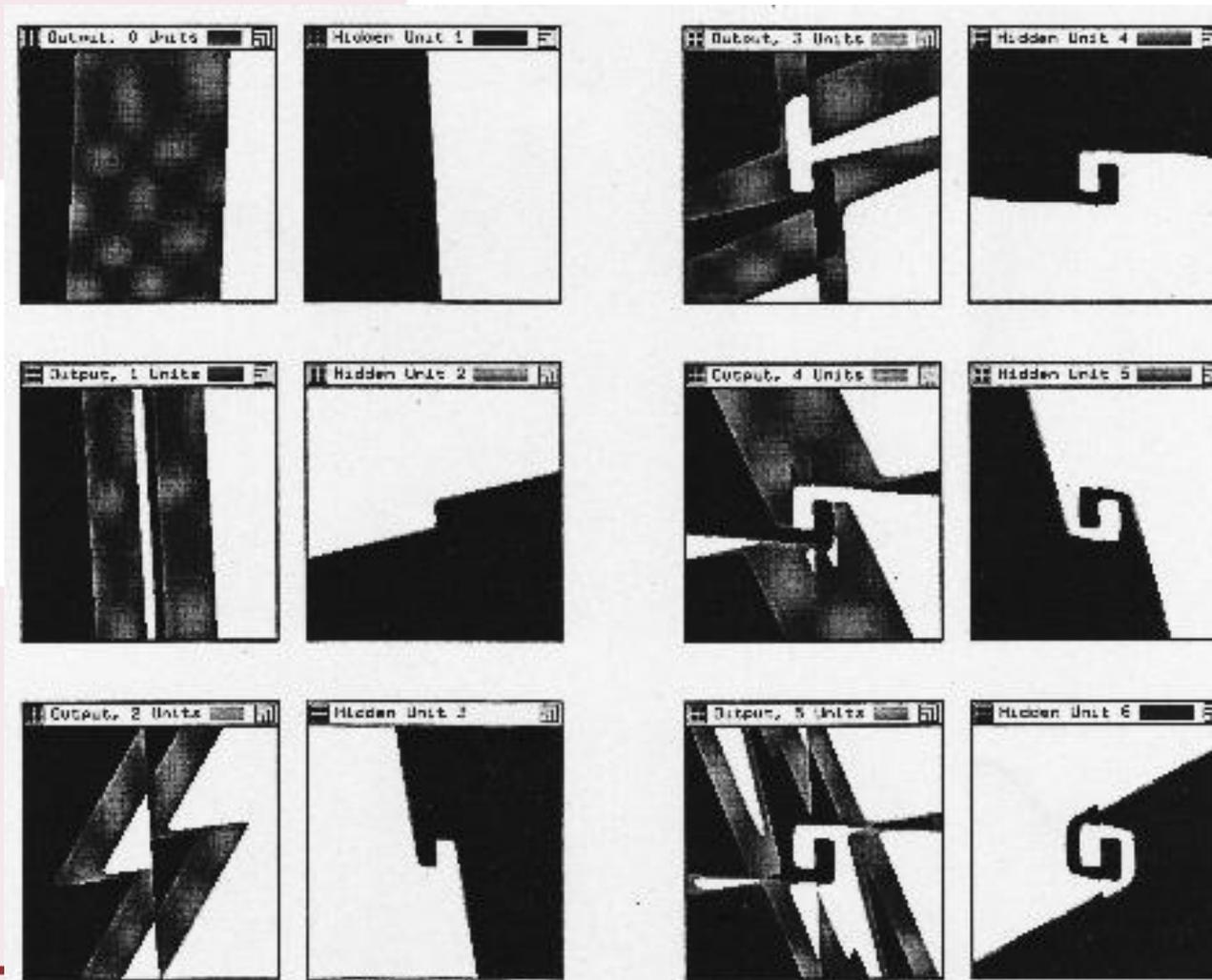


Exemplo: Problema das 2 Espirais

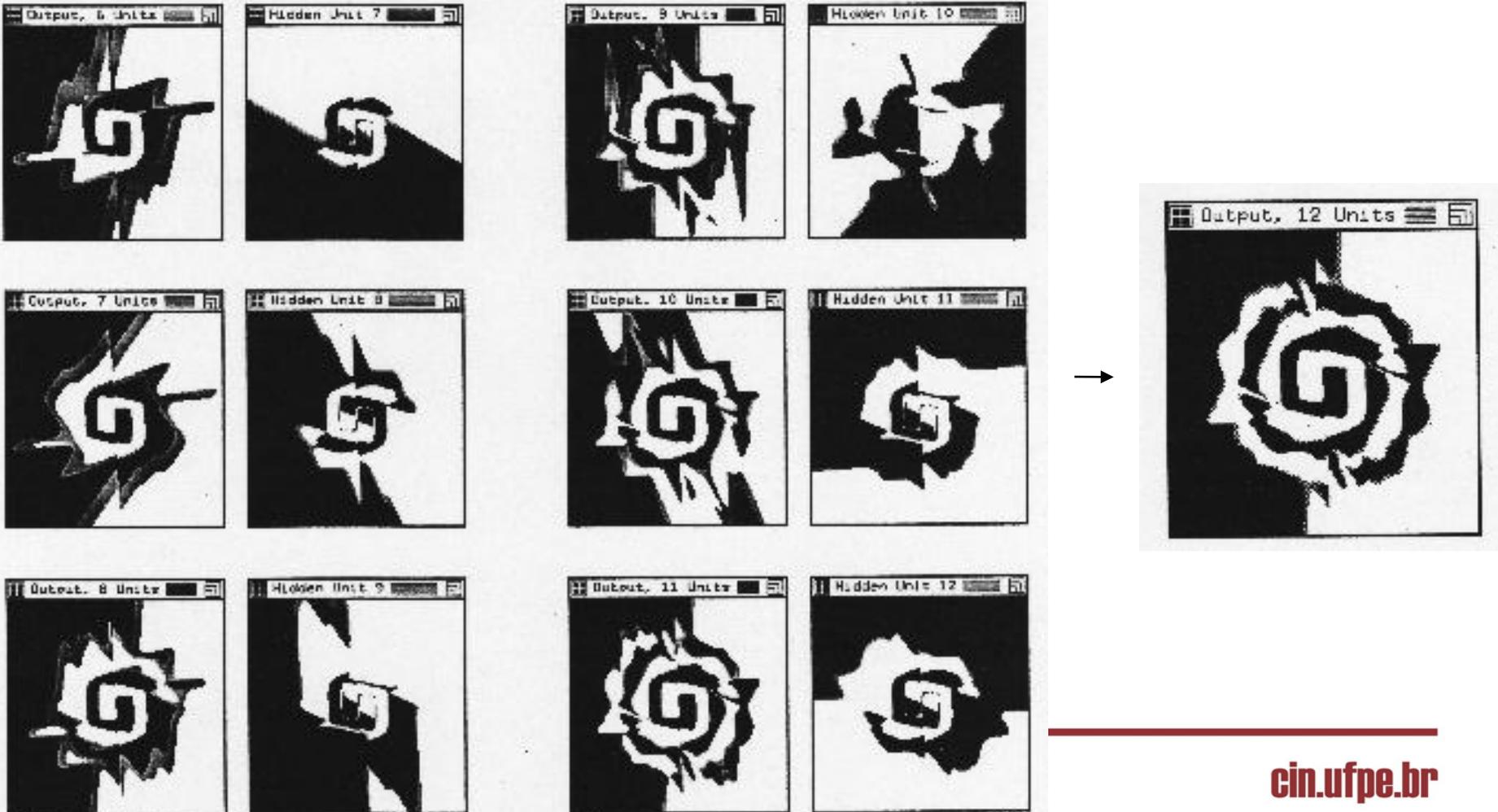
Um problema primário usado como benchmark para o backpropagation porque é um problema difícil de ser resolvido (treinado)



Problema das duas espirais



Problema das 2 Espirais



Cascade Correlation (CC)



Fahlman & Lebiere

Potenciais vantagens:

- > reduz o tempo de aprendizagem
- > transparente
- > cria uma rede estruturada



Cascade Correlation (CC)

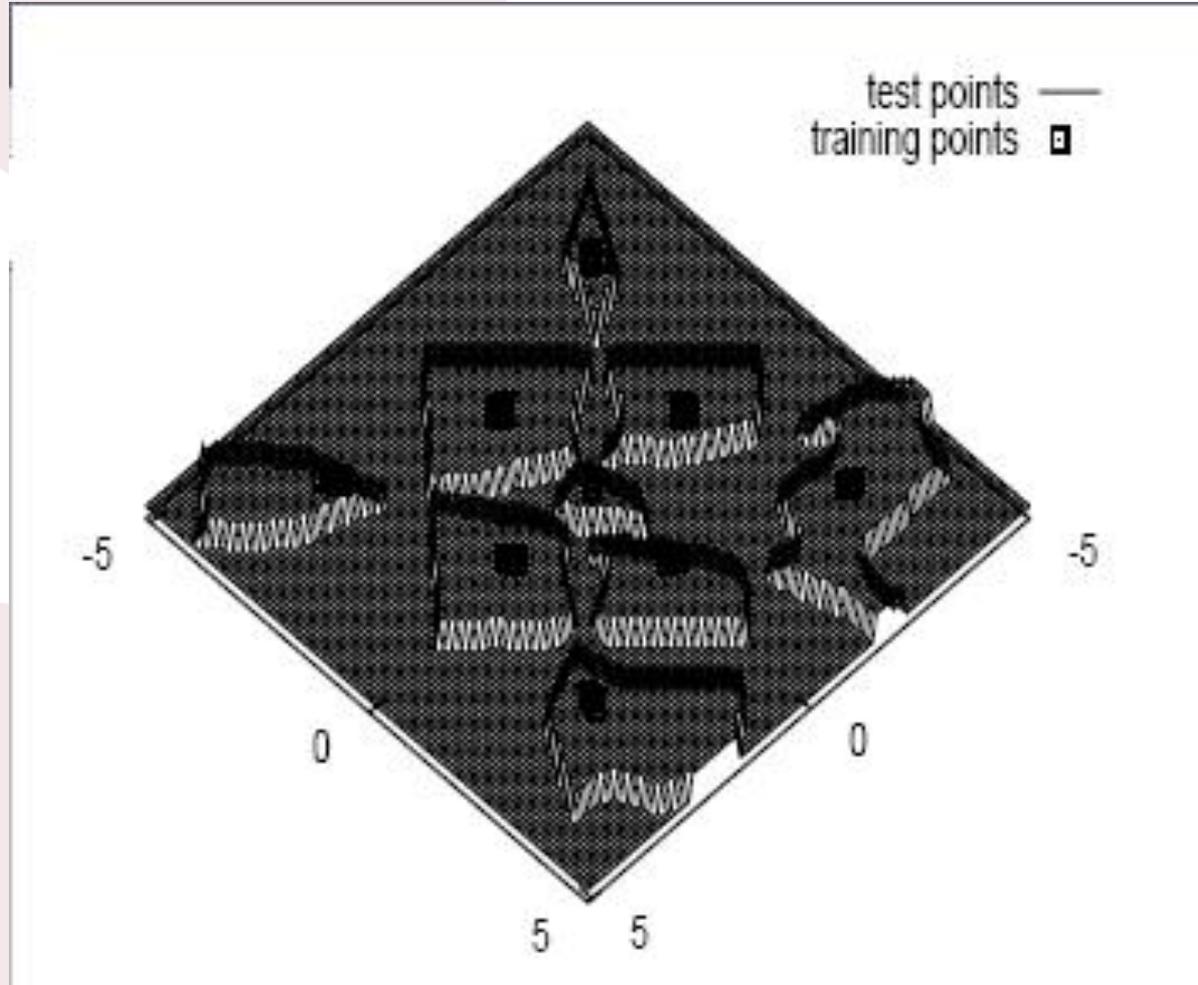


Desvantagem:

Pode levar à uma especialização maior do conjunto de treinamento



Exemplo de Regiões de Decisão: 9 classes



Applets Cascor



Arte com Cascor

- <http://Insclab.org/members/Frederic.Dandurand/art.html>

Cascor simulation applet

- http://www.psych.mcgill.ca/perpg/fac/shultz/cdp/Insc_applet.htm



A Arquitetura Tower



- Constrói uma torre de unidades
- Cada nova camada adicionada à rede recebe estímulos da camada de entrada e da camada imediatamente abaixo dela
- O algoritmo utilizado é do tipo perceptron (pocket with ratchet)



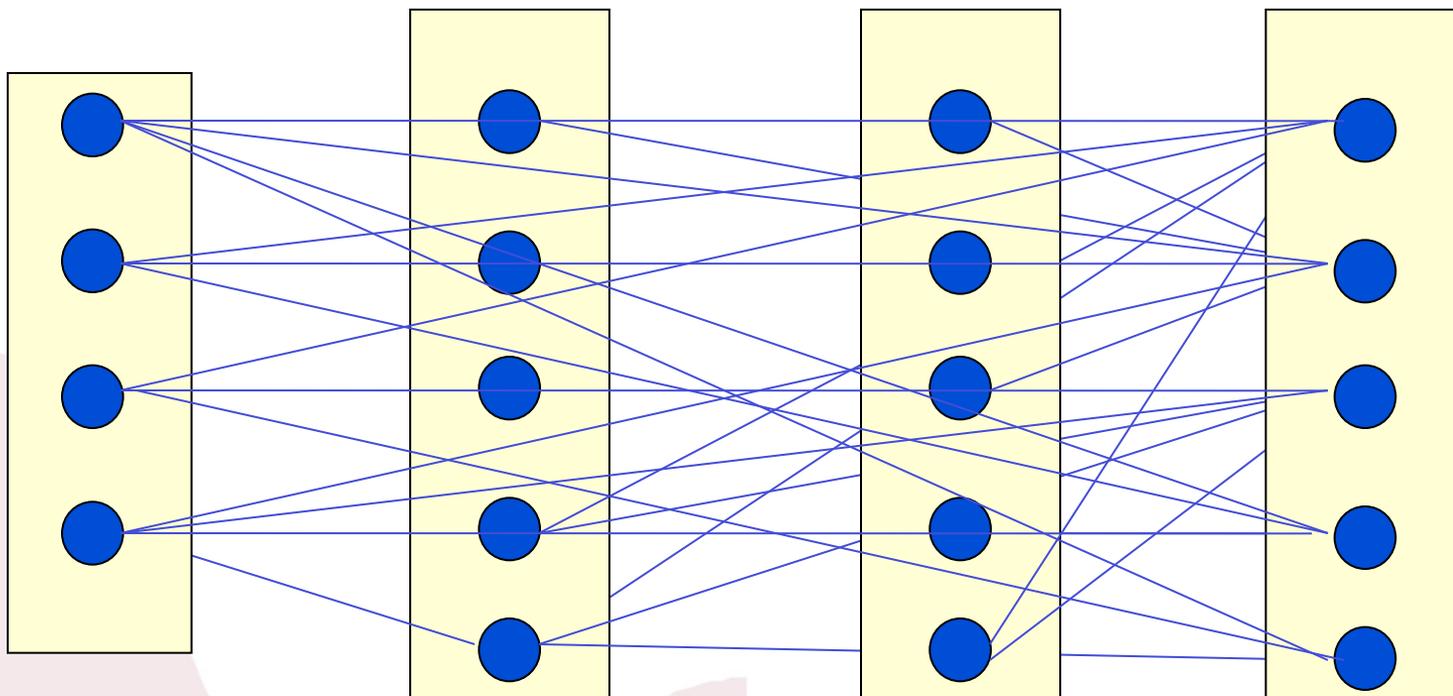
A Arquitetura Tower

Camada de Entrada

Camada Intermediária 1

Camada Intermediária 2

Camada de Saída



A Arquitetura Pyramid



- Constrói uma rede de forma similar à rede Tower
- Cada nova camada recebe estímulos tanto da camada de entrada como de todas as camadas previamente inseridas na rede
- A diferença chave em relação à Cascor é a forma de treinamento das novas unidades

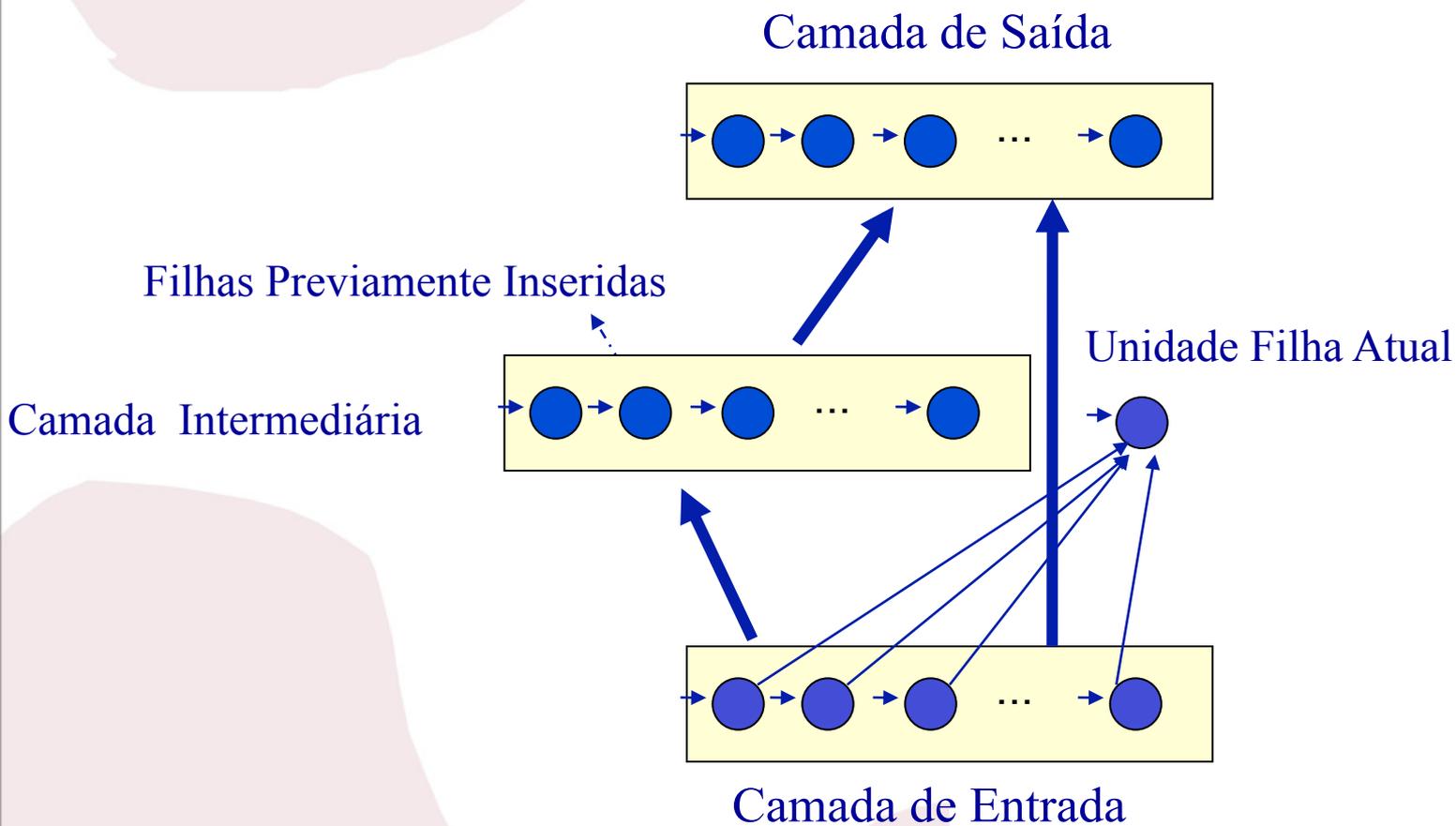
A Rede Upstart



- As unidades são inseridas em uma única camada
- Unidades “filhas” wrongly-on ou wrongly-off corrigem erros das unidades ascendentes que cometeram mais erros
- Algoritmo de treinamento tipo perceptron (Barycentric Correction Procedure - BCP)



A Rede Upstart



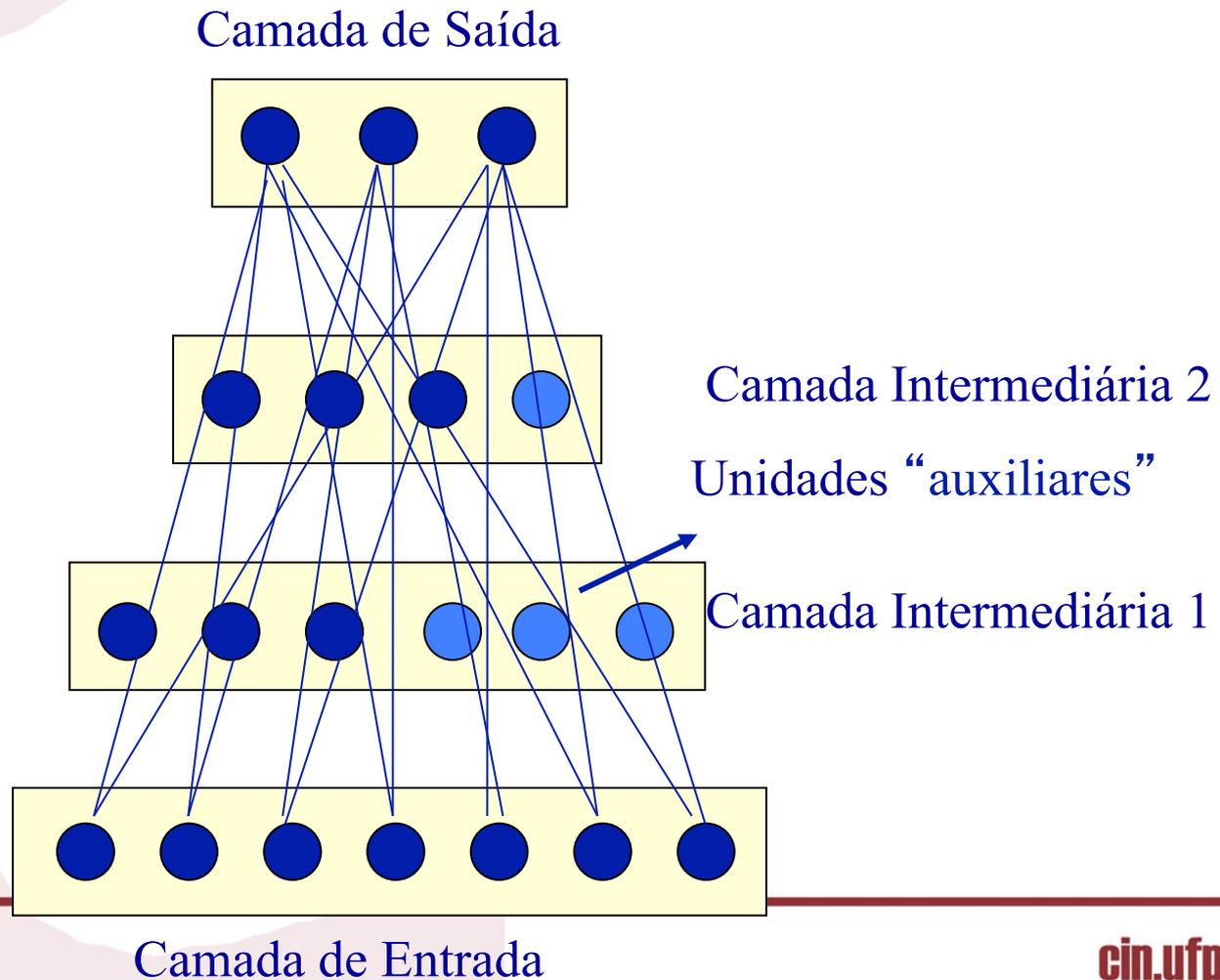
A Rede Tiling



- Unidades “mestres” são inseridas na rede de modo a classificar corretamente mais padrões que as unidades mestres das camadas anteriores
- Unidades auxiliares são inseridas nas camadas para se atingir uma representação fiel do problema



A Rede Tiling



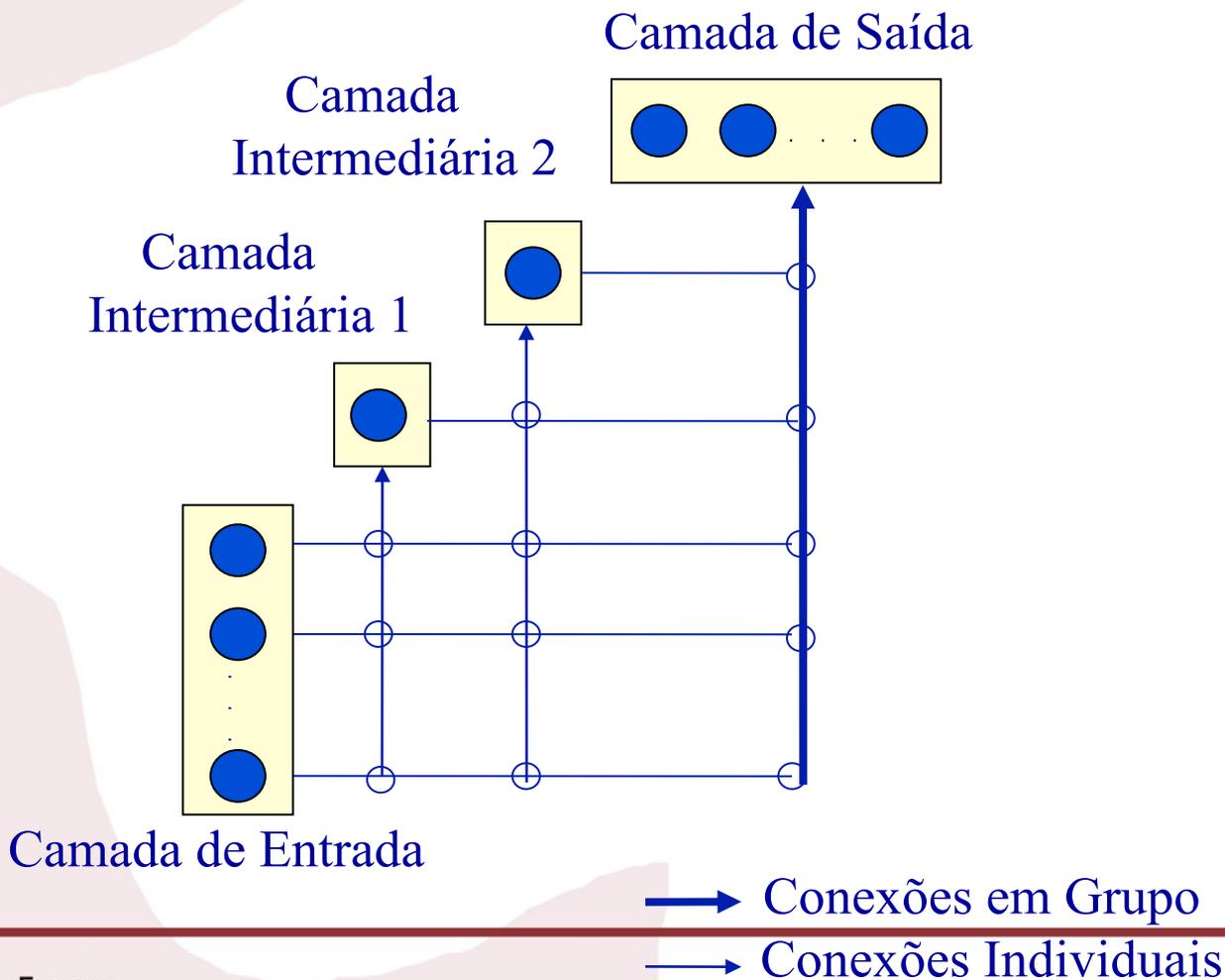
A Rede Perceptron-Cascade



- Baseia-se na estratégia da rede Upstart e constrói uma rede com topologia similar à Cascor
- Utiliza algoritmo do tipo perceptron (Thermal Perceptron)
- Convergência robusta, incluindo unidades de forma bastante otimizada



A Rede Perceptron-Cascade



A Rede DistAl



- Utiliza unidades esféricas definindo clusters
- Calcula distância entre pares de padrões, utilizando várias métricas de distância
- Tenta classificar o maior número de padrões possível com a inserção de cada unidade
- Treinamento do tipo one-shot



Problemas Práticos de Classificação



- Análise de crédito ao consumidor - CARD
- Diagnóstico de problemas cardíacos - HEART
- Reconhecimento de caracteres manuscritos



Problema Prático de Verificação Investigado



Verificação de Assinaturas



Classificação de Padrões (Estrutura Experimental)



- Os resultados relativos à MLP são extraídos do PROBEN1
- Foram realizadas simulações com as redes RBF seguindo o mesmo esquema experimental



Classificação de Padrões

(Estrutura Experimental)



- Investigação comparativa entre modelos construtivos e arquiteturas convencionais
- Propriedades de generalização, convergência, tamanhos de redes geradas e tempo de treinamento



Análise Comparativa dos Modelos (Classificação de Padrões)

REDE	UNID. INTERM	DESEMPENHO	DP
Cascor	8	84.00%	2.48
Tower	96	84.88%	2.41
Pyramid	96	86.05%	1.87
DistAl	27	84.35%	2.32
PercCasc	10	87.58%	2.12
Upstart	6	90.70%	2.03
Tiling	63	80.23%	2.74
MLP	8	86.00%	1.03
RBF	10	88.00%	0.67

Análise Comparativa dos Modelos

(Classificação de Padrões)



REDE	UNID. INTERM	DESEMPENHO	DP
Cascor	8	79.00%	1.87
Tower	84	78.49%	2.91
Pyramid	60	79.07%	2.97
DistAl	27	84.06%	4.10
PercCasc	10	86.80%	2.97
Upstart	7	86.05%	1.96
Tiling	51	79.05%	2.48
MLP	4	81.00%	0.86
RBF	40	82.00%	0.70



Análise Comparativa dos Modelos

(Classificação de Padrões)



REDE	UNID. INTERM	DESEMPENHO	DP
Cascor	8	82.00%	1.97
Tower	96	79.65%	2.26
Pyramid	60	79.07%	3.45
DistAl	27	82.90%	3.54
PercCasc	10	84.07%	2.11
Upstart	7	83.72%	2.53
Tiling	71	79.07%	2.08
MLP	24	81.00%	1.19
RBF	20	83.00%	0.70



Análise Comparativa dos Modelos (Classificação de Padrões)



REDE	UNID. INTERM	DESEMPENHO	DP
Cascor	8	82.00%	1.96
Tower	96	78.70%	2.73
Pyramid	96	79.57%	2.93
DistAI	40	79.13%	4.03
PercCasc	12	82.72%	2.64
Upstart	9	84.05%	2.48
Tiling	129	76.52%	1.98
MLP	8	80.00%	0.96
RBF	10	82.00%	1.20



Análise Comparativa dos Modelos

(Classificação de Padrões)



REDE	UNID. INTERM	DESEMPENHO	DP
Cascor	8	80.00%	1.66
Tower	60	78.26%	3.11
Pyramid	96	78.26%	2.69
DistAI	38	80.87%	2.92
PercCasc	12	82.56%	3.03
Upstart	9	81.45%	2.06
Tiling	101	71.74%	2.02
MLP	4	82.00%	1.14
RBF	20	82.00%	1.72



Análise Comparativa dos Modelos

(Classificação de Padrões)



REDE	UNID. INTERM	DESEMPENHO	DP
Cascor	8	74.00%	1.97
Tower	96	70.87%	3.36
Pyramid	96	76.09%	2.38
DistAI	37	82.61%	2.71
PercCasc	12	80.96%	2.02
Upstart	9	80.78%	2.65
Tiling	128	77.39%	2.18
MLP	24	76.00%	1.12
RBF	60	79.00%	0.48



Análise Comparativa dos Modelos

(Classificação de Padrões)

NC	Taxa de Reconhecimento			Taxa de Erro		
	Cascor	MLP	RBF	Cascor	MLP	RBF
0	91.1%	92.0%	91.1%	8.8%	8.0%	8.8%
0.15	84.7%	89.4%	85.7%	4.5%	5.3%	3.4%
0.25	83.2%	88.2%	80.3%	3.7%	5.3%	1.4%
0.35	81.4%	84.4%	74.9%	3.4%	3.7%	0.5%
0.45	80.3%	80.8%	66.7%	3.0%	2.6%	0.4%
0.55	79.1%	81.3%	55.8%	2.6%	3.0%	0.2%
0.65	77.8%	72.9%	-	2.2%	1.7%	-

Análise Comparativa dos Modelos

(Classificação de Padrões)



- Destaque das redes Perceptron-Cascade, Upstart, DistAl e Cascor

- As redes Tower e Pyramid consideram um fator de rejeição, tornando seus resultados mais robustos



Problema Prático de Verificação de Padrões



- Reconhecimento de Assinaturas
- Base de dados com imagens de assinaturas juntamente com características extraídas
- Exemplos de assinaturas genuínas, falsas habilidosas e falsas simples
- 21 classes foram consideradas

Exemplo de uma Classe de Assinaturas



Maria Salte Marcon

Assinatura Genuína

Maria Salte Marcon

Assinatura Falsa Habilidade

Maria Salte Marcon

Assinatura Falsa Simples



Arquiteturas Investigadas



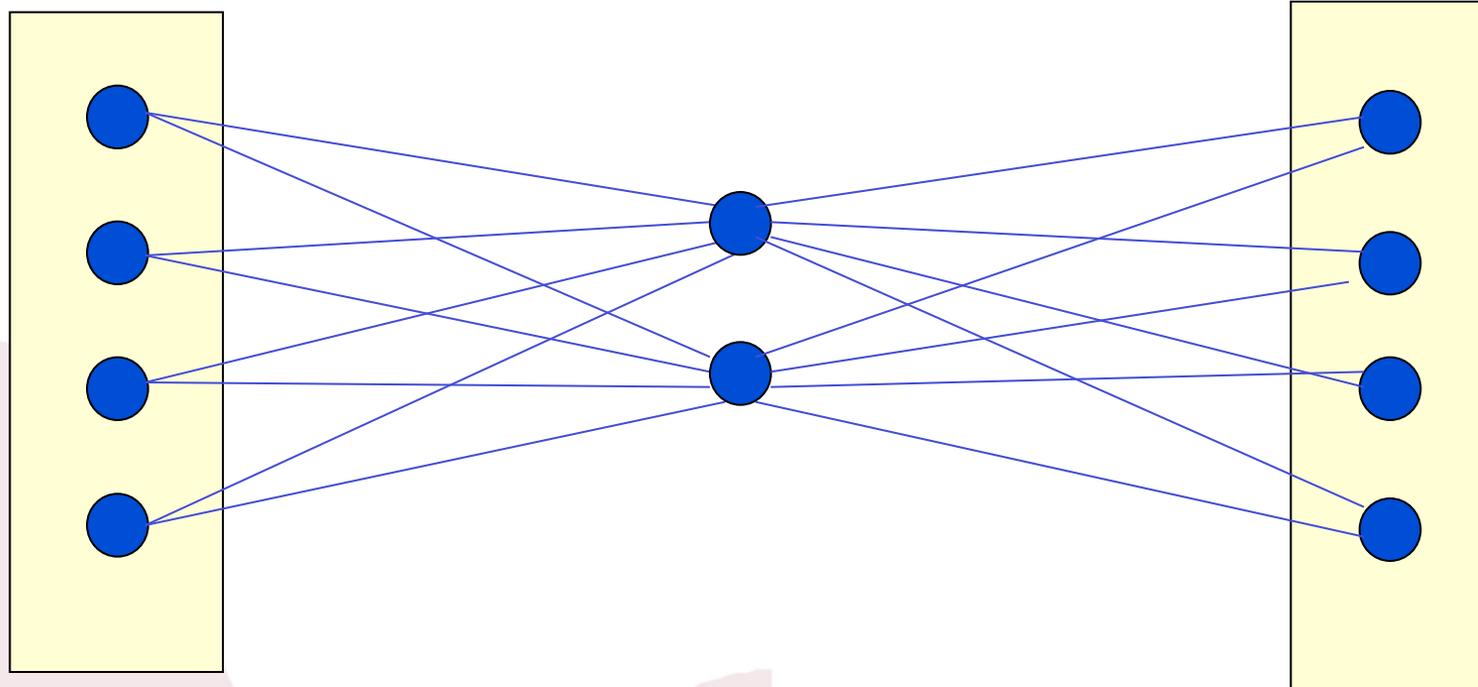
- Cascor autoassociativa
- MLP autoassociativa
- Tower-autoassociativa
- Pyramid-autoassociativa



Estrutura Autoassociativa

Camada de Entrada

Camada de Saída



Análise Comparativa dos Modelos

(Verificação de Padrões)



Rede	Unidades	I	II	III
MLP autoassociativa	10	60%	100%	100%
Cascor autoassociativa	8	70%	100%	100%
Tower-autoassociativa	6	100%	50%	100%
Pyramid-autoassociativa	12	100%	75%	100%

Desempenhos relativos à classe 0



Análise Comparativa dos Modelos

(Verificação de Padrões)



Rede	Unidades	I	II	III
MLP autoassociativa	10	90%	75%	100%
Cascor autoassociativa	5	70%	50%	100%
Tower-autoassociativa	12	100%	75%	100%
Pyramid-autoassociativa	12	80%	75%	75%

Desempenhos relativos à classe 5



Análise Comparativa dos Modelos (Verificação de Padrões)

Rede	Unidades	I	II	III
MLP autoassociativa	10	80%	75%	100%
Cascor autoassociativa	8	70%	75%	100%
Tower-autoassociativa	12	90%	75%	100%
Pyramid-autoassociativa	12	90%	75%	75%

Desempenhos relativos à classe 20

Conclusões



- Para problemas de Classificação, as redes construtivas se mostraram eficientes
- Melhores resultados com redes Perceptron-Cascade, Upstart, DistAI e Cascor
- Resultados muito próximos e muitas vezes superiores aos da MLP e a RBF



Conclusões



- Em relação à taxa de reconhecimento de assinaturas genuínas, as redes construtivas obtiveram, em geral, resultados melhores que a rede MLP autoassociativa (70% a 100%)



Conclusões

- A rede Pyramid-autoassociativa obteve os melhores resultados
- Em relação às taxas de rejeição de assinaturas falsas simples, as redes apresentaram resultados bastante satisfatórios e semelhantes aos da MLP (entre 75% e 100%)

Conclusões

Modelos construtivos podem seguramente ser considerados uma alternativa aos modelos mais tradicionais de redes feedforward

