

Universidade Federal de Itajubá  
Mestrado em Ciência e Tecnologia da Computação

# Avaliação de Técnicas de Aprendizado de Máquina no combate ao SPAM

Marcelo Vinícius Cysneiros Aragão – 34759

# Índice

- Introdução
- Definição do Problema
- Aprendizado de Máquina
  - Visão Geral
  - *MLP (Multilayer Perceptron)*
  - *RBF (Radial Basis Function) Network*
  - *SVM (Support Vector Machine)*
- Seleção de Características e *Data Sets*
- Bibliotecas e Medidas de Desempenho
- Experimentos e Resultados
- Conclusões
- Referências Bibliográficas



# Introdução

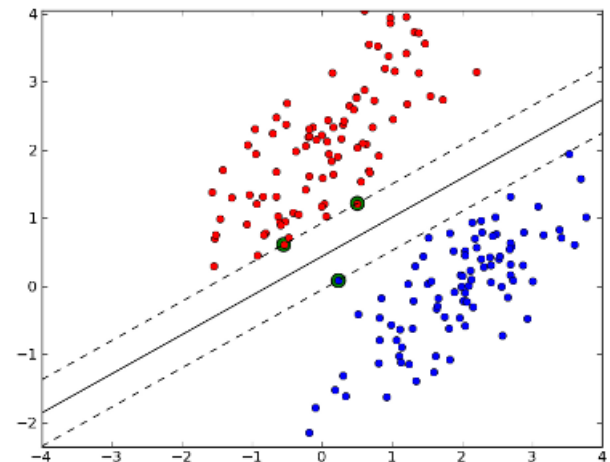
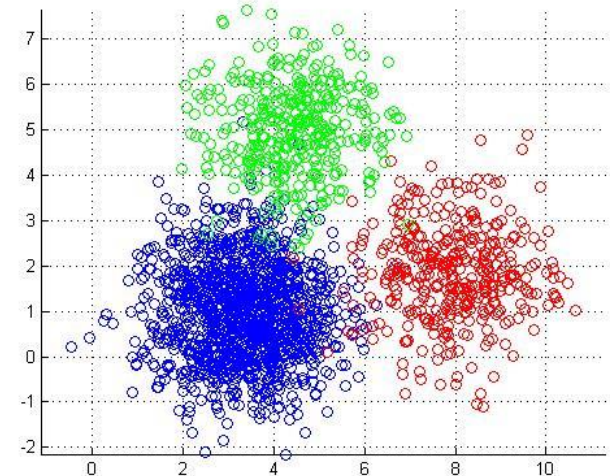
- Área de Concentração / Linha de Pesquisa
  - Matemática da Computação / Inteligência Artificial
- Histórico
  - Graduação → pré-processamento de SMS (através de expressões regulares) e classificação utilizando SVM
  - Projeto de pesquisa → pré-processamento de e-mails (cabeçalho/corpo/anexos) e classificação utilizando SOM
  - Mestrado → avaliação de diferentes técnicas de aprendizado de máquina e possível integração com o SAS (Isaac)

# Definição do Problema

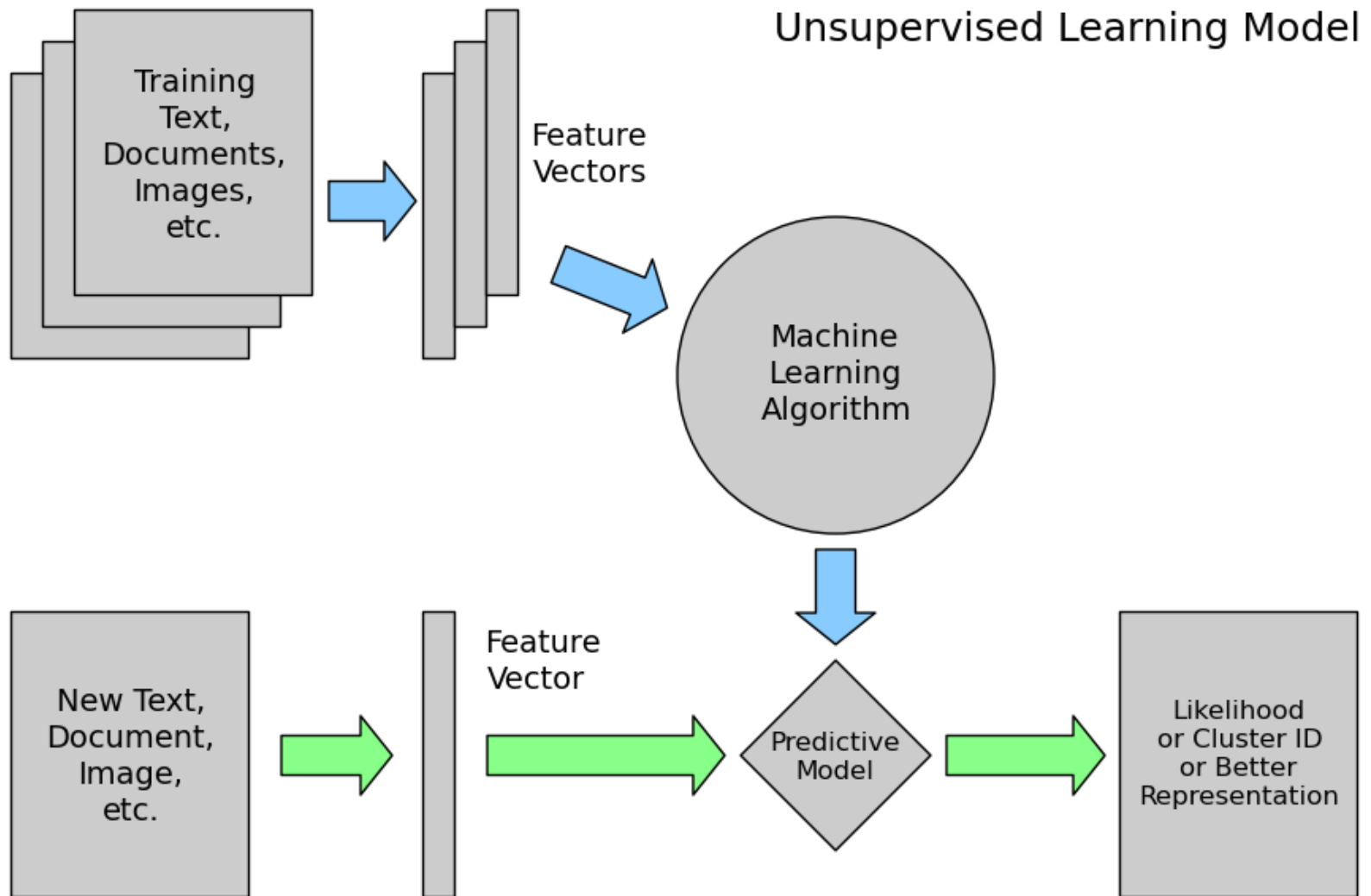
- Segundo pesquisa da Symantec [1], o volume de mensagens indesejadas que circularam pela rede em 2013 ultrapassou 66%.
- *SPAM*: mensagem não-requisitada e/ou enviada em massa
- *HAM*: qualquer mensagem que não seja *SPAM*
- Padrões que caracterizam *SPAM* são difíceis de serem detectados, tanto pelo homem quanto por computadores.
- Técnicas de *Blacklisting* possuem desvantagens intrínsecas (alto custo, legitimidade duvidosa)

# Aprendizado de Máquina

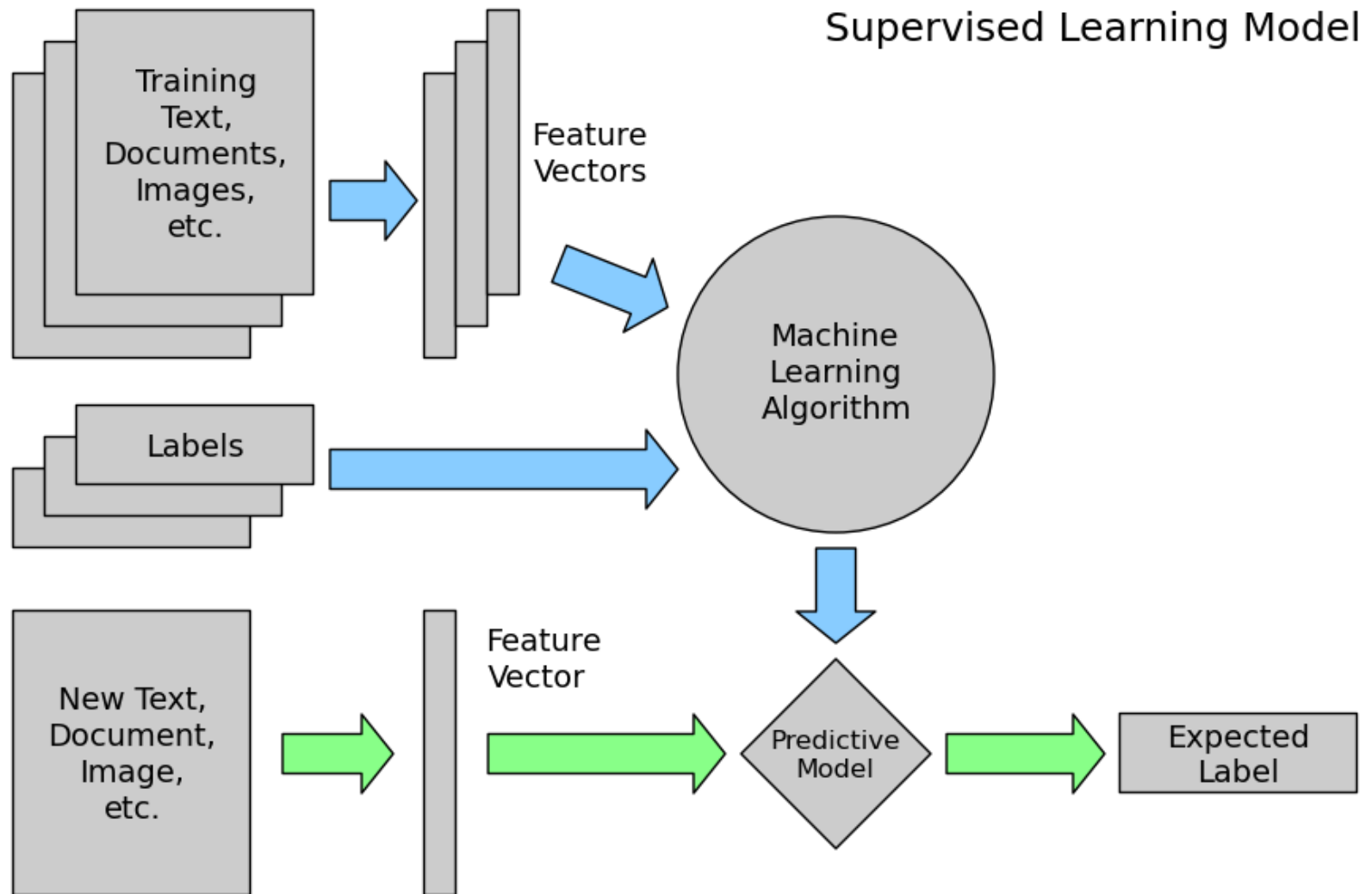
- Aprendizado não-supervisionado
  - Algoritmos de agrupamento
    - K-Means
    - Expectation-Maximization (EM)
  - Self-Organizing Map (SOM)
  - Adaptive Resonance Theory (ART)
- Aprendizado supervisionado
  - Artificial Neural Networks
    - Multi Layer Perceptron (MLP)
    - Radial Basis Function Networks (RBF)
  - Support Vector Machine (SVM)



# Aprendizado de Máquina

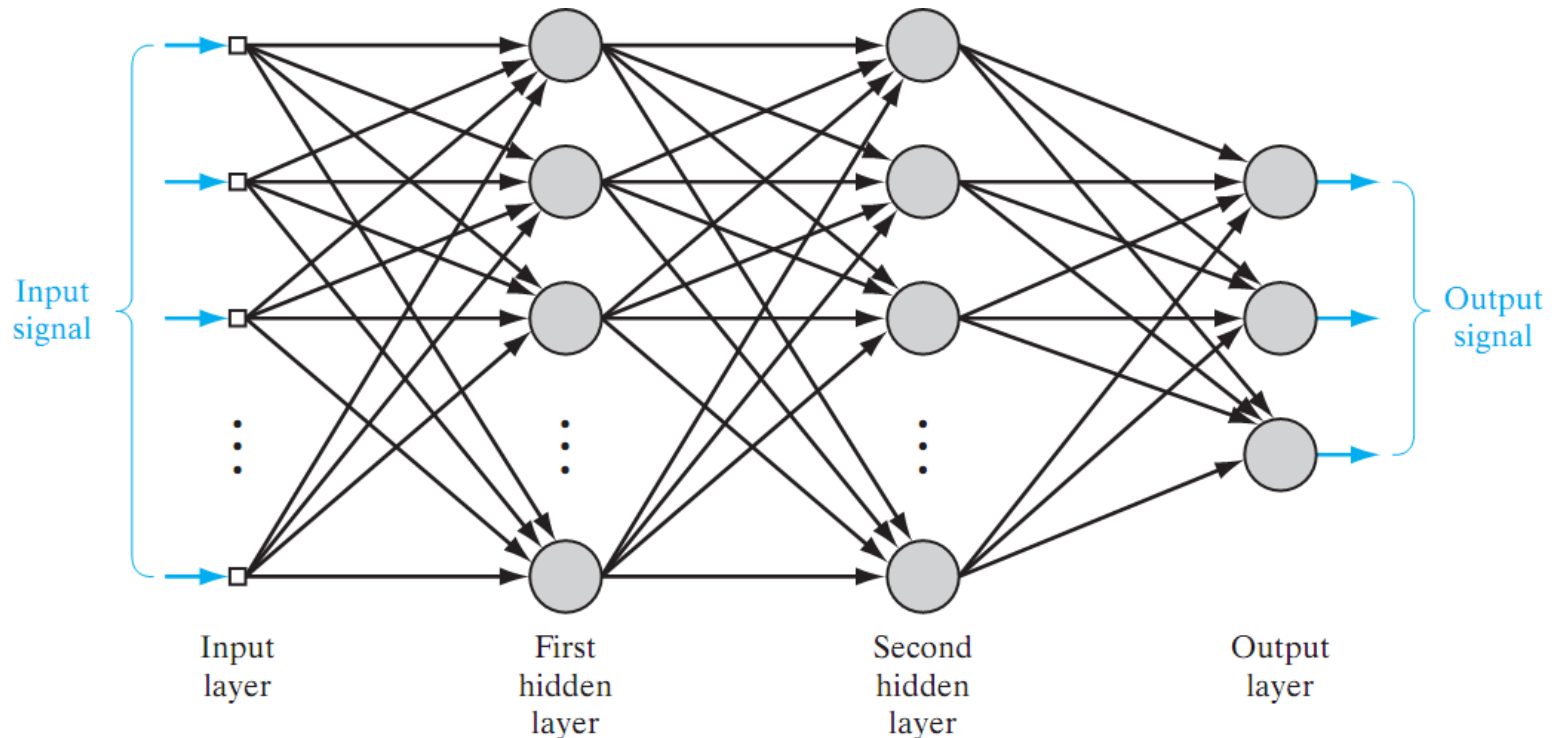


# Aprendizado de Máquina



# Multi Layer Perceptron (MLP)

- Possui pelo menos uma camada intermediária (escondida)
- É aplicável em diversos tipos de problemas
- Popularizou-se a partir do fim dos anos 80, com a publicação do livro *Parallel Distributed Processing* (Rumelhart et alii., 1988) [2].





# MLP: Backpropagation

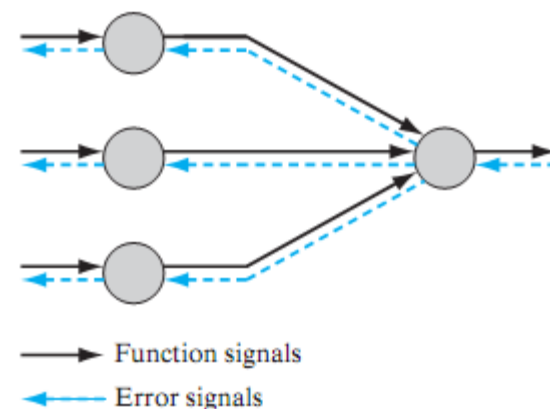
- Algoritmo mais comum de treinamento de redes MLP
- Também conhecido como Regra Delta generalizada

- Etapa *feed-forward*

- Propagação dos sinais de uma amostra desde a primeira até a última camada, visando obter as respostas da rede.
- Não há alteração de pesos sinápticos e limiares dos neurônios.
- É quando são calculados os erros de cada época.

- Etapa *back-propagation*

- Os erros calculados na etapa anterior são propagados de trás para frente, ajustando os pesos sinápticos e limiares de ativação.
- Implica na diminuição iterativa do erro da rede.



$$\begin{pmatrix} \text{Weight} \\ \text{correction} \\ \Delta w_{ji}(n) \end{pmatrix} = \begin{pmatrix} \text{learning-} \\ \text{rate parameter} \\ \eta \end{pmatrix} \times \begin{pmatrix} \text{local} \\ \text{gradient} \\ \delta_j(n) \end{pmatrix} \times \begin{pmatrix} \text{input signal} \\ \text{of neuron } j, \\ y_i(n) \end{pmatrix}$$

# MLP: Resilient Propagation

- Proposto por Martin Riedmiller e Heinrich Braun em 1993 [3], é uma melhoria do algoritmo *Backprop* no sentido de agilizar a convergência.
- Rprop<sup>-</sup> (sem *weight backtracking*)
  - Considera a variação do sinal do erro (ao invés da magnitude)
  - Caso o sinal seja alterado, o peso é multiplicado por  $\eta^-$ , onde  $0 < \eta^- < 1$
  - Caso o sinal permaneça igual, o peso é multiplicado por  $\eta^+$ , onde  $\eta^- < 1 < \eta^+$

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^+ * \Delta_{ij}^{(t-1)} & , \text{ if } \frac{\partial E}{\partial w_{ij}}^{(t-1)} * \frac{\partial E}{\partial w_{ij}}^{(t)} > 0 \\ \eta^- * \Delta_{ij}^{(t-1)} & , \text{ if } \frac{\partial E}{\partial w_{ij}}^{(t-1)} * \frac{\partial E}{\partial w_{ij}}^{(t)} < 0 \\ \Delta_{ij}^{(t-1)} & , \text{ else} \end{cases}$$

- Rprop<sup>+</sup> (com *weight backtracking*)
  - Caso a atualização do peso seja grande o suficiente para alterar seu sinal (“saltou sobre o mínimo”), então ela deve ser revertida

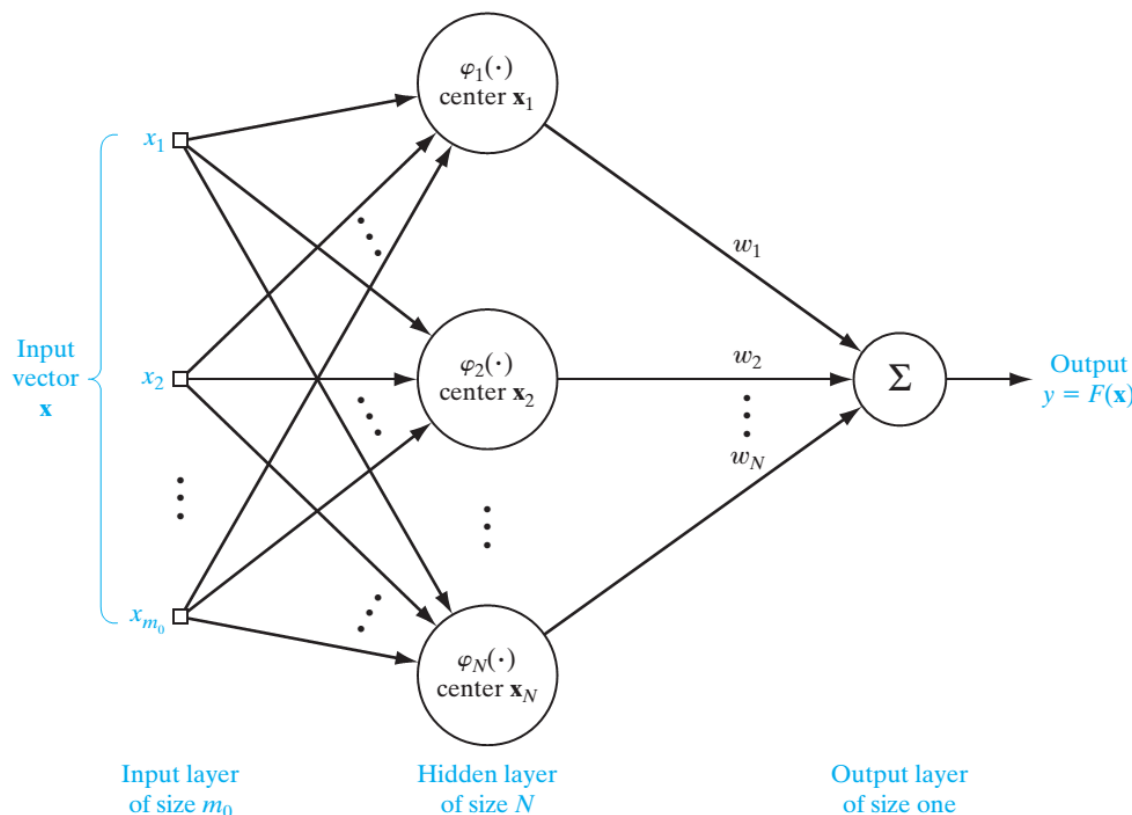
$$\Delta w_{ij}^{(t)} = -\Delta w_{ij}^{(t-1)} , \text{ if } \frac{\partial E}{\partial w_{ij}}^{(t-1)} * \frac{\partial E}{\partial w_{ij}}^{(t)} < 0$$

# MLP: iRprop<sup>+</sup> e iRprop<sup>-</sup>

- Proposto por Christian Igel e Michael Hüsken em 2000 [4], é uma melhoria (das duas versões) do algoritmo *Resilient Propagation*.
- Argumento: “o algoritmo *Rprop* se baseia na consideração de que uma mudança no sinal da derivada parcial implica que este saltou um mínimo local, mas não indica se a atualização do peso causou um aumento ou diminuição do erro”.
- Proposta: armazenar o erro da época anterior e/ou zerar o valor da derivada.
- iRprop<sup>-</sup> (sem *weight backtracking*):  $\frac{\partial E}{\partial w_{ij}}^{(t)} := 0$ 
  - Strongly influences the learning speed of the *Rprop* algorithm
- iRprop<sup>+</sup> (com *weight backtracking*):  $\text{if } E^{(t)} > E^{(t-1)} \text{ then } w_{ij}^{(t+1)} := w_{ij}^{(t)} - \Delta w_{ij}^{(t-1)}$ 
  - Combined with the improved weight-backtracking this yields the best performing algorithm in our investigation, iRprop<sup>+</sup>.

# RBF (Radial Basis Function)

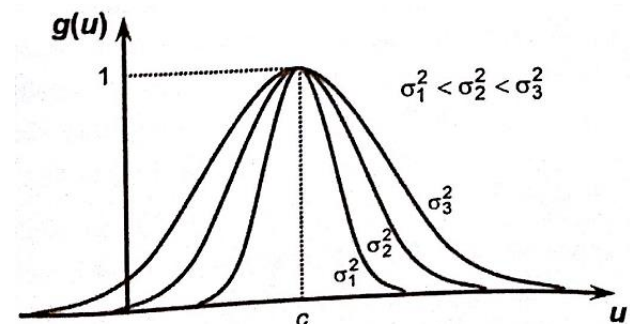
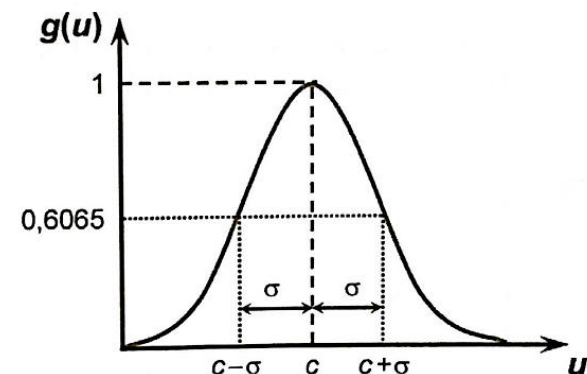
- Podem ser empregadas nos tipos de problemas tratados pelo MLP.
- A estrutura típica de uma rede RBF é composta por apenas uma camada intermediária, na qual as funções de ativação são do tipo gaussiana.



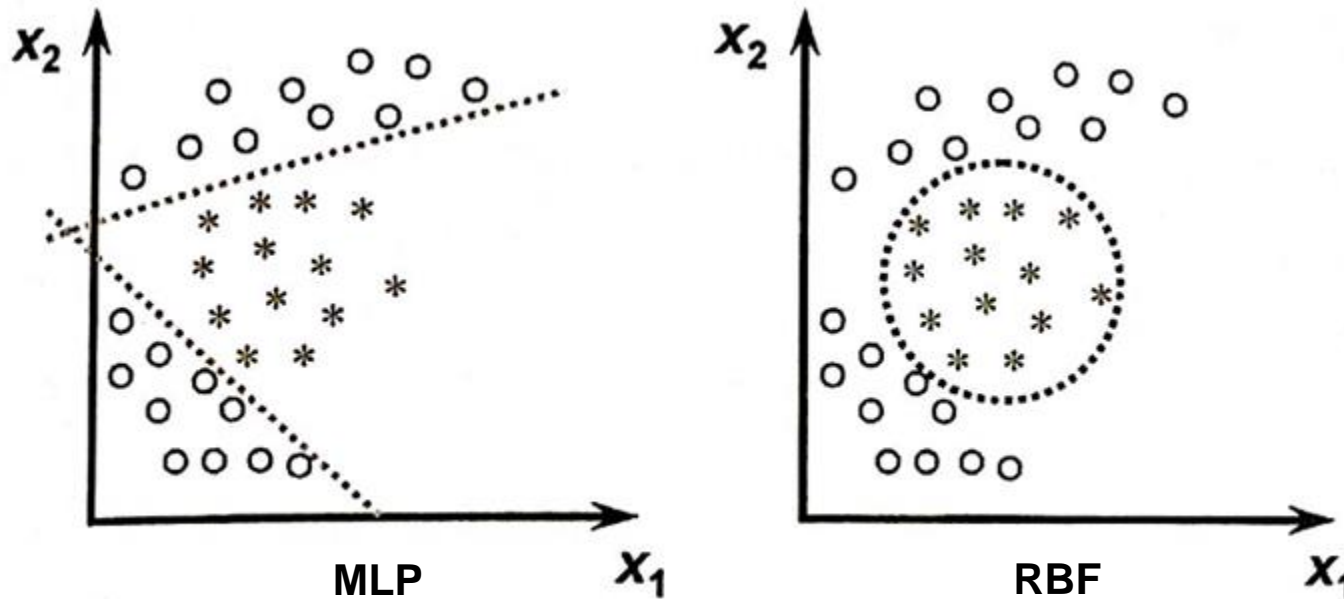
# RBF (Radial Basis Function)

- Assim como no MLP (*backprop*), o treinamento das redes RBF é realizado em duas etapas.
- Entretanto, o processo de treinamento inicia-se pelos neurônios da camada intermediária, e encerra-se nos neurônios da camada de saída.
- Etapa não-supervisionada
  - Ajustes dos pesos sinápticos da camada intermediária levando em conta somente as características dos dados de entrada.
  - Estes ajustes estão relacionados à alocação das funções de base radial e definição de seus parâmetros (centro e variância).
- Etapa supervisionada
  - Ajuste dos pesos sinápticos e limiares de ativação da camada de saída.
  - Feito através da aplicação da Regra Delta generalizada.

$$g(u) = e^{-\frac{(u-c)^2}{2\sigma^2}}$$



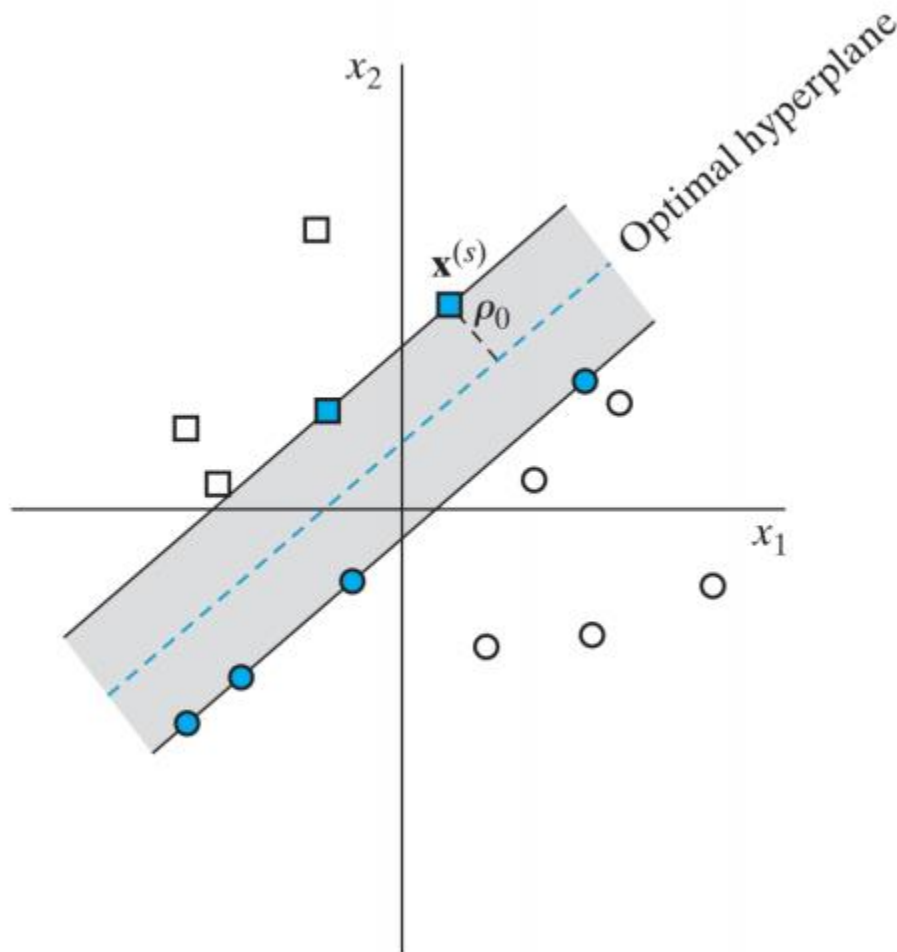
# MLP *versus* RBF



- “O MLP computa as fronteiras de delimitação de classes por intermédio de uma combinação de hiperplanos.”
- “Já na RBF, as fronteiras delimitadoras são definidas por campos receptivos hiperesféricos, considerando a distância radial em que os padrões se encontram em relação ao centro das hiperesferas.”

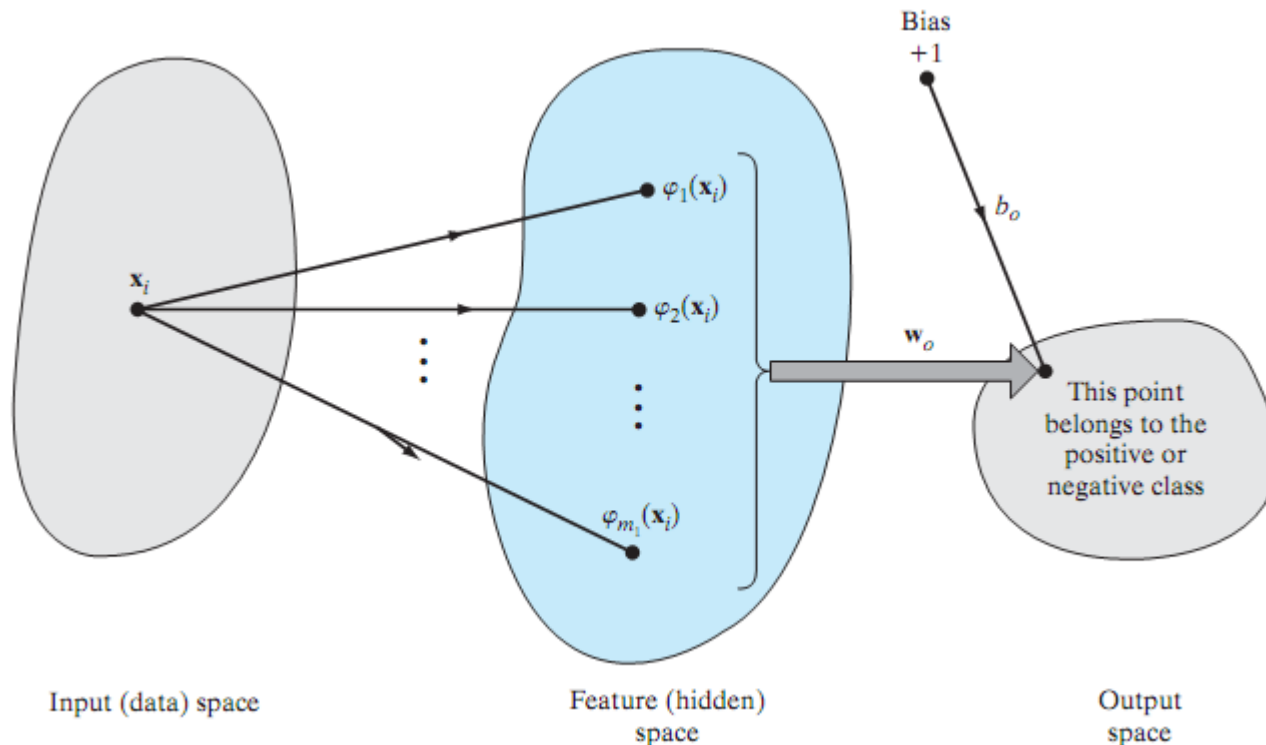
# SVM (Support Vector Machine)

- É um modelo de aprendizado supervisionado que resolve problemas de identificação de padrões, interpolação e/ou extrapolação.
- Dado um conjunto de exemplos de treinamento, a SVM constrói um hiperplano como uma superfície decisória de tal forma que a margem de separação entre os exemplos positivos e negativos seja maximizada.



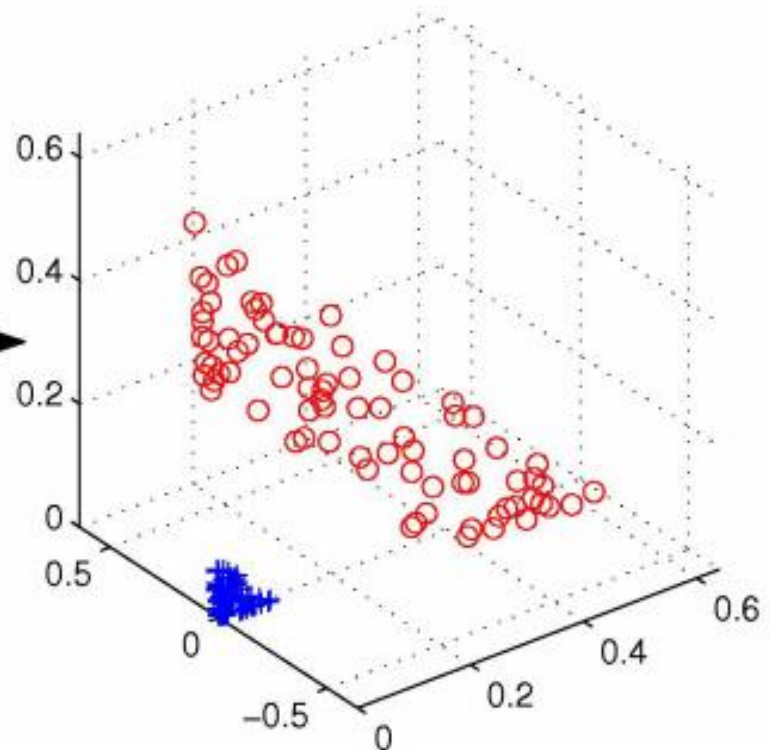
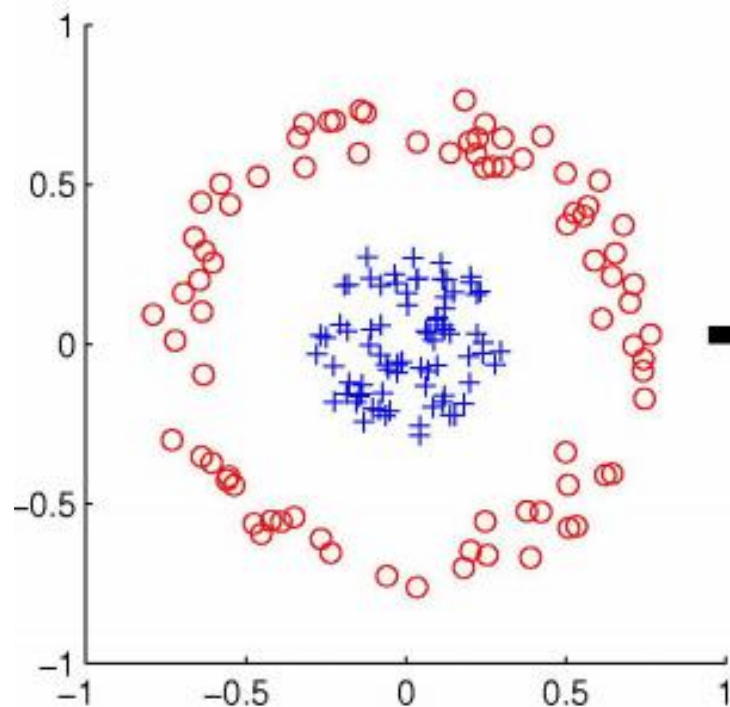
# SVM (Support Vector Machine)

- Ideia central:
  - Utilizar transformações matemáticas no espaço de entrada não-linearmente separável para obter um espaço de características linearmente separável.





# SVM (Support Vector Machine)



**Padrões não-linearmente separáveis em  $R^2$**

**Padrões linearmente separáveis em  $R^3$**

# SVM (Support Vector Machine)

- Os dados da figura da esquerda não são linearmente separáveis.
- Entretanto, pode-se mapeá-los para um espaço tridimensional e tentar separar os dados de forma linear.

$$\begin{aligned}\phi: \quad \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ (x_1, x_2) &\longmapsto (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)\end{aligned}$$

- Desta forma, as fronteiras de decisão serão hiperplanos em  $\mathbb{R}^3$ .
- O que usamos é o é a Matriz de Gram K dos dados.
- Uma vez que sabemos K, podemos descartar os dados originais.

$$K = \begin{bmatrix} x_1^T x_1 & x_1^T x_2 & \cdots \\ x_2^T x_1 & \ddots & \\ \vdots & & \end{bmatrix}_{n \times n} = XX^T$$

- Podemos utilizar o nosso algoritmo linear sobre uma versão dos dados transformados para obter um algoritmo não linear com pouco esforço. [11]

# SVM (Support Vector Machine)

- Exemplo: problema XOR

Pode-se utilizar um núcleo polinomial:

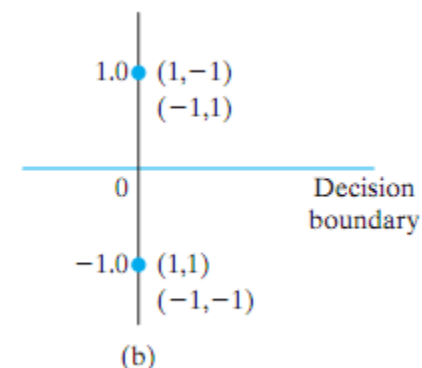
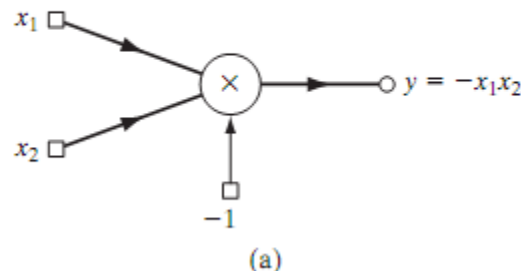
$$k(\mathbf{x}, \mathbf{x}_i) = 1 + (\mathbf{x}^T \mathbf{x}_i)^2$$

$$k(\mathbf{x}, \mathbf{x}_i) = 1 + x_1^2 x_{i1}^2 + 2x_1 x_2 x_{i1} x_{i2} + x_2^2 x_{i2}^2 + 2x_1 x_{i1} + 2x_2 x_{i2}$$

O que resulta no seguinte espaço de características:

$$\Phi(\mathbf{x}) = [1, x_1^2, \sqrt{2}x_1 x_2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2]^T$$

Saída:  $y = -x_1 x_2$



Input vector $\mathbf{x}$	Desired response $d$
$(-1, -1)$	$-1$
$(-1, +1)$	$+1$
$(+1, -1)$	$+1$
$(+1, +1)$	$-1$

# Seleção de Características

- Como representar um e-mail de forma conveniente para alimentar um método de aprendizado de máquina?
- Através de métodos estatísticos para seleção de características. [5]
- Foram gerados vetores com 8, 16, ... , 512 e 1024 características.
- Chi-Quadrado (CHI2)
  - Utilizado para testar se a ocorrência de um termo específico e a ocorrência de uma classe específica são independentes. Termos dependentes devem ser considerados.
- Informação Mútua (MI)
  - É uma medida de dependência mútua de duas variáveis aleatórias.
- Seleção de Características Baseada na Frequência (FD / DF)
  - Pode ser definida como a frequência de documentos (o número de documentos na classe que contêm o termo) ou como frequência em uma coleção (o número tokens  $t$  que ocorrem nos documentos em  $c$ ).

# Conjuntos de Dados

- Tendo em mãos:
  - Técnicas de Aprendizado de Máquina
  - Técnicas de Seleção de Características
- São necessários dados para executar os experimentos
  - Data Set UNIFEI
    - E-mails coletados no próprio servidor da universidade
    - É composto por 4.194 *hams* e 1.892 *spams* (2,2 : 1)
  - Data Set SpamAssassin [6]
    - Repositório público disponível na Web
    - É composto por 4.420 *hams* e 15.857 *spams* (1 : 3,6)

# Libs e Medidas de Desempenho

- Foram utilizadas as seguintes bibliotecas para execução dos testes:

Técnica	Método de Treinamento	Biblioteca(s)	Linguagem
MLP	<i>Backpropagation</i>	Otávio / Isaac	Java
	<i>Resilient Propagation</i>	Encog [7]	
RBF	-	Encog [7]	
SVM	-	LibSVM [8]	

- Para avaliar os resultados, foram consideradas as seguintes métricas:
  - Tipo de Característica (CHI2, FD, MI)
  - Quantidade de características (8, 16, 32, 64, 128, 256, 512, 1024)
  - Precisão na classificação (%)
  - Erro na classificação (%)
  - Tempo de treinamento (hh:min:seg)
  - Tempo de classificação (hh:min:seg)
  - Tempo total (hh:min:seg)

# Experimentos

- Foram executadas baterias de teste, chamadas por arquivos *batch* do Windows.
- Para cada técnica, foram realizados dois tipos de teste:
  - 100/100: avalia a capacidade da técnica em aprender e generalizar todo o conteúdo que lhe foi apresentado.
  - 50/50: avalia a capacidade da técnica em aprender com metade do conteúdo e inferir sobre a outra metade.
- No caso do SVM, também foram realizados mais testes:
  - *One Class* (HAM): treina-se HAM e classifica-se SPAM
  - *One Class* (SPAM) treina-se SPAM e classifica-se HAM

# Resultados

## MLP Backprop (Otávio/Isaac)

SpamAssassin_MLP_100.log.min							SpamAssassin_MLP_50.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total	Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	70,31	29,69	00:00:04	00:00:00	00:00:04	CHI2	8	70,28	29,72	00:00:02	00:00:00	00:00:02
CHI2	16	71,07	28,93	00:00:04	00:00:00	00:00:04	CHI2	16	70,93	29,07	00:00:02	00:00:00	00:00:02
CHI2	32	71,10	28,90	00:00:06	00:00:00	00:00:06	CHI2	32	71,21	28,79	00:00:02	00:00:00	00:00:02
CHI2	64	71,81	28,19	00:00:07	00:00:00	00:00:07	CHI2	64	71,65	28,35	00:00:04	00:00:00	00:00:04
CHI2	128	44,30	55,70	00:00:11	00:00:00	00:00:11	CHI2	128	77,24	22,76	00:00:06	00:00:00	00:00:06
CHI2	256	85,41	14,59	00:00:22	00:00:00	00:00:22	CHI2	256	86,57	13,43	00:01:03	00:00:00	00:01:03
CHI2	512	83,78	16,22	00:00:55	00:00:00	00:00:55	CHI2	512	86,32	13,68	00:01:30	00:00:00	00:01:30
CHI2	1024	85,55	14,45	00:01:54	00:00:00	00:01:54	CHI2	1024	86,21	13,79	00:02:06	00:00:00	00:02:06
FD	8	62,13	37,87	00:00:32	00:00:00	00:00:32	FD	8	52,80	47,20	00:00:05	00:00:00	00:00:05
FD	16	81,95	18,05	00:01:05	00:00:00	00:01:05	FD	16	67,61	32,39	00:00:03	00:00:00	00:00:03
FD	32	82,62	17,38	00:00:23	00:00:00	00:00:23	FD	32	73,15	26,85	00:00:03	00:00:00	00:00:03
FD	64	85,54	14,46	00:00:50	00:00:00	00:00:50	FD	64	49,77	50,23	00:00:06	00:00:00	00:00:06
FD	128	75,86	24,14	00:00:13	00:00:00	00:00:13	FD	128	87,31	12,69	00:00:39	00:00:00	00:00:39
FD	256	77,41	22,59	00:00:20	00:00:00	00:00:20	FD	256	86,80	13,20	00:00:41	00:00:00	00:00:41
FD	512	81,30	18,70	00:01:12	00:00:00	00:01:12	FD	512	83,45	16,55	00:00:46	00:00:00	00:00:46
FD	1024	92,89	7,11	00:27:10	00:00:00	00:27:10	FD	1024	90,03	9,97	00:05:27	00:00:00	00:05:27
MI	8	58,77	41,23	00:00:05	00:00:00	00:00:05	MI	8	76,89	23,11	00:00:08	00:00:00	00:00:08
MI	16	78,55	21,45	00:00:09	00:00:00	00:00:09	MI	16	81,00	19,00	00:00:08	00:00:00	00:00:08
MI	32	76,09	23,91	00:00:08	00:00:00	00:00:08	MI	32	83,33	16,67	00:00:11	00:00:00	00:00:11
MI	64	87,62	12,38	00:01:23	00:00:00	00:01:23	MI	64	80,46	19,54	00:00:07	00:00:00	00:00:07
MI	128	89,98	10,02	00:02:57	00:00:00	00:02:57	MI	128	85,03	14,97	00:00:18	00:00:00	00:00:18
MI	256	89,88	10,12	00:02:51	00:00:00	00:02:51	MI	256	89,46	10,54	00:01:17	00:00:00	00:01:17
MI	512	90,74	9,26	00:08:42	00:00:00	00:08:42	MI	512	79,82	20,18	00:00:45	00:00:00	00:00:45
MI	1024	79,36	20,64	00:01:52	00:00:00	00:01:52	MI	1024	91,25	8,75	00:09:29	00:00:00	00:09:29



# Resultados

## MLP Backprop (Otávio/Isaac)

UNIFEI_MLP_100.log.min							UNIFEI_MLP_50.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total	Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	74,59	25,41	00:00:02	00:00:00	00:00:02	CHI2	8	74,56	25,44	00:00:01	00:00:00	00:00:01
CHI2	16	75,73	24,27	00:00:03	00:00:00	00:00:03	CHI2	16	75,81	24,19	00:00:02	00:00:00	00:00:02
CHI2	32	74,28	25,72	00:00:04	00:00:00	00:00:04	CHI2	32	74,15	25,85	00:00:02	00:00:00	00:00:02
CHI2	64	79,14	20,86	00:00:10	00:00:00	00:00:10	CHI2	64	78,62	21,38	00:00:06	00:00:00	00:00:06
CHI2	128	62,31	37,69	00:00:19	00:00:00	00:00:19	CHI2	128	61,74	38,26	00:00:12	00:00:00	00:00:12
CHI2	256	79,53	20,47	00:00:41	00:00:01	00:00:42	CHI2	256	78,45	21,55	00:00:23	00:00:00	00:00:23
CHI2	512	83,14	16,86	00:03:02	00:00:00	00:03:02	CHI2	512	85,09	14,91	00:03:18	00:00:00	00:03:18
CHI2	1024	83,90	16,10	00:06:09	00:00:01	00:06:10	CHI2	1024	78,30	21,70	00:03:03	00:00:01	00:03:04
FD	8	70,91	29,09	00:01:20	00:00:00	00:01:20	FD	8	70,91	29,09	00:00:48	00:00:00	00:00:48
FD	16	71,59	28,41	00:03:42	00:00:00	00:03:42	FD	16	70,16	29,84	00:00:08	00:00:00	00:00:08
FD	32	74,43	25,57	00:03:12	00:00:00	00:03:12	FD	32	73,07	26,93	00:00:36	00:00:00	00:00:36
FD	64	71,68	28,32	00:00:20	00:00:00	00:00:20	FD	64	73,76	26,24	00:00:54	00:00:00	00:00:54
FD	128	70,94	29,06	00:00:31	00:00:00	00:00:31	FD	128	79,09	20,91	00:01:41	00:00:00	00:01:41
FD	256	82,67	17,33	00:07:15	00:00:01	00:07:16	FD	256	78,81	21,19	00:00:49	00:00:00	00:00:49
FD	512	76,50	23,50	00:02:21	00:00:01	00:02:22	FD	512	78,11	21,89	00:02:25	00:00:01	00:02:26
FD	1024	82,94	17,06	00:28:29	00:00:01	00:28:30	FD	1024	88,55	11,45	00:39:32	00:00:01	00:39:33
MI	8	25,84	74,16	00:00:13	00:00:00	00:00:13	MI	8	25,86	74,14	00:00:07	00:00:00	00:00:07
MI	16	26,85	73,15	00:00:47	00:00:00	00:00:47	MI	16	26,80	73,20	00:00:08	00:00:00	00:00:08
MI	32	74,84	25,16	00:09:17	00:00:00	00:09:17	MI	32	75,20	24,80	00:05:25	00:00:00	00:05:25
MI	64	75,19	24,81	00:01:26	00:00:00	00:01:26	MI	64	77,18	22,82	00:01:06	00:00:00	00:01:06
MI	128	79,86	20,14	00:04:19	00:00:00	00:04:19	MI	128	75,37	24,63	00:00:39	00:00:00	00:00:39
MI	256	81,91	18,09	00:13:22	00:00:00	00:13:22	MI	256	79,06	20,94	00:02:01	00:00:00	00:02:01
MI	512	77,08	22,92	00:06:06	00:00:00	00:06:06	MI	512	75,06	24,94	00:02:18	00:00:00	00:02:18
MI	1024	82,25	17,75	00:29:12	00:00:00	00:29:12	MI	1024	81,48	18,52	00:09:08	00:00:00	00:09:08

# Resultados MLP Backprop (Encog)

SpamAssassin_MLP_BPROP_100.log.min							SpamAssassin_MLP_BPROP_50.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total	Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	77,98	22,02	00:00:06	00:00:00	00:00:06	CHI2	8	63,17	36,83	00:00:02	00:00:00	00:00:02
CHI2	16	76,94	23,06	00:00:56	00:00:00	00:00:56	CHI2	16	67,73	32,27	00:00:24	00:00:00	00:00:24
CHI2	32	80,08	19,92	00:10:44	00:00:00	00:10:44	CHI2	32	64,37	35,63	00:00:09	00:00:00	00:00:09
CHI2	64	78,15	21,85	00:00:47	00:00:00	00:00:47	CHI2	64	50,28	49,72	00:00:44	00:00:00	00:00:44
CHI2	128	81,04	18,96	00:01:13	00:00:00	00:01:13	CHI2	128	52,18	47,82	00:00:07	00:00:00	00:00:07
CHI2	256	83,68	16,32	00:00:30	00:00:00	00:00:30	CHI2	256	57,29	42,71	00:01:17	00:00:00	00:01:17
CHI2	512	81,02	18,98	00:03:06	00:00:00	00:03:06	CHI2	512	52,81	47,19	00:00:21	00:00:00	00:00:21
CHI2	1024	84,32	15,68	00:02:48	00:00:00	00:02:48	CHI2	1024	51,54	48,46	00:03:38	00:00:00	00:03:38
FD	8	77,76	22,24	00:00:30	00:00:00	00:00:30	FD	8	69,38	30,62	00:00:32	00:00:00	00:00:32
FD	16	70,59	29,41	00:00:29	00:00:00	00:00:29	FD	16	68,93	31,07	00:00:03	00:00:00	00:00:03
FD	32	83,86	16,14	00:00:12	00:00:00	00:00:12	FD	32	52,03	47,97	00:00:16	00:00:00	00:00:16
FD	64	83,94	16,06	00:00:37	00:00:00	00:00:37	FD	64	52,46	47,54	00:00:08	00:00:00	00:00:08
FD	128	83,30	16,70	00:00:21	00:00:00	00:00:21	FD	128	60,88	39,12	00:00:06	00:00:00	00:00:06
FD	256	86,80	13,20	00:00:27	00:00:00	00:00:27	FD	256	68,87	31,13	00:00:08	00:00:00	00:00:08
FD	512	83,28	16,72	00:01:50	00:00:00	00:01:50	FD	512	53,09	46,91	00:00:50	00:00:00	00:00:50
FD	1024	85,48	14,52	00:02:21	00:00:00	00:02:21	FD	1024	65,04	34,96	00:03:05	00:00:00	00:03:05
MI	8	79,17	20,83	00:00:26	00:00:00	00:00:26	MI	8	69,08	30,92	00:00:04	00:00:00	00:00:04
MI	16	82,75	17,25	00:00:11	00:00:00	00:00:11	MI	16	54,86	45,14	00:00:03	00:00:00	00:00:03
MI	32	84,50	15,50	00:00:14	00:00:00	00:00:14	MI	32	68,90	31,10	00:00:09	00:00:00	00:00:09
MI	64	83,17	16,83	00:00:11	00:00:00	00:00:11	MI	64	55,33	44,67	00:00:02	00:00:00	00:00:02
MI	128	79,67	20,33	00:00:08	00:00:00	00:00:08	MI	128	68,87	31,13	00:00:05	00:00:00	00:00:05
MI	256	83,86	16,14	00:00:11	00:00:00	00:00:11	MI	256	65,53	34,47	00:00:23	00:00:00	00:00:23
MI	512	82,87	17,13	00:00:47	00:00:00	00:00:47	MI	512	51,96	48,04	00:00:13	00:00:00	00:00:13
MI	1024	79,92	20,08	00:02:12	00:00:00	00:02:12	MI	1024	53,48	46,52	00:01:24	00:00:00	00:01:24

# Resultados MLP Backprop (Encog)

UNIFEI_MLP_BPROP_100.log.min							UNIFEI_MLP_BPROP_50.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total	Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	71,84	28,16	00:00:09	00:00:00	00:00:09	CHI2	8	74,94	25,06	00:00:03	00:00:00	00:00:03
CHI2	16	75,29	24,71	00:00:18	00:00:00	00:00:18	CHI2	16	56,43	43,57	00:00:10	00:00:00	00:00:10
CHI2	32	75,37	24,63	00:00:29	00:00:00	00:00:29	CHI2	32	77,46	22,54	00:00:20	00:00:00	00:00:20
CHI2	64	79,42	20,58	00:00:44	00:00:00	00:00:44	CHI2	64	80,00	20,00	00:00:24	00:00:00	00:00:24
CHI2	128	81,00	19,00	00:02:23	00:00:00	00:02:23	CHI2	128	84,07	15,93	00:01:14	00:00:00	00:01:14
CHI2	256	76,06	23,94	00:02:31	00:00:00	00:02:31	CHI2	256	84,05	15,95	00:02:05	00:00:00	00:02:05
CHI2	512	83,85	16,15	00:02:58	00:00:00	00:02:58	CHI2	512	85,35	14,65	00:01:14	00:00:01	00:01:15
CHI2	1024	79,87	20,13	00:14:47	00:00:00	00:14:47	CHI2	1024	84,20	15,80	00:11:54	00:00:00	00:11:54
FD	8	78,10	21,90	00:01:03	00:00:00	00:01:03	FD	8	76,29	23,71	00:02:51	00:00:00	00:02:51
FD	16	64,00	36,00	00:04:06	00:00:00	00:04:06	FD	16	76,49	23,51	00:02:21	00:00:00	00:02:21
FD	32	62,91	37,09	00:05:16	00:00:00	00:05:16	FD	32	77,39	22,61	00:02:31	00:00:00	00:02:31
FD	64	78,08	21,92	00:01:39	00:00:00	00:01:39	FD	64	83,37	16,63	00:01:50	00:00:00	00:01:50
FD	128	78,10	21,90	00:02:46	00:00:00	00:02:46	FD	128	85,52	14,48	00:01:08	00:00:00	00:01:08
FD	256	78,09	21,91	00:04:06	00:00:00	00:04:06	FD	256	85,99	14,01	00:03:36	00:00:00	00:03:36
FD	512	80,25	19,75	00:03:17	00:00:00	00:03:17	FD	512	84,29	15,71	00:01:44	00:00:00	00:01:44
FD	1024	78,31	21,69	00:03:36	00:00:00	00:03:36	FD	1024	81,47	18,53	00:12:26	00:00:00	00:12:26
MI	8	78,04	21,96	00:17:40	00:00:00	00:17:40	MI	8	86,33	13,67	01:07:09	00:00:00	01:07:09
MI	16	78,04	21,96	01:42:56	00:00:00	01:42:56	MI	16	86,41	13,59	00:00:28	00:00:00	00:00:28
MI	32	79,68	20,32	00:04:15	00:00:00	00:04:15	MI	32	85,51	14,49	00:03:12	00:00:00	00:03:12
MI	64	77,90	22,10	00:05:18	00:00:00	00:05:18	MI	64	85,50	14,50	00:00:38	00:00:00	00:00:38
MI	128	67,34	32,66	00:03:11	00:00:00	00:03:11	MI	128	14,49	85,51	00:03:58	00:00:00	00:03:58
MI	256	79,18	20,82	00:00:42	00:00:01	00:00:43	MI	256	79,39	20,61	00:01:56	00:00:00	00:01:56
MI	512	70,69	29,31	00:01:06	00:00:00	00:01:06	MI	512	84,68	15,32	00:04:53	00:00:00	00:04:53
MI	1024	72,35	27,65	00:02:43	00:00:01	00:02:44	MI	1024	83,97	16,03	00:02:40	00:00:00	00:02:40

# Resultados MLP Rprop (Encog)

SpamAssassin_MLP_RPROP_100.log.min								SpamAssassin_MLP_RPROP_50.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total		Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	70,95	29,05	00:00:01	00:00:00	00:00:01		CHI2	8	57,09	42,91	00:00:00	00:00:00	00:00:00
CHI2	16	68,88	31,12	00:00:28	00:00:00	00:00:28		CHI2	16	57,84	42,16	00:00:04	00:00:00	00:00:04
CHI2	32	73,21	26,79	00:00:25	00:00:00	00:00:25		CHI2	32	52,96	47,04	00:00:10	00:00:00	00:00:10
CHI2	64	66,30	33,70	00:01:26	00:00:00	00:01:26		CHI2	64	63,41	36,59	00:00:06	00:00:00	00:00:06
CHI2	128	55,04	44,96	00:01:08	00:00:00	00:01:08		CHI2	128	28,75	71,25	00:00:01	00:00:00	00:00:01
CHI2	256	86,27	13,73	00:00:02	00:00:00	00:00:02		CHI2	256	83,00	17,00	00:00:04	00:00:00	00:00:04
CHI2	512	25,37	74,63	00:02:51	00:00:00	00:02:51		CHI2	512	84,79	15,21	00:00:01	00:00:00	00:00:01
CHI2	1024	89,51	10,49	00:00:16	00:00:00	00:00:16		CHI2	1024	54,33	45,67	00:00:02	00:00:00	00:00:02
FD	8	69,55	30,45	00:00:06	00:00:00	00:00:06		FD	8	49,34	50,66	00:00:03	00:00:00	00:00:03
FD	16	69,87	30,13	00:00:04	00:00:00	00:00:04		FD	16	78,57	21,43	00:00:01	00:00:00	00:00:01
FD	32	68,71	31,29	00:00:26	00:00:00	00:00:26		FD	32	72,93	27,07	00:00:00	00:00:00	00:00:00
FD	64	59,83	40,17	00:00:20	00:00:00	00:00:20		FD	64	63,62	36,38	00:00:00	00:00:00	00:00:00
FD	128	87,81	12,19	00:00:03	00:00:00	00:00:03		FD	128	55,40	44,60	00:00:01	00:00:00	00:00:01
FD	256	86,80	13,20	00:00:02	00:00:00	00:00:02		FD	256	89,60	10,40	00:00:01	00:00:00	00:00:01
FD	512	93,65	6,35	00:00:08	00:00:00	00:00:08		FD	512	94,26	5,74	00:00:01	00:00:00	00:00:01
FD	1024	92,81	7,19	00:00:14	00:00:00	00:00:14		FD	1024	94,59	5,41	00:00:06	00:00:00	00:00:06
MI	8	59,28	40,72	00:00:06	00:00:00	00:00:06		MI	8	73,55	26,45	00:00:00	00:00:00	00:00:00
MI	16	87,68	12,32	00:00:02	00:00:00	00:00:02		MI	16	85,08	14,92	00:00:01	00:00:00	00:00:01
MI	32	88,60	11,40	00:00:01	00:00:00	00:00:01		MI	32	55,33	44,67	00:00:00	00:00:00	00:00:00
MI	64	89,32	10,68	00:00:00	00:00:00	00:00:00		MI	64	55,23	44,77	00:00:00	00:00:00	00:00:00
MI	128	91,82	8,18	00:00:01	00:00:00	00:00:01		MI	128	50,41	49,59	00:00:01	00:00:00	00:00:01
MI	256	93,58	6,42	00:00:02	00:00:00	00:00:02		MI	256	95,25	4,75	00:00:01	00:00:00	00:00:01
MI	512	92,33	7,67	00:00:08	00:00:00	00:00:08		MI	512	93,17	6,83	00:00:01	00:00:00	00:00:01
MI	1024	89,88	10,12	00:00:06	00:00:00	00:00:06		MI	1024	91,78	8,22	00:00:04	00:00:00	00:00:04

# Resultados MLP Rprop (Encog)

UNIFEI_MLP_RPROP_100.log.min							UNIFEI_MLP_RPROP_50.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total	Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	0,00	100,00	00:00:01	00:00:00	00:00:01	CHI2	8	17,61	82,39	00:00:01	00:00:00	00:00:01
CHI2	16	29,19	70,81	00:00:11	00:00:00	00:00:11	CHI2	16	68,90	31,10	00:00:05	00:00:00	00:00:05
CHI2	32	43,35	56,65	00:00:17	00:00:00	00:00:17	CHI2	32	69,34	30,66	00:00:16	00:00:00	00:00:16
CHI2	64	76,32	23,68	00:00:08	00:00:00	00:00:08	CHI2	64	73,08	26,92	00:00:23	00:00:00	00:00:23
CHI2	128	48,85	51,15	00:03:16	00:00:00	00:03:16	CHI2	128	73,25	26,75	00:01:06	00:00:00	00:01:06
CHI2	256	82,88	17,12	00:00:02	00:00:00	00:00:02	CHI2	256	28,41	71,59	00:02:27	00:00:00	00:02:27
CHI2	512	3,06	96,94	00:09:11	00:00:00	00:09:11	CHI2	512	82,06	17,94	00:00:03	00:00:00	00:00:03
CHI2	1024	0,47	99,53	00:10:20	00:00:01	00:10:21	CHI2	1024	82,43	17,57	00:00:06	00:00:00	00:00:06
FD	8	77,45	22,55	00:00:12	00:00:00	00:00:12	FD	8	85,97	14,03	00:00:11	00:00:00	00:00:11
FD	16	60,62	39,38	00:00:19	00:00:00	00:00:19	FD	16	85,30	14,70	00:00:33	00:00:00	00:00:33
FD	32	52,48	47,52	00:00:21	00:00:00	00:00:21	FD	32	55,65	44,35	00:00:22	00:00:00	00:00:22
FD	64	30,13	69,87	00:05:18	00:00:00	00:05:18	FD	64	14,43	85,57	00:00:41	00:00:00	00:00:41
FD	128	84,67	15,33	00:00:03	00:00:00	00:00:03	FD	128	92,78	7,22	00:00:12	00:00:00	00:00:12
FD	256	89,79	10,21	00:00:17	00:00:00	00:00:17	FD	256	89,36	10,64	00:00:02	00:00:00	00:00:02
FD	512	89,52	10,48	00:00:16	00:00:00	00:00:16	FD	512	86,06	13,94	00:00:04	00:00:00	00:00:04
FD	1024	78,13	21,87	00:00:11	00:00:01	00:00:12	FD	1024	82,13	17,87	00:00:10	00:00:00	00:00:10
MI	8	78,04	21,96	00:00:23	00:00:00	00:00:23	MI	8	85,95	14,05	00:00:09	00:00:00	00:00:09
MI	16	78,05	21,95	00:00:13	00:00:00	00:00:13	MI	16	85,54	14,46	00:00:27	00:00:00	00:00:27
MI	32	89,35	10,65	00:00:09	00:00:00	00:00:09	MI	32	87,48	12,52	00:00:36	00:00:00	00:00:36
MI	64	86,48	13,52	00:00:07	00:00:00	00:00:07	MI	64	69,07	30,93	00:00:02	00:00:00	00:00:02
MI	128	94,01	5,99	00:00:05	00:00:00	00:00:05	MI	128	77,88	22,12	00:00:01	00:00:00	00:00:01
MI	256	83,13	16,87	00:00:04	00:00:00	00:00:04	MI	256	79,02	20,98	00:00:04	00:00:00	00:00:04
MI	512	88,46	11,54	00:00:20	00:00:01	00:00:21	MI	512	85,82	14,18	00:00:04	00:00:01	00:00:05
MI	1024	76,41	23,59	00:00:12	00:00:00	00:00:12	MI	1024	89,84	10,16	00:00:10	00:00:01	00:00:11

# Resultados – SVM (LibSVM)

SpamAssassin_SVM_100.log.min							SpamAssassin_SVM_50.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total	Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	93,92	6,08	00:00:03	00:00:03	00:00:06	CHI2	8	78,58	21,42	00:00:01	00:00:01	00:00:02
CHI2	16	95,60	4,40	00:00:03	00:00:03	00:00:06	CHI2	16	78,32	21,68	00:00:01	00:00:00	00:00:01
CHI2	32	95,89	4,11	00:00:06	00:00:06	00:00:12	CHI2	32	82,01	17,99	00:00:02	00:00:01	00:00:03
CHI2	64	96,79	3,21	00:00:10	00:00:16	00:00:26	CHI2	64	81,03	18,97	00:00:02	00:00:03	00:00:05
CHI2	128	97,68	2,32	00:00:17	00:00:28	00:00:45	CHI2	128	83,15	16,85	00:00:03	00:00:06	00:00:09
CHI2	256	99,29	0,71	00:00:28	00:00:47	00:01:15	CHI2	256	82,76	17,24	00:00:07	00:00:11	00:00:18
CHI2	512	98,54	1,46	00:00:49	00:01:23	00:02:12	CHI2	512	81,34	18,66	00:00:13	00:00:21	00:00:34
CHI2	1024	97,05	2,95	00:01:31	00:02:44	00:04:15	CHI2	1024	81,94	18,06	00:00:20	00:00:38	00:00:58
FD	8	97,55	2,45	00:00:04	00:00:03	00:00:07	FD	8	79,17	20,83	00:00:01	00:00:01	00:00:02
FD	16	97,15	2,85	00:00:04	00:00:05	00:00:09	FD	16	75,37	24,63	00:00:01	00:00:01	00:00:02
FD	32	95,13	4,87	00:00:06	00:00:07	00:00:13	FD	32	77,02	22,98	00:00:01	00:00:02	00:00:03
FD	64	92,23	7,77	00:00:12	00:00:15	00:00:27	FD	64	77,72	22,28	00:00:03	00:00:05	00:00:08
FD	128	88,47	11,53	00:00:17	00:00:27	00:00:44	FD	128	80,55	19,45	00:00:04	00:00:07	00:00:11
FD	256	84,80	15,20	00:00:27	00:00:47	00:01:14	FD	256	77,58	22,42	00:00:07	00:00:13	00:00:20
FD	512	81,35	18,65	00:00:44	00:01:21	00:02:05	FD	512	80,79	19,21	00:00:12	00:00:22	00:00:34
FD	1024	78,23	21,77	00:01:25	00:02:42	00:04:07	FD	1024	80,49	19,51	00:00:21	00:00:39	00:01:00
MI	8	94,29	5,71	00:00:04	00:00:02	00:00:06	MI	8	75,34	24,66	00:00:00	00:00:01	00:00:01
MI	16	93,60	6,40	00:00:04	00:00:04	00:00:08	MI	16	75,94	24,06	00:00:01	00:00:02	00:00:03
MI	32	90,94	9,06	00:00:06	00:00:07	00:00:13	MI	32	75,27	24,73	00:00:01	00:00:02	00:00:03
MI	64	90,97	9,03	00:00:11	00:00:15	00:00:26	MI	64	77,12	22,88	00:00:03	00:00:05	00:00:08
MI	128	87,18	12,82	00:00:16	00:00:25	00:00:41	MI	128	79,96	20,04	00:00:04	00:00:08	00:00:12
MI	256	83,80	16,20	00:00:26	00:00:45	00:01:11	MI	256	78,64	21,36	00:00:07	00:00:11	00:00:18
MI	512	80,64	19,36	00:00:42	00:01:27	00:02:09	MI	512	79,63	20,37	00:00:12	00:00:21	00:00:33
MI	1024	77,90	22,10	00:01:23	00:02:48	00:04:11	MI	1024	79,83	20,17	00:00:22	00:00:40	00:01:02

# Resultados – SVM (LibSVM)

UNIFEI_SVM_100.log.min							UNIFEI_SVM_50.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total	Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	92,34	7,66	00:00:02	00:00:01	00:00:03	CHI2	8	52,56	47,44	00:00:00	00:00:01	00:00:01
CHI2	16	93,13	6,87	00:00:02	00:00:02	00:00:04	CHI2	16	50,12	49,88	00:00:00	00:00:01	00:00:01
CHI2	32	95,73	4,27	00:00:05	00:00:04	00:00:09	CHI2	32	50,81	49,19	00:00:01	00:00:01	00:00:02
CHI2	64	98,88	1,12	00:00:26	00:00:32	00:00:58	CHI2	64	62,07	37,93	00:00:06	00:00:07	00:00:13
CHI2	128	99,63	0,37	00:00:59	00:01:26	00:02:25	CHI2	128	65,88	34,12	00:00:15	00:00:22	00:00:37
CHI2	256	99,91	0,09	00:05:32	00:03:53	00:09:25	CHI2	256	66,45	33,55	00:00:40	00:00:58	00:01:38
CHI2	512	99,93	0,07	00:15:31	00:08:31	00:24:02	CHI2	512	66,9	33,1	00:01:22	00:02:15	00:03:37
CHI2	1024	99,96	0,04	00:31:52	00:15:36	00:47:28	CHI2	1024	66,62	33,38	00:02:39	00:04:29	00:07:08
FD	8	99,02	0,98	00:01:41	00:00:30	00:02:11	FD	8	66,97	33,03	00:00:16	00:00:07	00:00:23
FD	16	98,38	1,62	00:02:10	00:00:45	00:02:55	FD	16	66,84	33,16	00:00:16	00:00:09	00:00:25
FD	32	98,27	1,73	00:02:56	00:01:11	00:04:07	FD	32	67,23	32,77	00:00:24	00:00:18	00:00:42
FD	64	98,02	1,98	00:04:28	00:02:12	00:06:40	FD	64	67,15	32,85	00:00:44	00:00:39	00:01:23
FD	128	97,12	2,88	00:06:34	00:04:01	00:10:35	FD	128	66,76	33,24	00:00:54	00:01:17	00:02:11
FD	256	95,47	4,53	00:10:47	00:07:17	00:18:04	FD	256	66,3	33,7	00:01:31	00:02:04	00:03:35
FD	512	92,06	7,94	00:18:50	00:13:14	00:32:04	FD	512	66,7	33,3	00:02:35	00:03:54	00:06:29
FD	1024	88,52	11,48	00:35:28	00:26:15	01:01:43	FD	1024	66,13	33,87	00:04:35	00:07:41	00:12:16
MI	8	86,25	13,75	00:01:12	00:00:29	00:01:41	MI	8	66,3	33,7	00:00:15	00:00:07	00:00:22
MI	16	86,35	13,65	00:01:32	00:00:46	00:02:18	MI	16	66,87	33,13	00:00:13	00:00:10	00:00:23
MI	32	95,23	4,77	00:03:58	00:01:32	00:05:30	MI	32	66,41	33,59	00:00:37	00:00:20	00:00:57
MI	64	97,41	2,59	00:05:00	00:02:29	00:07:29	MI	64	66,33	33,67	00:00:43	00:00:42	00:01:25
MI	128	96,24	3,76	00:07:15	00:04:19	00:11:34	MI	128	66,81	33,19	00:01:01	00:01:22	00:02:23
MI	256	93,72	6,28	00:11:18	00:07:35	00:18:53	MI	256	65,98	34,02	00:01:30	00:02:14	00:03:44
MI	512	90,18	9,82	00:19:13	00:13:49	00:33:02	MI	512	66,44	33,56	00:02:32	00:03:38	00:06:10
MI	1024	85,53	14,47	00:35:18	00:27:31	01:02:49	MI	1024	66,56	33,44	00:04:51	00:07:48	00:12:39

# Resultados - RBF (Encog)

SpamAssassin_RBF_50.log.min							SpamAssassin_RBF_100.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total	Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	80,30	19,70	00:00:01	00:00:00	00:00:01	CHI2	8	67,90	32,10	00:00:00	00:00:00	00:00:00
CHI2	16	80,26	19,74	00:00:06	00:00:00	00:00:06	CHI2	16	66,36	33,64	00:00:02	00:00:00	00:00:02
CHI2	32	80,29	19,71	00:03:25	00:00:00	00:03:25	CHI2	32	68,24	31,76	00:01:27	00:00:00	00:01:27
CHI2	64	0,00	100,00	01:24:49	00:00:01	01:24:50	CHI2	64	0,00	100,00	00:39:51	00:00:00	00:39:51
FD	8	70,57	29,43	00:00:01	00:00:00	00:00:01	FD	8	55,13	44,87	00:00:00	00:00:00	00:00:00
FD	16	70,94	29,06	00:00:05	00:00:00	00:00:05	FD	16	54,67	45,33	00:00:02	00:00:00	00:00:02
FD	32	71,11	28,89	00:03:09	00:00:01	00:03:10	FD	32	55,46	44,54	00:01:38	00:00:00	00:01:38
FD	64	71,42	28,58	01:30:30	00:00:02	01:30:32	FD	64	55,23	44,77	00:43:47	00:00:01	00:43:48
MI	8	80,23	19,77	00:00:00	00:00:00	00:00:00	MI	8	69,65	30,35	00:00:00	00:00:00	00:00:00
MI	16	80,81	19,19	00:00:07	00:00:00	00:00:07	MI	16	66,61	33,39	00:00:03	00:00:00	00:00:03
MI	32	81,12	18,88	00:03:20	00:00:00	00:03:20	MI	32	61,44	38,56	00:01:36	00:00:01	00:01:37
MI	64	81,47	18,53	01:22:34	00:00:02	01:22:36	MI	64	0,00	100,00	00:43:40	00:00:01	00:43:41
UNIFEI_RBF_100.log.min							UNIFEI_RBF_50.log.min						
Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total	Tipo	# Carac.	% Acerto	% Erro	Treinamento	Classificação	Total
CHI2	8	76,19	23,81	00:00:01	00:00:00	00:00:01	CHI2	8	74,11	25,89	00:00:00	00:00:00	00:00:00
CHI2	16	77,38	22,62	00:00:03	00:00:00	00:00:03	CHI2	16	76,15	23,85	00:00:02	00:00:00	00:00:02
CHI2	32	0,00	100,00	00:02:43	00:00:00	00:02:43	CHI2	32	0,00	100,00	00:01:20	00:00:00	00:01:20
CHI2	64	0,00	100,00	02:25:39	00:00:03	02:25:42	CHI2	64	0,00	100,00	01:08:08	00:00:01	01:08:09
FD	8	78,11	21,89	00:00:02	00:00:00	00:00:02	FD	8	86,36	13,64	00:00:01	00:00:00	00:00:01
FD	16	78,12	21,88	00:00:27	00:00:00	00:00:27	FD	16	85,65	14,35	00:00:11	00:00:00	00:00:11
FD	32	78,17	21,83	00:15:06	00:00:01	00:15:07	FD	32	85,73	14,27	00:06:12	00:00:01	00:06:13
FD	64	78,18	21,82	05:11:15	00:00:05	05:11:20	FD	64	84,72	15,28	02:48:28	00:00:03	02:48:31
MI	8	78,04	21,96	00:00:02	00:00:00	00:00:02	MI	8	85,69	14,31	00:00:01	00:00:00	00:00:01
MI	16	78,05	21,95	00:00:29	00:00:00	00:00:29	MI	16	85,49	14,51	00:00:11	00:00:00	00:00:11
MI	32	78,10	21,90	00:18:35	00:00:01	00:18:36	MI	32	85,67	14,33	00:06:02	00:00:01	00:06:03
MI	64	78,18	21,82	06:22:37	00:00:05	06:22:42	MI	64	85,22	14,78	03:43:13	00:00:02	03:43:15



# Conclusões

- Os tempos de treinamento foram muito menores nas redes MLP do que na SVM e, principalmente, do que na rede RBF, especialmente em situações onde a quantidade de características é “grande” ( $> 64$ ).
- Os tempos de classificação mostraram-se aceitáveis para todas as técnicas.
- A técnica SVM foi a que apresentou melhor generalização de conhecimento.
- Nas redes MLP, o algoritmo RPROP mostrou maior velocidade de convergência, mas o BACKPROP obteve maior precisão na classificação.
- Os métodos de seleção de características apresentaram desempenhos variáveis, apesar de a proposta do *Mutual Information* (MI) soar mais promissora no contexto da representação de e-mails.

# Estudos Futuros

- Domínio das técnicas e realização de mais testes para otimização de seus parâmetros, como por exemplo:
  - Taxa de aprendizado, épocas de treinamento e número de neurônios em cada camada oculta do MLP.
  - Estabelecimento de centros e variâncias da rede RBF (possivelmente através de técnicas de inteligência artificial).
  - Taxas de penalização e precisão da SVM.
- Testes com outros conjuntos de dados.
- Implementação e teste do algoritmo RPROP no MLP desenvolvido pelo Otávio / Isaac.

# Referências Bibliográficas

- [1] Symantec Corporation, "Internet Security Threat Report 2014 :: Volume 19" (2014) p.79. Disponível em <[http://www.symantec.com/content/en/us/enterprise/other\\_resources/b-istr\\_appendices\\_v19\\_221284438.en-us.pdf](http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_appendices_v19_221284438.en-us.pdf)>.
- [2] Rumelhart, David E., James L. McClelland, and PDP Research Group. Parallel distributed processing. Vol. 1. IEEE, 1988.
- [3] Riedmiller, Martin, and Heinrich Braun. "A direct adaptive method for faster backpropagation learning: The RPROP algorithm." Neural Networks, 1993., IEEE International Conference on. IEEE, 1993.
- [4] Igel, Christian, and Michael Hüsken. "Improving the Rprop learning algorithm." Proceedings of the second international ICSC symposium on neural computation (NC 2000). Vol. 2000. ICSC Academic Press, 2000.
- [5] Mendez, Jose Ramon, et al. "A comparative performance study of feature selection methods for the anti-spam filtering domain." Advances in Data Mining. Applications in Medicine, Web Mining, Marketing, Image and Signal Mining. Springer Berlin Heidelberg, 2006. 106-120.

# Referências Bibliográficas

- [6] Apache SpamAssassin™ public mail corpus. Disponível em [<https://spamassassin.apache.org/publiccorpus/>](https://spamassassin.apache.org/publiccorpus/).
- [7] Heaton, Jeff. "Encog: Library of Interchangeable Machine Learning Models for Java and C#." Journal of Machine Learning Research 16 (2015): 1243-47. Print.
- [8] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: A library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST) 2.3 (2011): 27.
- [9] Haykin, Simon, and Neural Network. "A comprehensive foundation." Neural Networks 2.2004 (2004).
- [10] da Silva, Ivan Nunes, Danilo Hernane Spatti, and Rogério Andrade Flauzino. "Redes Neurais Artificiais para engenharia e ciências aplicadas curso prático." Artliber (2010).
- [11] Jordan, Michael. "Statistical Learning Theory: Course Lectures." University of California, Berkeley. Computer Science Division (2004). Disponível em [<http://www.cs.berkeley.edu/~jordan/courses/281B-spring04/lectures/lec3.pdf >](http://www.cs.berkeley.edu/~jordan/courses/281B-spring04/lectures/lec3.pdf).