
Trabalho de Redes Neurais

Lucas Abadde, Luiz Lima e Pedro Carletti

Breast-cancer

—

O conjunto de dados
reúne **9 características**
relacionadas à
ocorrência do câncer
de mama

University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia



1. Age

Idade da paciente no momento do diagnóstico.

Tipo: categórico

Valores: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99

2. Menopause

Situação de menopausa da paciente no momento do diagnóstico.

Tipo: categórico

Valores: lt40, ge40, premeno





3. Tumor-size

Tamanho do tumor, em mm.

Tipo: categórico

Valores: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59

4. Inv-nodes

Número de gânglios linfáticos axilares apresentando câncer de mama no momento do exame histológico.

Tipo: categórico

Valores: 0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, 27-29, 30-32, 33-35, 36-39





5. Node-caps

Penetração do tumor na cápsula do gânglio linfático.

Tipo: binário

Valores: yes, no

6. Deg-malig

Classificação histológica do tumor:

1. Mais tecido normal e crescimento lento;
2. Entre graus 1 e 3;
3. Muito diferente do tecido normal e crescimento rápido

Tipo: categórico

Valores: 1, 2, 3





7. Breast

Lado do seio diagnosticado com tumor.

Tipo: binário

Valores: left, right

8. Breast-quad

Quadrante do seio onde se localiza o tumor, considerando o mamilo como centro

Tipo: categórico

Valores: left-up, left-low, right-up, right-low, central





9. Irradiat

Se a paciente já passou por radioterapia.

Tipo: binário

Valores: yes, no

A classificação esperada é sobre a recorrência da doença

University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia

10. Class

Sintomas reaparecem na paciente após tratamento

Tipo: binário

Valores: no-recurrence-events,
recurrence-events



—

Existem **286** amostras,
sendo **201** não
recorrentes e **85**
recorrentes

Voted Perceptron

—

**Simples e eficiente para
classificação linear com
grandes margens**

Algoritmo de treino

Input: a labeled training set $\langle (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m) \rangle$
 number of epochs T

Output: a list of weighted perceptrons $\langle (\mathbf{v}_1, c_1), \dots, (\mathbf{v}_k, c_k) \rangle$

- Initialize: $k := 0, \mathbf{v}_1 := \mathbf{0}, c_1 := 0$.
- Repeat T times:
 - For $i = 1, \dots, m$:
 - * Compute prediction: $\hat{y} := \text{sign}(\mathbf{v}_k \cdot \mathbf{x}_i)$
 - * If $\hat{y} = y$ then $c_k := c_k + 1$.
 else $\mathbf{v}_{k+1} := \mathbf{v}_k + y_i \mathbf{x}_i$;
 $c_{k+1} := 1$;
 $k := k + 1$.

Algoritmo de teste

Given: the list of weighted perceptrons: $\langle (\mathbf{v}_1, c_1), \dots, (\mathbf{v}_k, c_k) \rangle$
 an unlabeled instance: \mathbf{x}
compute a predicted label \hat{y} as follows:

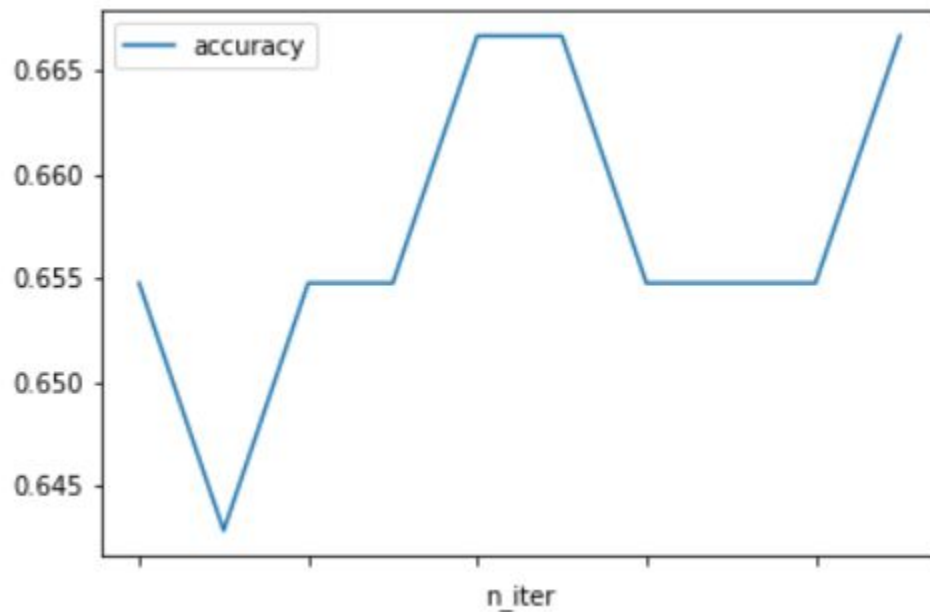
$$s = \sum_{i=1}^k c_i \operatorname{sign}(\mathbf{v}_i \cdot \mathbf{x}); \quad \hat{y} = \operatorname{sign}(s).$$

Resultados

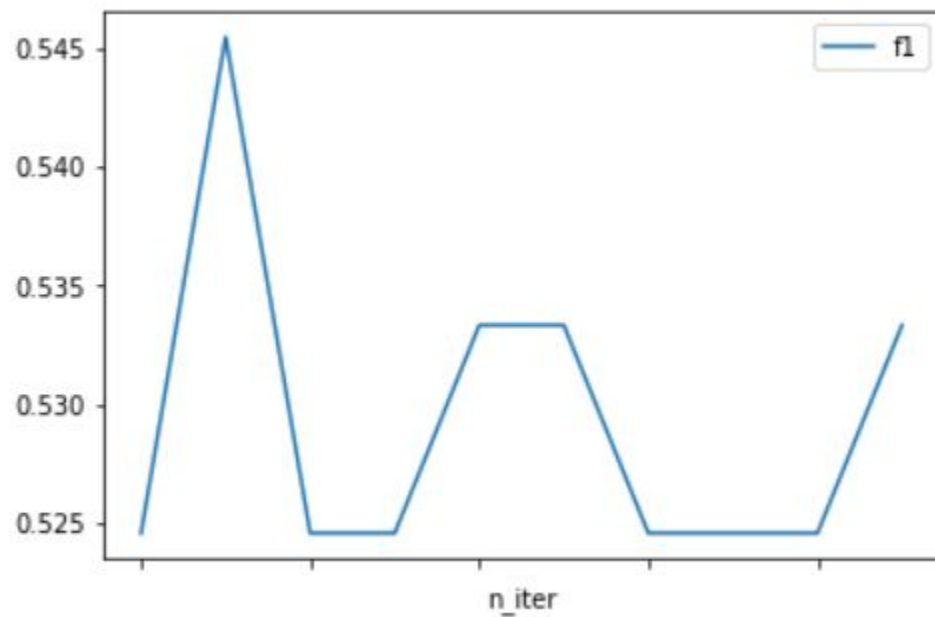
Variando o número de iterações

	n_iter	conf_mat	accuracy	precision	recall	f1
0	1	[[39, 20], [9, 16]]	0.654762	0.444444	0.64	0.524590
1	2	[[36, 23], [7, 18]]	0.642857	0.439024	0.72	0.545455
2	4	[[39, 20], [9, 16]]	0.654762	0.444444	0.64	0.524590
3	8	[[39, 20], [9, 16]]	0.654762	0.444444	0.64	0.524590
4	16	[[40, 19], [9, 16]]	0.666667	0.457143	0.64	0.533333
5	32	[[40, 19], [9, 16]]	0.666667	0.457143	0.64	0.533333
6	64	[[39, 20], [9, 16]]	0.654762	0.444444	0.64	0.524590
7	128	[[39, 20], [9, 16]]	0.654762	0.444444	0.64	0.524590
8	256	[[39, 20], [9, 16]]	0.654762	0.444444	0.64	0.524590
9	512	[[40, 19], [9, 16]]	0.666667	0.457143	0.64	0.533333

Número de iterações x Taxa de acerto



F1 score



Conclusões

Modelo e dataset

O dataset não é linearmente separável por grandes margens.

No Voted Perceptron mais iterações não representam mais qualidade*

—

Obrigado!