

# Human Activity Recognition on Smartphones using Multiclass Hardware-Friendly Support Vector Machine

# Importância

Monitoramento de atividades para idosos (healthcare)

# Trabalhos Relacionados

Naive Bayes

SVMs

Threshold-based

Markov chains

Matrix Back Propagation (aritmética de ponto-fixa)

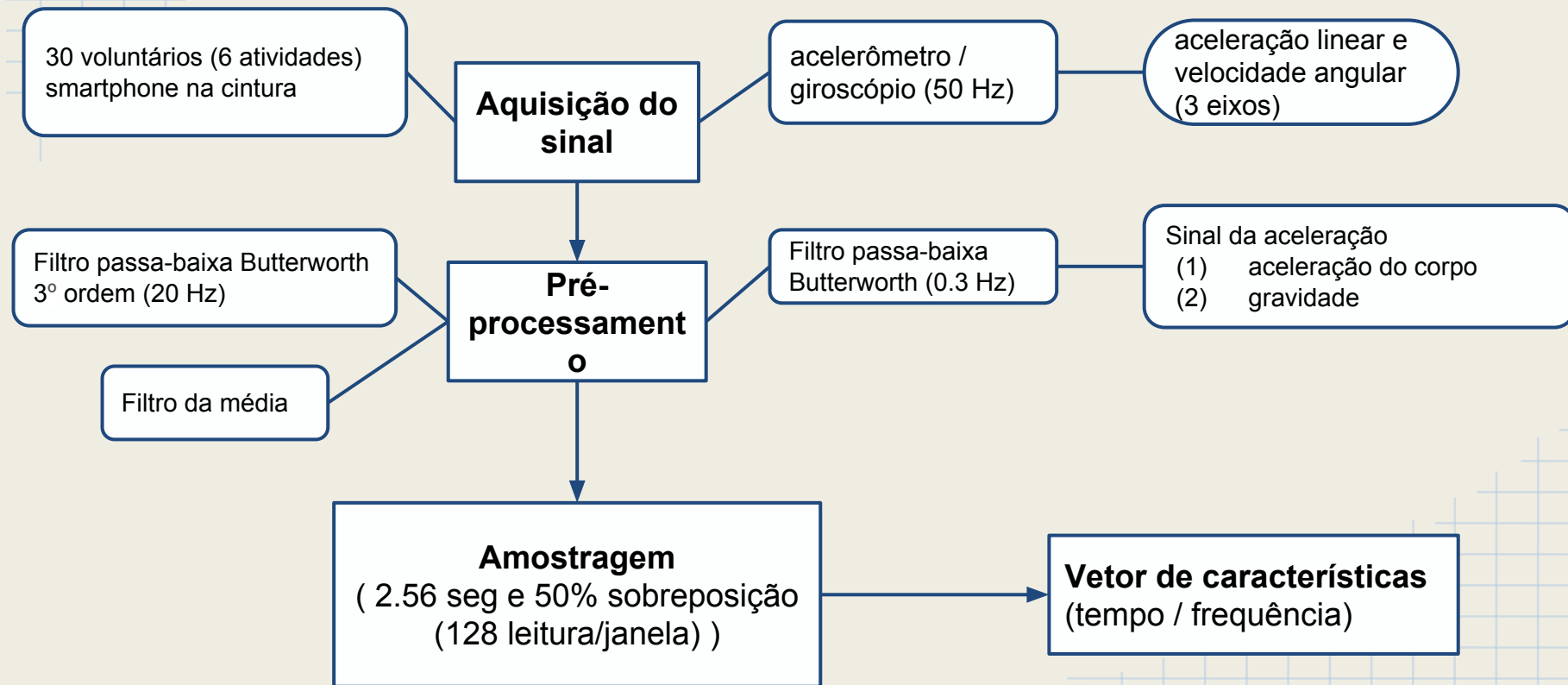
HF-SVM (explora aritmética de ponto-fixa na fase *feed-forward* do SVM)

One-vs-All (máximo)

One-vs-One

- $K(K - 1) / 2$  classificadores
- um classificador para cada par de classe
- a classe com maior número de +1

# Metodologia



# Aquisição do sinal

## Acelerômetro

Pode medir a orientação de um item estacionário em relação à superfície da Terra.

Ao acelerar em uma determinada direção, é incapaz de distinguir entre esta aceleração e a fornecida através da atração gravitacional da Terra.

Assim como, a aceleração ( $\text{m/s}^2$ ), inclinação, ângulo de inclinação, rotação, vibração, colisão e gravidade.

Pode **mudar automaticamente o visor do celular na vertical ou horizontal.**

Detecção de movimento (ao chacoalhar, ao bater etc.)

# Aquisição do sinal

## Giroscópio

Mede a velocidade de rotação em torno de um determinado eixo. Usando os princípios fundamentais do momento angular, o giroscópio ajuda a indicar a orientação.

Podem monitorar a orientação, direção, movimento angular e rotação (rad/s).

**Executa funções de reconhecimento de gestos** (no smartphone)

***Detecta a rotação do aparelho baseado em três eixos. Gira a tela de visualização.***

Detecção da rotação (giro, virada etc.)

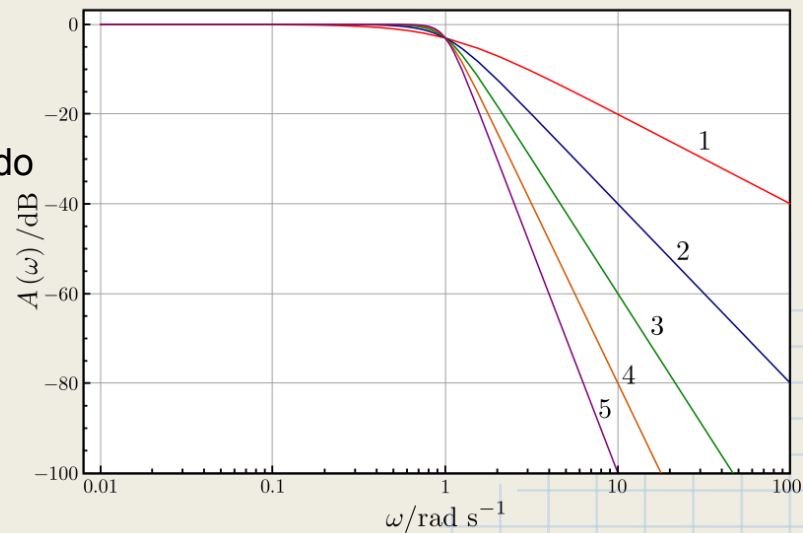
# Pré-processamento

## Filtro da Média (median filter)

Usa o mesmo princípio em PDI?

## 3rd order low pass Butterworth filter

É um tipo de projeto de filtros eletrônicos. Ele é desenvolvido de modo a ter uma resposta em frequência o mais plana o quanto for matematicamente possível na banda passante.



# Pré-processamento

## **Fast Fourier Transform**

É uma operação matemática que, quando aplicada a um sinal qualquer, que está no domínio do tempo, o transforma numa representação desse sinal no domínio das frequências, onde são representadas as frequências resultantes dessa transformação, suas respectivas fases e as amplitudes que compõem esse sinal.

**Jerk** (derivada da aceleração) ?



# Vetor de Características

mean(): Mean value

std(): Standard deviation

mad(): **Median absolute deviation**

max(): Largest value in array

min(): Smallest value in array

sma(): **Signal magnitude area**

energy(): **Energy measure. Sum of the squares divided by the number of values.**

iqr(): **Interquartile range**

entropy(): **Signal entropy**

arCoeff(): **Autoregression coefficients with Burg order equal to 4**

correlation(): **correlation coefficient between two signals**

maxInds(): index of the frequency component with largest magnitude

# Vetor de Características

meanFreq(): **Weighted average of the frequency components to obtain a mean frequency ???**

skewness(): **skewness of the frequency domain signal**

kurtosis(): **kurtosis of the frequency domain signal**

bandsEnergy(): **Energy of a frequency** interval within the 64 bins of the FFT of each window. ???

angle(): Angle between to vectors.

# Vetor de características

## Median absolute deviation

For a univariate data set  $X_1, X_2, \dots, X_n$ , the MAD is defined as the median of the absolute deviations from the data's median:

$$\text{MAD} = \text{median}_i ( |X_i - \text{median}_j(X_j)| ),$$

that is, starting with the residuals (deviations) from the data's median, the MAD is the median of their absolute values.

## Example

Consider the data (1, 1, 2, **2**, 4, 6, 9). It has a median value of 2. The absolute deviations about 2 are (1, 1, 0, 0, 2, 4, 7) which in turn have a median value of 1 (because the sorted absolute deviations are (0, 0, 1, **1**, 2, 4, 7)). So the median absolute deviation for this data is 1.

# Vetor de características

## Signal magnitude area

The SMA value of a set of values (or a continuous-time waveform) is the normalized integral of the original values.

In the case of a set of  $n$  values  $\{x_1, x_2, \dots, x_n\}$  matching a time length  $T$ , the SMA

$$x_{\text{sma}} = \sum_{i=1}^n x_i$$

In the continuous domain, we have for example, with a 3-axis signal with an offset correction  $a$  for each axis, the following equation:

$$f_{\text{sma}} = \frac{1}{T} \int_0^T |x(t) - a_x| + |y(t) - a_y| + |z(t) - a_z| dt$$

Somatório dos eixos das acelerações?

# Vetor de características

## Interquartile range

O intervalo interquartil é a distância entre o terceiro e o primeiro quartis

**primeiro quartil** (designado por  $Q1/4$ ) = quartil inferior = é o valor aos 25% da amostra ordenada = 25º percentil

**terceiro quartil** (designado por  $Q3/4$ ) = quartil superior = valor a partir do qual se encontram 25% dos valores mais elevados = valor aos 75% da amostra ordenada = 75º percentil

# Vetor de características

## Signal entropy

A entropia é uma forma de medir a quantidade de informação. Intuitivamente uma mensagem tem tanta mais informação quanto maior for o seu grau de aleatoriedade ou imprevisibilidade. Assim, por exemplo, a entropia ou quantidade de informação da variável aleatória  $x$  de probabilidade  $p_X(x)$ , é definida por

$$H(x) = E[-\log_2 p_X(x)] = - \sum_{x \in \Omega_x} p_X(x) \log_2 p_X(x).$$

# Vetor de características

## Autoregression coefficients with Burg order equal to 4???

An autoregressive model (AR) is also known in the filter design industry as an infinite impulse response filter (IIR) or an all pole filter, and is sometimes known as a maximum entropy model in physics applications. There is "memory" or feedback and therefore the system can generate internal dynamics.

The definition that will be used here is as follows

$$x_t = \sum_{i=1}^N a_i x_{t-i} + \varepsilon_t$$

where  $a_i$  are the autoregression coefficients,  $x_t$  is the series under investigation, and  $N$  is the order (length) of the filter which is generally very much less than the length of the series.

- In statistics and signal processing, an autoregressive (AR) model is a representation of a type of random process; as such, it describes certain time-varying processes in nature, economics, etc. The autoregressive model specifies that the output variable depends linearly on its own previous values. It is a special case of the more general ARMA model of time series.

# Vetor de características

Skewness de um sinal  $x$ , denotado por  $S_x$ , é o momento estatístico de terceira ordem, normalizado pelo desvio padrão elevado ao cubo:

$$S_x = \frac{1}{\sigma_x^3} \frac{1}{N} \sum_{n=1}^N (x_n - \mu_x)^3$$

onde  $x$  é o sinal medido,  $N$  número de amostras;  $(\mu)_x$  é a média de  $x$ , e  $(\sigma)_x$  seu desvio padrão. Skewness assemelha-se o valor médio no sentido que a terceira potência mantém números negativos. **Um valor skewness diferente de zero indica que a função de densidade de probabilidade do sinal não é simétrica em torno da média.** Normalmente para sinais dinâmicos, o skewness será zero. Uma diferença entre skewness e média é que valores grandes são enfatizados no skewness pela terceira potência. Por exemplo, skewness será mais sensível a picos de energia no sinal em apenas uma direção (voltagem positiva ou negativa).



# Vetor de características

## Kurtosis of the frequency domain signal

O kurtosis espectral é uma ferramenta estatística heurística para detectar a presença de transientes em um sinal e sua localização no domínio da frequência.

Transiente, em engenharia elétrica, é um surto de tensão elétrica que ocorre num intervalo de tempo muito pequeno.

Em Estatística descritiva, a curtose é uma medida de dispersão que caracteriza o pico ou "achatamento" da curva da função de distribuição de probabilidade 1 . É normalmente definida como:

$$\frac{m_4(\mu)}{\sigma^4} + (-3)$$

onde  $m_4(\mu)$  é o quarto Momento central e  $\sigma$  é o Desvio-padrão.

$$m_4(\mu)$$

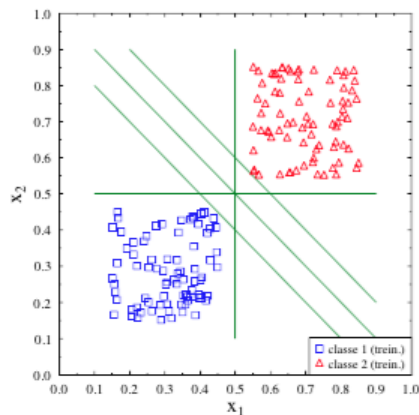
# Vetor de características

**Energy of a frequency interval within the 64 bins of the FFT of each window.**

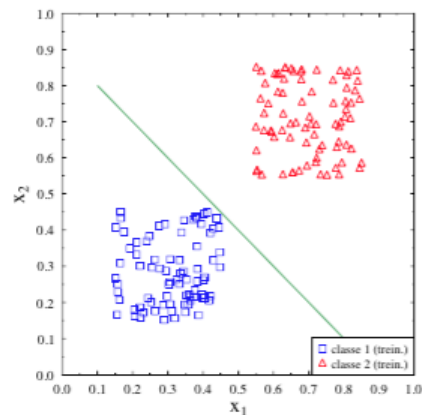
O tamanho da janela influencia a resolução temporal ou frequência, ou precisão da representação do sinal. A resolução da frequência pode ser aumentada mudando o tamanho do FFT, que é o número de bins da janela de análise.

O tamanho FFT define o número de bins usado para dividir a janela em faixas iguais, ou bins. Por isso, um bin é amostra de espectro, e define a resolução da frequência da janela.

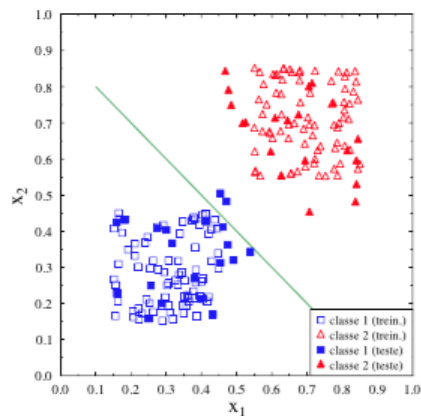
# SVM



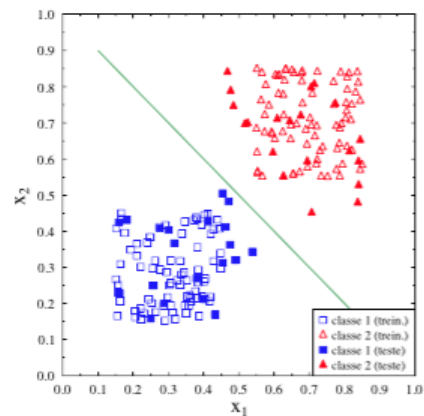
(a) Diversos hiperplanos capazes de resolver o problema linearmente separável. Nessa figura são mostrados apenas os padrões de treinamento.



(b) Problema resolvido por um hiperplano muito próximo a uma das classes.

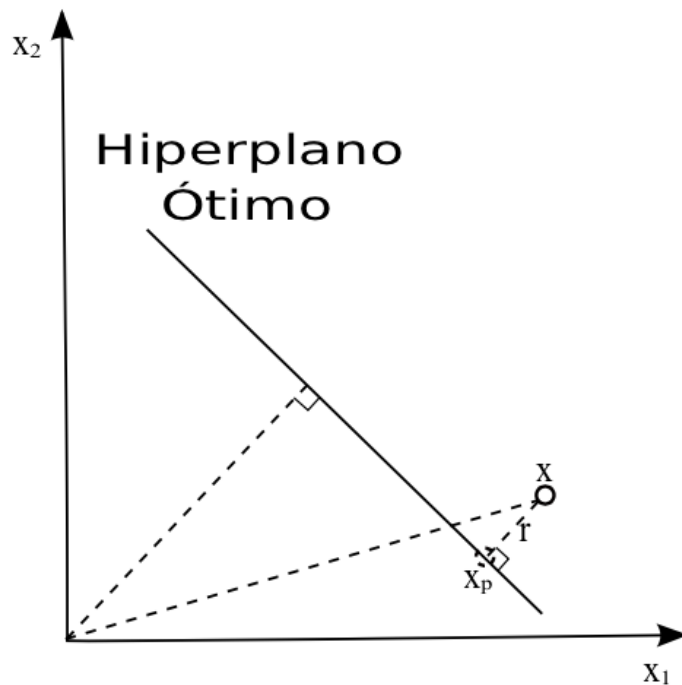


(c) O mesmo problema descrito em (a), porém aqui são mostrados os padrões de treinamento e teste. Observe que padrões do conjunto de teste apresentam-se do lado oposto ao da sua classe.



(d) Problema resolvido por um hiperplano mais adequado, pois apresenta-se equidistante das duas classes.

# SVM



**Figura 2.3:** Interpretação geométrica da distância de um padrão  $x$  ao hiperplano ótimo.

# SVM

Maximização da margem de separação  $\rho$  implica simultaneamente na minimização da dimensão VC

$$\max \rho = \max \frac{2}{\|\mathbf{w}\|}, \quad (2.14)$$

que ainda pode ser apresentado como

$$\min \|\mathbf{w}\| \iff \min \sqrt{\mathbf{w}^T \mathbf{w}} \iff \min \frac{1}{2} \mathbf{w}^T \mathbf{w}, \quad (2.15)$$

com base na minimização da norma do vetor de pesos  $\|\mathbf{w}\|$  ou de termos decorrentes.

A partir de agora deve-se considerar a função  $\tau(\mathbf{w})$  a ser minimizada, como sendo a seguinte:

$$\tau(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w}. \quad (2.16)$$

# SVM

Classificador SVM com Margem Rígida

Método de Lagrange

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i^o d_i \mathbf{x}_i^T \mathbf{x} + b_o.$$

Um hiperplano que separe sem erros todos os padrões pertencentes às duas classes, nem sempre existe.

# SVM

## Classificador SVM com Margem Flexível

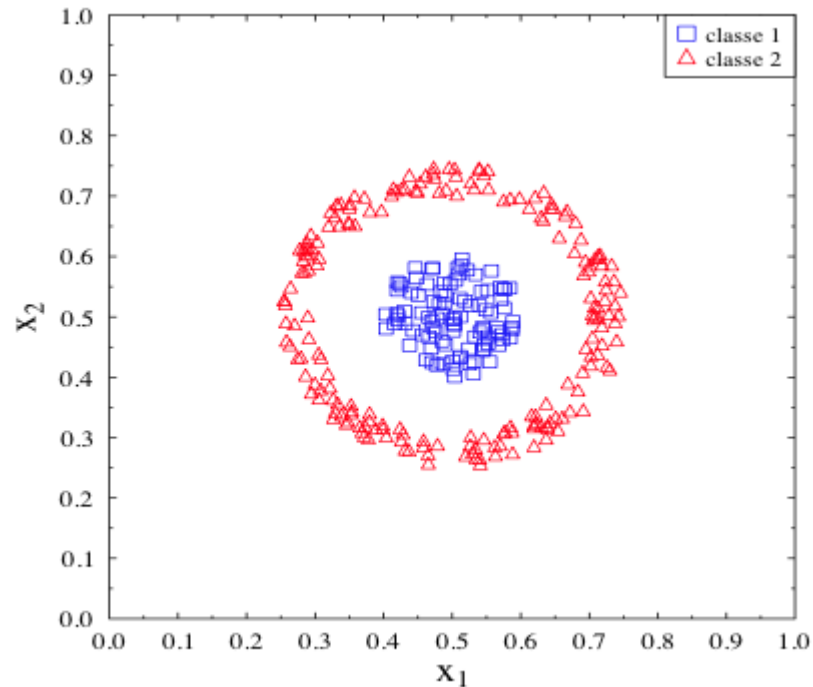
$$d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, \dots, n,$$

## Método de Lagrange

$$\begin{aligned} \max L(\alpha) = \max & \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j \right\} \\ & s.a. \sum_{i=1}^n \alpha_i d_i = 0 \\ & s.a. 0 \leq \alpha_i \leq C \text{ para } i = 1, \dots, n \end{aligned}$$

# SVM

O conjunto de dados deve ser transformado para um espaço de características de elevada dimensão.





# SVM

## Truque de Kernel

O truque de kernel permite que não seja necessário realizar o mapeamento para que se obtenha o vetor  $\phi(\mathbf{x})$ , desde que se conheça uma função que descreva o produto interno  $\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$

<i>Kernel</i>	Descrição
Linear	$K(\mathbf{x}, \mathbf{x}_i) = \mathbf{x}_i^T \cdot \mathbf{x}$
Polinomial	$K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}_i^T \cdot \mathbf{x} + 1)^d$ , em que $d$ é o grau do polinômio.
Gaussiano (RBF)	$K(\mathbf{x}, \mathbf{x}_i) = \exp \left\{ -\ \mathbf{x} - \mathbf{x}_i\ ^2 / \sigma^2 \right\}$ , em que $\sigma$ é uma constante.

# SVM

## Obtenção dos Parâmetros Ótimos para o Classificador SVM

Solução baseada em Programação Quadrática (usada no artigo)

$$\min L(\alpha) = \frac{1}{2} \alpha^T Q \alpha - \alpha^T \mathbf{1} \quad (2.64)$$

sujeito a

$$\alpha^T \mathbf{d} = 0 \quad (2.65)$$

$$0 \leq \alpha_i \leq C \quad i = 1 \dots n,$$

em que  $Q$  é uma matriz  $n \times n$ , sendo seus elementos descritos por  $q_{i,j} = d_i d_j K(\mathbf{x}_i, \mathbf{x}_j)$  tal que  $i, j = 1 \dots n$ . Em geral, a matriz  $Q$  é densa e semi-definida positiva, e caso apresente-se com um valor elevado para  $n$ , tal matriz pode ser grande demais para ser armazenada (KAUFMAN, 1999). Em virtude disto, um método *online* pode ser útil, pois este não exige uma grande quantidade de memória.

# SVM

alfa\_i pode ser usado para encontrar as classes na fase Feed-forward?