

# INF1771 - INTELIGÊNCIA ARTIFICIAL

## TRABALHO 4 – APRENDIZADO DE MÁQUINA

### Descrição:

O trabalho 4 consiste em utilizar 4 algoritmos de aprendizado supervisionado de máquina (Árvores de Decisão, K-Nearest Neighbor, Support Vector Machines e Redes Neurais) para classificar atividades físicas realizadas por determinadas pessoas com base em dados fisiológicos adquiridos através de sensores.

Os dados que serão utilizados neste trabalho foram obtidos utilizando a braçadeira BodyMedia SenseWear Pro (<http://sensewear.bodymedia.com/>). Este equipamento, mostrado na figura abaixo, permite o monitoramento contínuo de dados fisiológicos de um usuário fora de um laboratório.



O equipamento é usado na parte de trás do braço direito e possui o seguinte conjunto de sensores:

- **Acelerômetro:** Sensor de dois eixos que medem o movimento realizado pelo corpo da pessoa.
- **Fluxo de Calor:** Sensor que mede a quantidade de calor a ser dissipado pelo corpo.
- **Resposta Galvânica da Pele:** Sensor que mede a condutividade elétrica entre dois pontos no braço do portador.
- **Temperatura da Pele:** Sensor que mede a temperatura da pele da pessoa.
- **Temperatura Próxima do Corpo:** Sensor que mede a temperatura do ar em torno da braçadeira do usuário.

O conjunto de dados de treinamento consiste em aproximadamente 10.000 horas de dados coletados de usuários. Cada registro da base de treinamento consiste nos dados coletados em intervalos de 1 minuto. Cada registro é composto por 16 valores, dos quais apenas 9 são dados fisiológicos. Os atributos que compõem os registros são:

- **Característica1:** Idade
- **Característica2:** Canhoto ou destro (0 ou 1)
- **Anotação:** Código da atividade física sendo realizada.
- **Gênero:** Masculino ou feminino (0 ou 1)
- **Sensor1:** Resposta Galvânica da Pele (condutividade através da pele).
- **Sensor2:** Fluxo de Calor (medida de calor perdido para o ambiente).
- **Sensor3:** Temperatura Próxima do Corpo.
- **Sensor4:** Pedômetro (número de passos).
- **Sensor5:** Temperatura da Pele.
- **Sensor6:** Medida da soma das diferenças absolutas de aceleração vertical.
- **Sensor7:** Medida média de aceleração vertical.
- **Sensor8:** Medida da soma das diferenças absolutas de aceleração horizontal.
- **Sensor9:** Medida média de aceleração horizontal.
- **UserID:** Código identificador do usuário.
- **SessionID:** Código identificador da sessão.
- **SessionTime:** Tempo referente a sessão.

As classes são identificadas pelo campo **anotação**. Quando a anotação é igual a 0 significa que o usuário não anotou o que ele estava fazendo naquele momento. Portanto, você deve escolher se irá simplesmente ignorar os exemplos não rotulados ou irá considerá-los exemplos negativos.

Infelizmente o fornecedor dos dados não informa exatamente qual é a atividade representada por cada código do campo anotação. Conhecemos apenas o significado dos seguintes códigos:

- **5102:** Dormindo.
- **3104:** Assistindo TV.

Entretanto, existem outras classes de atividades, as quais estão corretamente rotuladas por um código. No trabalho todas as classes devem ser consideradas, mesmo sem sabermos exatamente o que elas significam.

O conjunto de dados de treinamento pode ser acessado no seguinte link:

<http://edirlei.eternix.com.br/aulas/ia/Trabalho4Treinamento.zip>

O conjunto de dados de teste pode ser acessado no seguinte link:

<http://edirlei.eternix.com.br/aulas/ia/Trabalho4Teste.zip>

O trabalho consiste em treinar 4 classificadores diferentes com o conjunto de dados de treinamento e depois fazer classificação do conjunto de dados de teste para verificar se o classificador aprendeu a identificar corretamente as atividades físicas com base nos dados fisiológicos captados pelos sensores. Os algoritmos de aprendizado supervisionado que devem ser utilizados são:

- Árvores de Decisão.
- K-Nearest Neighbor (KNN)
- Support Vector Machine (SVM)
- Rede Neural (usando backpropagation)

O objetivo do trabalho é observar como cada um dos classificadores se comporta com este conjunto de dados e qual apresenta o melhor resultado. No final do trabalho, deve ser entregue um relatório descrevendo o resultado dos experimentos realizados para cada um dos algoritmos, incluindo a taxa de acertos, tempo gasto no processo de treinamento e tempo gasto no processo de classificação.

### **Informações Adicionais:**

- Não é necessário implementar todos os classificadores. É permitida a utilização de bibliotecas externas, como por exemplo, a LibSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>).
- Não é necessário desenvolver um programa totalmente integrado. Você pode ter um pequeno programa para cada classificador.
- Provavelmente será necessário desenvolver um pequeno programa para formatar os dados de treinamento e de teste para eles possam ser acessados mais facilmente durante o processo de treinamento e classificação.
- Os programas podem ser implementados em qualquer linguagem (C, C++, C#, Java...).
- Você deve decidir quais atributos serão utilizados pelos classificadores. A classificação pode piorar ou melhorar dependendo do conjunto de atributos utilizado.
- Provavelmente será necessário normalizar os dados para se conseguir melhores resultados.
- **IMPORTANTE:** O conjunto de dados utilizado neste trabalho também foi utilizado no concurso “Physiological Data Modeling Contest” que foi realizado na “International Conference on Machine Learning (ICML)” em 2004. A página do concurso pode ser uma importante fonte de informações para ajudar no entendimento dos dados: <http://www.cs.utexas.edu/users/sherstov/pdmc/>

### **Forma de Avaliação:**

Será avaliada a qualidade do relatório final e a correta utilização dos classificadores. Neste trabalho, o mais importante será o relatório final contendo os resultados dos testes. Entretanto, também será necessário apresentar os programas para comprovar os dados do relatório.

**Data de Entrega:**

29/06

**Forma de Entrega:**

O programa e o relatório devem ser apresentados na aula do dia 29/06 (quarta) e enviados até o dia 29/06 para o email [edirlei.slima@gmail.com](mailto:edirlei.slima@gmail.com). Como estamos no final do semestre, não serão aceitos trabalhos enviados depois desta data.