



## Exercício Prático 4: Classificador Naive Bayes

### Introdução

Neste exercício você implementará o método Naive Bayes e verá como ele utiliza os dados para fazer classificações de amostras não vistas. Antes de começar este exercício, é recomendável que você revise os conceitos apresentados em aula.

### Arquivos incluídos neste exercício

`ex04.m` – Script geral do exercício

`ex04Dados.mat` – Base de dados com amostras de partida

[ \* ] `calcularProbabilidades.m` – Função para calcular a probabilidade de ocorrência de cada atributo em cada classe

[ \* ] `classificacao.m` – Função que calcula a probabilidade condicional Bayesiana de uma amostra pertencer a uma classe dado o conjunto de atributos

\* indica os arquivos que você precisará completar.

O arquivo `ex04.m` conduzirá todo o processo desse exercício.

### O Problema

Rodonildo é um jogador nato de *League of Legends*<sup>1</sup> e esteve coletando dados nas partidas em que jogou. O objetivo de Rodonildo é prever o vencedor de uma determinada batalha a partir de algumas informações. Na coleta de dados que Rodonildo fez, ele utilizou amostras compostas pelos 5 atributos binários (1 = **sim** e 0 = **não**) a seguir:

1. `primeiroAbate`: indica se a primeira morte do jogo foi realizada pelo time de Rodonildo;
2. `primeiraTorre`: indica se a primeira torre destruída do jogo foi derrubada pelo time de Rodonildo (Figura 1a);
3. `primeiroInibidor`: indica se o primeiro inibidor destruído do jogo foi derrubado pelo time de Rodonildo (Figura 1b);

---

<sup>1</sup>Jogo de estratégia que envolve a batalha entre dois times. Para maiores detalhes, consulte <http://br.leagueoflegends.com/>.

4. `primeiroDragao`: indica se o personagem Dragão foi abatido primeiro pelo time de Rodonildo (Figura 1c);
5. `primeiroBaron`: indica se o personagem Baron foi abatido primeiro pelo time de Rodonildo (Figura 1d).

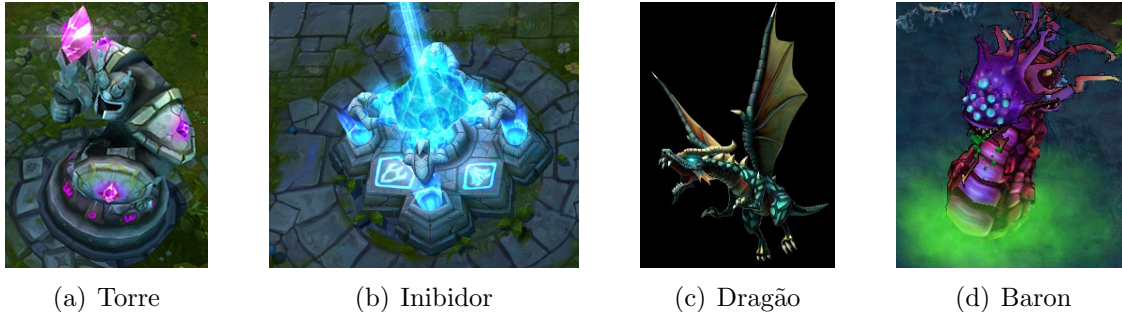


Figura 1: Objetos e Criaturas de League of Legends

Por exemplo, a amostra  $\mathbf{x} = [0 \ 0 \ 1 \ 1 \ 0]$  e  $\mathbf{y} = 0$ , representa um jogo no qual o time de Rodonildo destruiu primeiro um inibidor inimigo e derrotou o dragão antes da equipe inimiga. Por sua vez, a equipe adversária fez o primeiro abate do jogo, destruiu a primeira torre e derrotou o Baron. Essa partida foi vencida pela equipe adversária.

Após longo período de coleta de dados, Rodonildo precisa da sua ajuda para prever o resultado de outras partidas utilizando as informações armazenadas. A sua função é implementar o classificador Naive Bayes para predizer qual será o resultado das próximas partidas de Rodonildo, condicionado aos valores dos atributos.

## Probabilidade das classes

Na primeira etapa do procedimento `ex04.m`, a base de dados com as partidas jogadas por Rodonildo é carregada e são encontradas as probabilidades de ocorrências de cada classe. A seguir, o procedimento `ex04.m` chama a função para calcular a probabilidade de ocorrência de cada atributo em cada classe.

*Você precisará completar a função `calcularProbabilidades.m`.*

Se a sua função para encontrar as probabilidades estiver correta, espera-se que a probabilidade do time de Rodonildo ter feito o primeiro abate em partidas que o time dele venceu  $P(\text{PrimeiroAbate} = 1 | \text{Classe} = 1)$  seja aproximadamente igual à 52,96%.

## Classificador Naive Bayes

A próxima etapa do procedimento `ex04.m` é realizar, de fato, a classificação das amostras com base nas probabilidades encontradas no passo anterior. A classificação é realizada verificando se a amostra em questão tem maior probabilidade de pertencer à classe 1 ou à classe 0. Para calcular a probabilidade de uma amostra pertencer a uma determinada classe, é necessário utilizar as probabilidades de ocorrências de atributos previamente computadas. O cálculo pode ser expresso como:

$$P(y_j|\vec{x}) = \hat{P}(y_j) \prod_{x_i \in \vec{x}} \hat{P}(x_i|y_j)$$

Portanto, a probabilidade de uma amostra  $\vec{x}$  pertencer a uma classe  $j$  é obtida a partir da probabilidade geral da classe  $j$  ( $\hat{P}(y_j)$ ) multiplicada pelo produtório da probabilidade de ocorrência de cada atributo  $x_i$  com relação a classe  $j$  ( $\hat{P}(x_i|y_j)$ ).

Se a rotina de classificação estiver correta, espera-se que a acurácia obtida ao classificar a própria base de amostras de jogos que Ronildo participou seja aproximadamente igual à 76,60%.

*Você precisará completar a função **classificacao.m**.*

## Novas amostras

Após toda a etapa de treinamento e classificação estar concluída, o último passo do procedimento **ex04.m** é permitir que novas amostras sejam classificadas. Para facilitar a classificação, o programa perguntará a você as informações referentes aos atributos e, ao final, responderá se essa partida em questão tem mais chance vitória ou derrota com base no classificador desenvolvido.