# Parte 1 – Métodos de Classificação

## Escolha das features

De forma a tentarmos obter um erro dentro dos limites de tolerância (≤ 30%) tendo em conta os exemplos de classificação (True ou False) de palavras dados no enunciado do projeto e as palavras para avaliação incluídas no ficheiro de texto, optámos por escolher as seguintes *features*:

🡪 Tamanho da palavra;

🡪 Paridade do número de vogais;

🡪 Primeira letra da palavra é uma vogal;

🡪 Palavra tem acentuação;

🡪 Palavra tem letras repetidas;

## 1.2 Análise dos métodos de aprendizagem

Com estas *features* testámos diferentes métodos de aprendizagem através dos seguintes algoritmos, de onde obtivemos as seguintes percentagens de sucesso de classificação:

|  |  |  |
| --- | --- | --- |
| Algoritmo | Teste 1 | Teste 2 |
| Neighbors | 74% | 72% |
| Decision Tree | 84% | 84% |
| Linear | 73% | 79% |

Antes de uma decisão final, voltámos a analisar os métodos de aprendizagem com as seguintes *features*:

🡪 Número ímpar de letras;

🡪 Começar e terminal em vogal;

🡪 O sufixo da palavra ser uma palavra de 3 letras já classificada;

# 1.3 Conclusão

Após revermos os resultados obtidos, verificamos que nenhuma destas últimas *features* alcançavam uma percentagem maior do que as que já tínhamos obtido anteriormente, continuando assim o ***Decision Tree*** a obter a maior percentagem de sucesso.

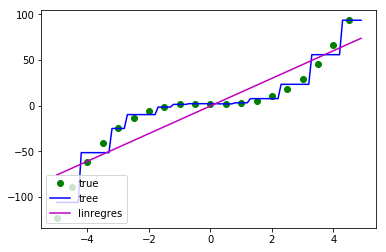
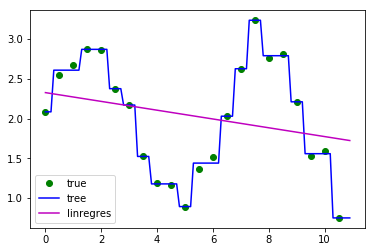
Mesmo utilizando os parâmetros ***max\_depth*** (profundidade máxima da árvore de decisão) e ***min\_samples\_split***, as percentagens de sucesso nunca foram superiores.

Concluindo assim que o melhor método de aprendizagem utilizando as *features* definidas, para classificar as palavras, é o ***Decision Tree***.

# Parte 2 – Métodos de Regressão

## Análise dos métodos de aprendizagem

### Linear Regression e Tree Regression

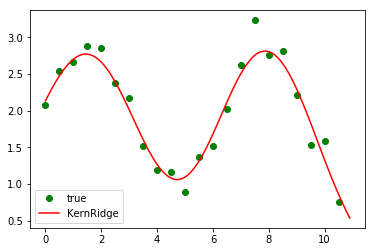


Teste de Regressão 1 - FAILED

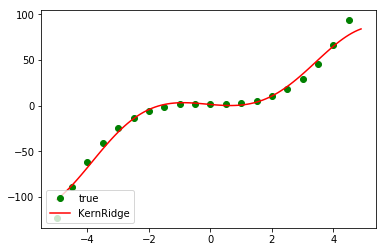
Teste de Regressão 2 - FAILED

Teste de regressão 1 - Linear e Tree

### Kernel Ridge

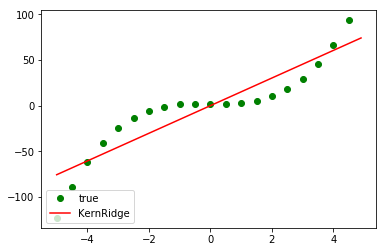
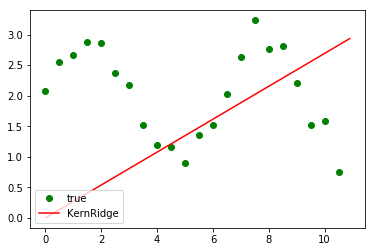


Teste de Regressão 1 (parâmetro kernel = rbf) - OK



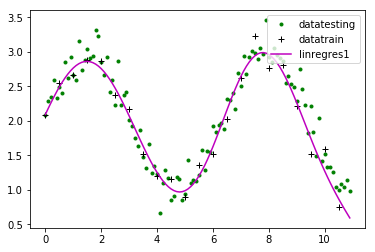
Teste de Regressão 2 (parâmetro kernel = rbf) - FAILED

Teste de Regressão 1 (parâmetro kernel = linear) - FAILED

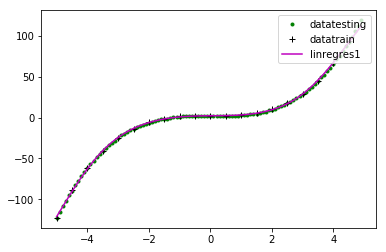


Teste de Regressão 2 (parâmetro kernel = linear) - OK

### GridSearchCV with Kernel Ridge



Teste de Regressão 1 (parâmetro kernel = rbf) - OK



Teste de Regressão 2 (parâmetro kernel = rbf) - OK