Universidade Federal de Sergipe

Centro de Ciências Exatas e Tecnológicas

Departamento de Computação

2ª Avaliação Prática de Inteligência Artificial

Alunos:

Antônio José Alves Neto

Iúri Batista Teles

Kelvin Marx do Nascimento Modesto

William Santana Nascimento

Matheus Pereira Alves

Wesley Oliveira Souza

Paulo Henrique dos Santos

São Cristóvão, maio de 2016

Resolução

**1-Visualização:** Esta é uma etapa comumente ignorada, mas de extrema importância.

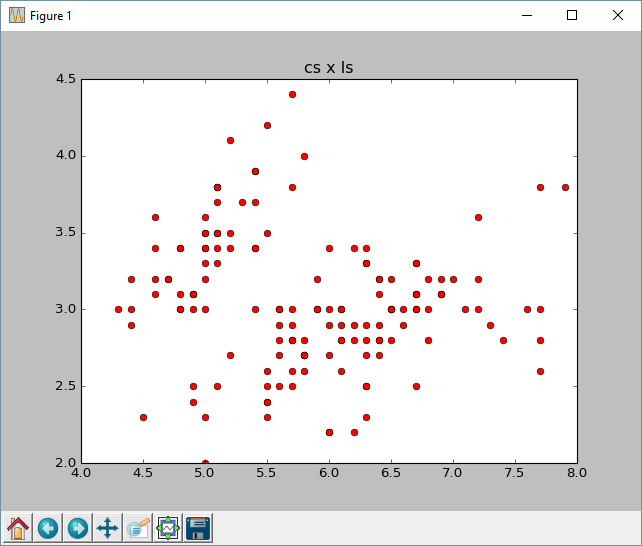
Muitas vezes é possível perceber padrões antes mesmo de decidir o que fazer. Muita

gente aplica os métodos de reconhecimento de padrões sem sequer olhar para os

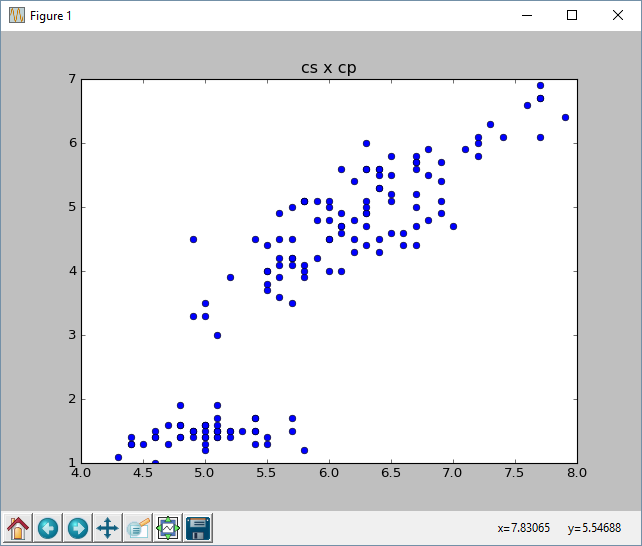
dados que tem. Nosso conjunto de dados tem 4 dimensões, não dá para desenhar.

1. **Faça um gráfico de cada uma das 6 projeções bi-dimensionais possíveis:**

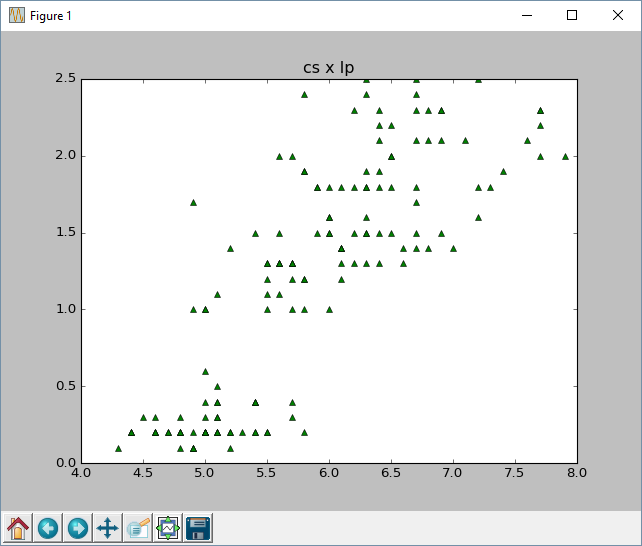
i.cs x ls



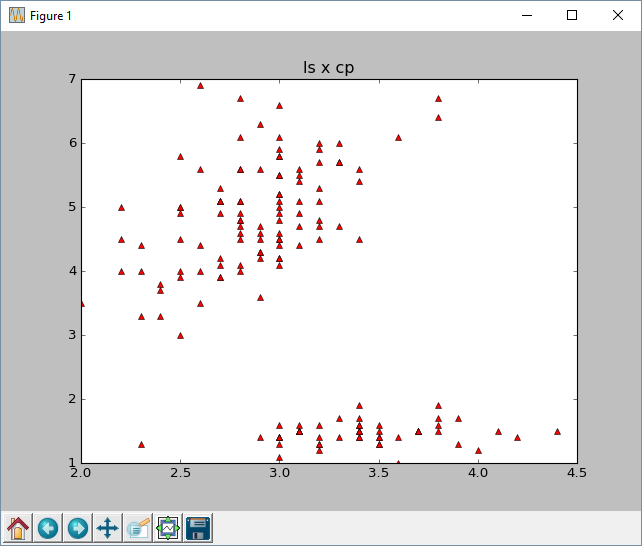
ii.cs x cp



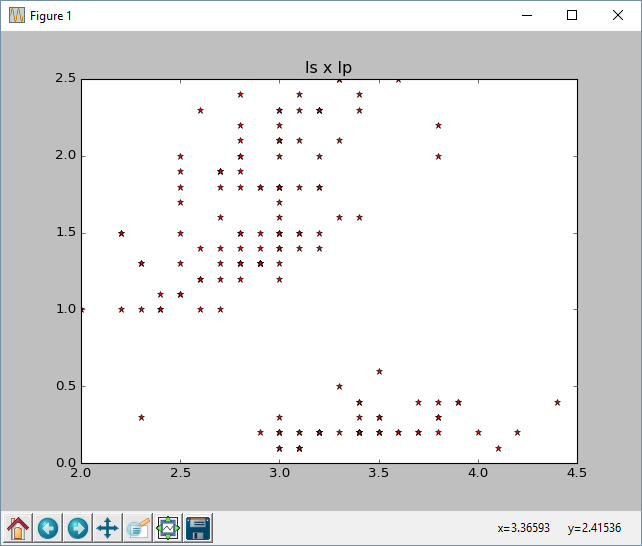
iii. cs x lp



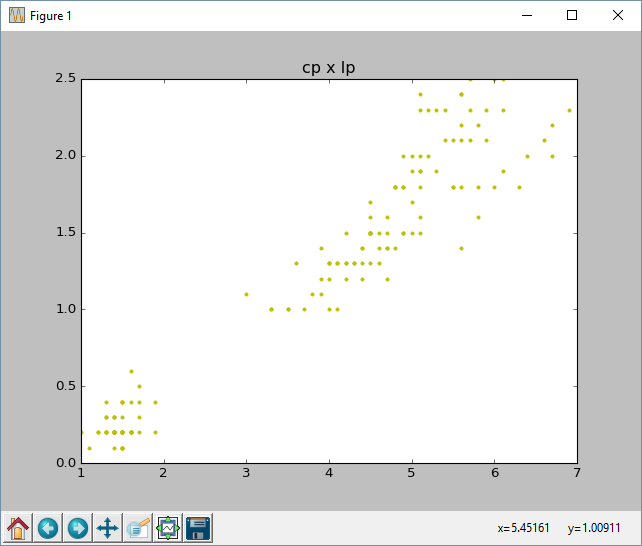
iv. ls x cp



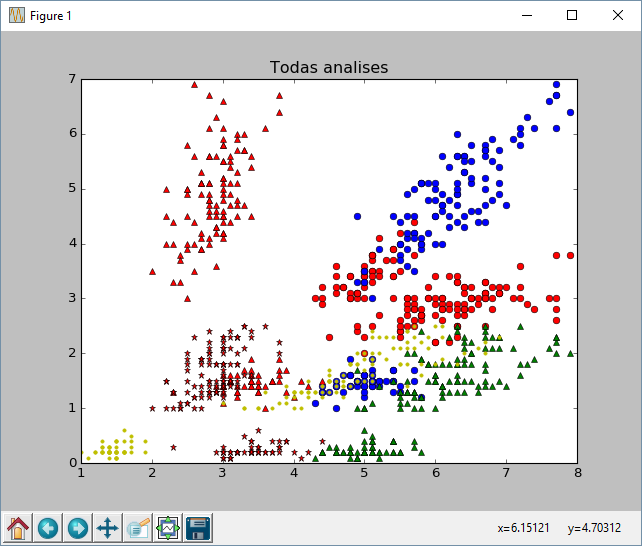
v. ls x lp



vi. cp x lp



Combinação dos **i** ao **vi**.



**b) Responda:**

**i. Você percebe algum padrão nos dados?**

Na plotagem com maior dimensão nota-se um faixa de grupo que pode ser separado mais facilmente, porém nota-se também que ocorre uma mistura parcial dos dados.

**ii. Existe algum grupo que é mais fácil de separar?**

Como existe uma mistura parcial, a separação dos dados por completo parece ser algo mais difícil.

**iii. Qual dimensão parece ser mais informativa? (útil para classificação)**

A apresentação do gráfico na terceira dimensão irá apresentar um efeito mais informativo e ilustrativo com relação a aglutinação dos dados da flor.

2-**Aprendizagem Não-Supervisionada (agrupamento de dados):** Nesta forma de

aprendizagem os rótulos de classe (quinta coluna do nosso arquivo) não são

utilizados no treinamento, apenas as características. Os algoritmos buscam

“regularidades” nos dados.

a)Implementação

**i.Aplique um agrupamento hierárquico (aglomerativo ou divisivo) aos dados.**

Código: Hierarchical\_Clustering.py

**ii.Aplique o algoritmo k-means aos dados.**

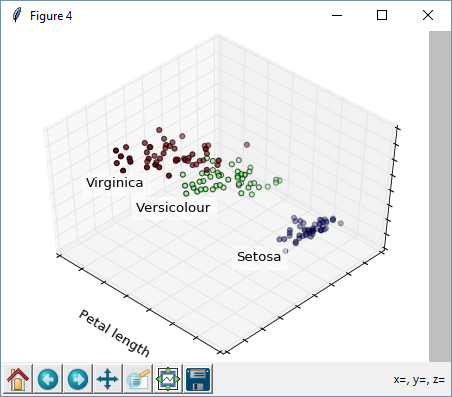
Código: k-means.py

b)

**i.Os grupos resultantes correspondem aos tipos das plantas?**

Hierarchical\_Clustering - não corretamente

k-means - sim



**ii.Há algum grupo mais destacado dos demais?**

A **Setosa** apresenta uma separação dos demais grupos mais acentuada.

**iii.A escolha da medida de distância entre grupos altera o resultado obtido?**

Sim, razoavelmente.

**iv.Se você usar somente uma das características para realizar o**

**agrupamento ainda é possível identificar grupos?**

Sim

**1.Qual a melhor característica a escolher?**

O comprimento da pétala – cp, pois ocorre maior distinção entre os dados.

**2.Que utilidade isto teria?**

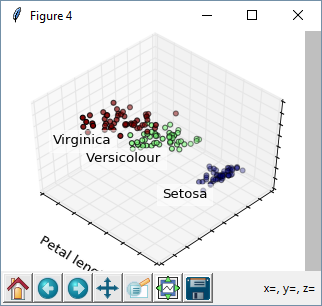
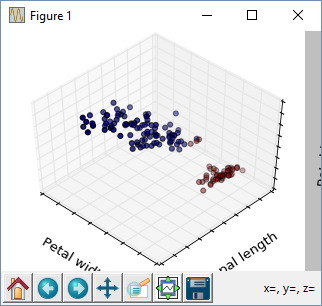
Uma diminuição no processamento, uma vez que são consideradas menos dimensões. Em algoritmos complexos e um maior numero de dados, tal fator pode ser aceitável pela taxa de erro viável.

**v.Para o k-means, o que acontece se:**

**1.k=2**

Ocorre aglutinação em apenas dois grupos distintos. Ocasionando assim a impossibilidade de identificar entre iris-versicolor e iris-virginica.

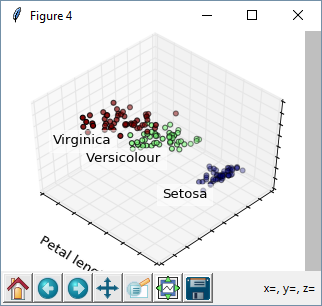
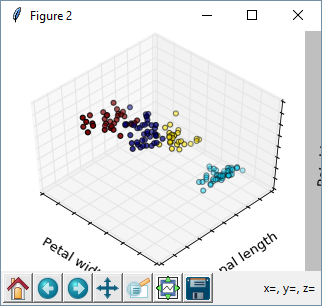
Ex:



**2.k=4**

Ocorre um diferenciamento a inclusão de um grupo extra, o que ocasiona maior indice de falha na classificação das flores.

Ex:



3-**Aprendizagem Supervisionada:** Neste caso utilizamos as informações que já

possuímos sobre qual tipo de planta cada exemplo é para treinar um classificador.

a)**Implementação:** para o treinamento dos classificadores é preciso pensar em

como representar os dados, tanto as características quanto os rótulos de

classe.

i. **Treine uma árvore de decisão utilizando os dados.**

Código: Arvore\_Decisao.py

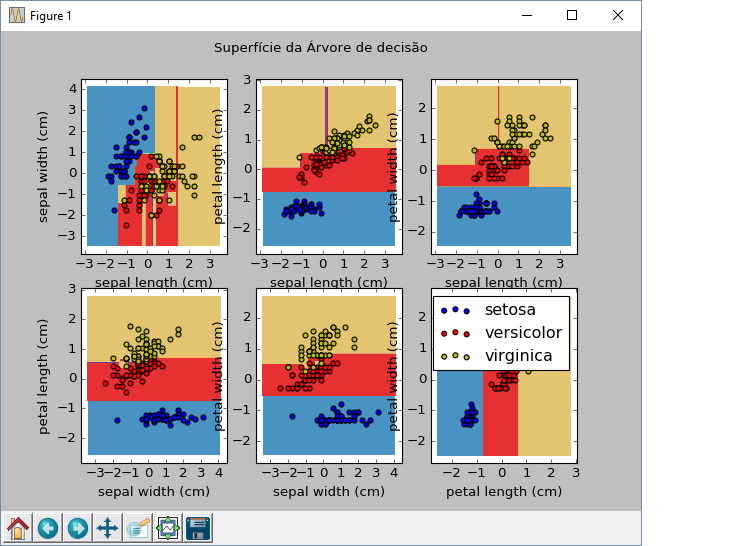
ii. **Treine uma rede neural MLP utilizando os dados.**

Código: Neural\_MLP2.py

b) **Para a árvore de decisão:**

i. **Qual o primeiro “corte” feito pela árvore? Você esperava isso?**

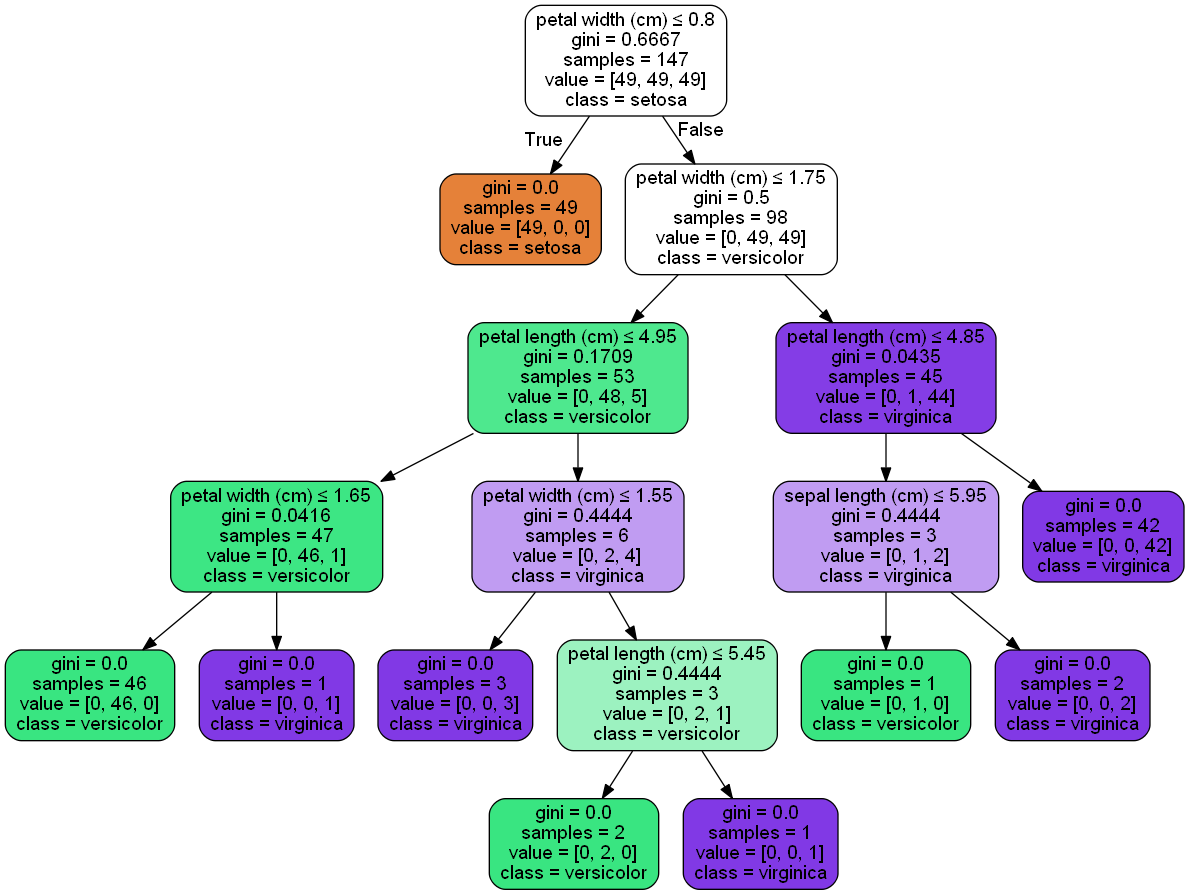
Não, imaginava um corte menos preciso, uma vez que os dados estão apenas em 2D(a lib foi bem desenvolvida por sinal...).



ii. **É possível classificar perfeitamente todos os pontos? Quantas**

**perguntas são necessárias?**

Sim, 8 perguntas.



iii. **Escolha N subconjuntos aleatórios de exemplos e treine N árvores com**

**no máximo K perguntas cada. Classifique os pontos usando as N**

**árvores ao mesmo tempo e diga que ele pertence à classe que a**

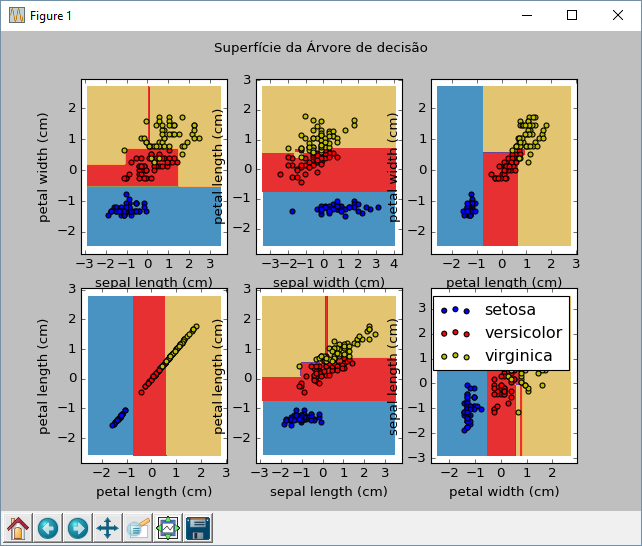
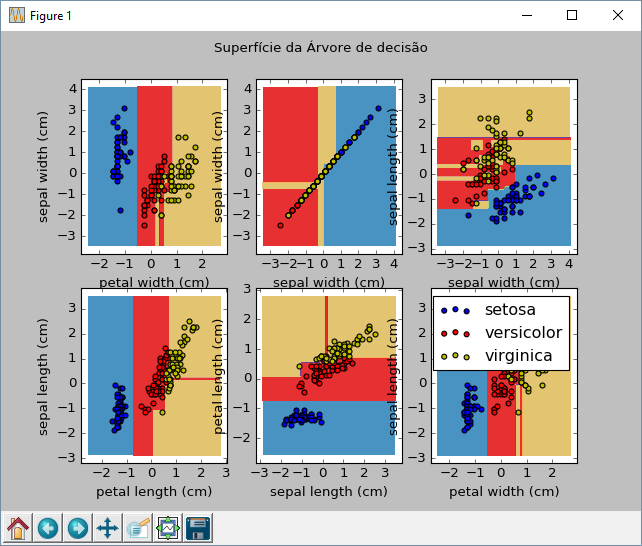
**maioria das árvores concordam.**

1. **Como foi o resultado?**
2. **Você consegue perceber como as escolhas de N e K afetam o**

**resultado e o custo computacional?**

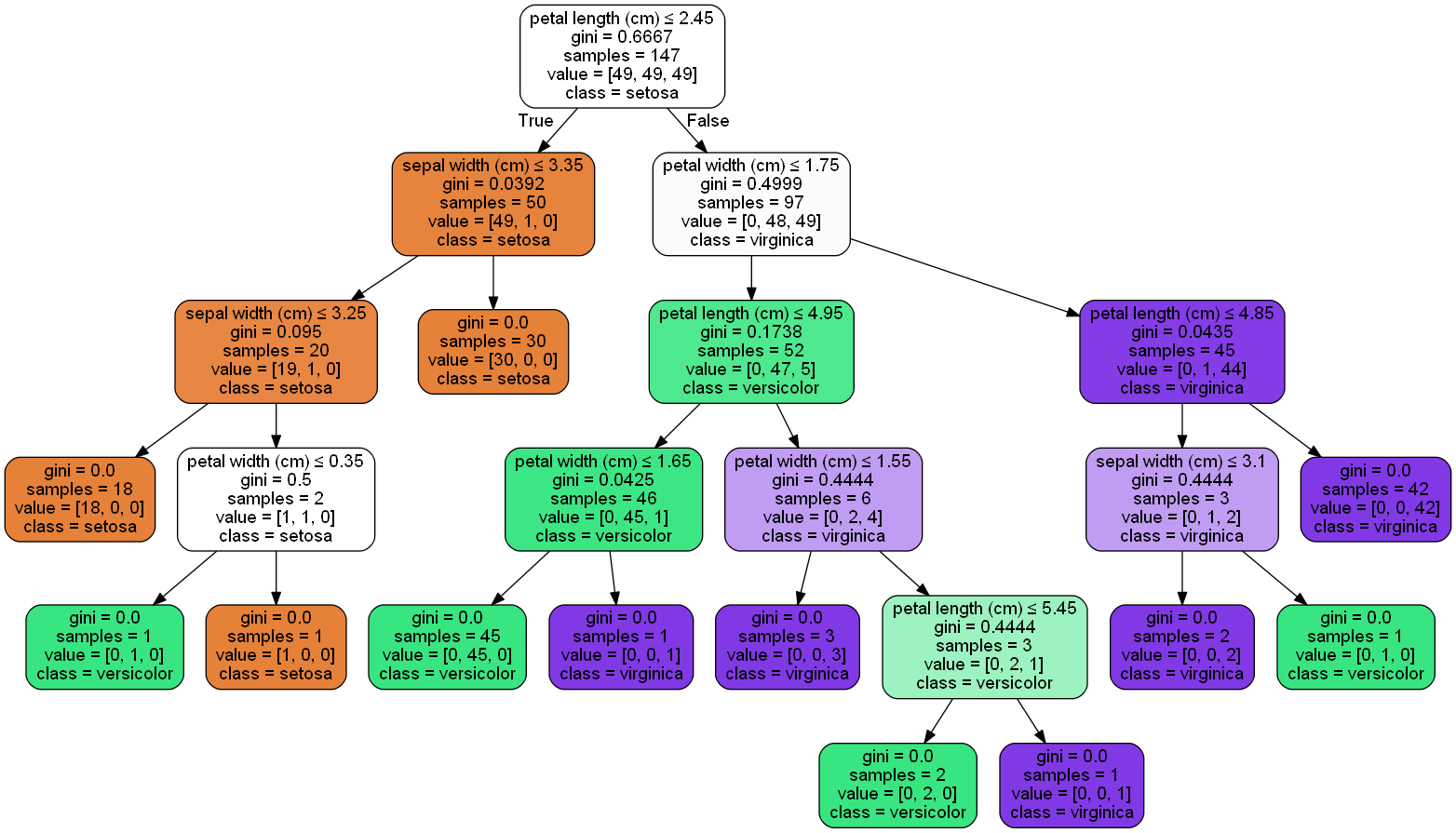
Alterando aleatoriamente o subconjunto alterar drasticamente a eficiência do corte.

Exemplos:



Bem como a variância dos grupos de teste afetas na criação e eficiência da Arvore de decisão.

Exemplo:



c)

i. **Como o número de neurônios na camada escondida afeta o**

**desempenho?**

Afeta na precisão e taxa de falha dos resultados.

ii. **Como a escolha na forma de representar a saída (rótulos) afeta o**

**desempenho?**

**(não entendi muito bem)**

A forma de representar uma característica interfere na probabilidade da presença ou ausência das outras características. A saída de cada camada interfere da complexidade da camadas posteriores caso ocorra um maior nível de permutação(“sinapse”) entre os neurônio.

iii. **Como a escolha na forma de representar a entrada afeta o**

**desempenho?**

Similar a saída da resposta anterior.

d)**Validação:** Utilizar métodos supervisionados para realizar classificações

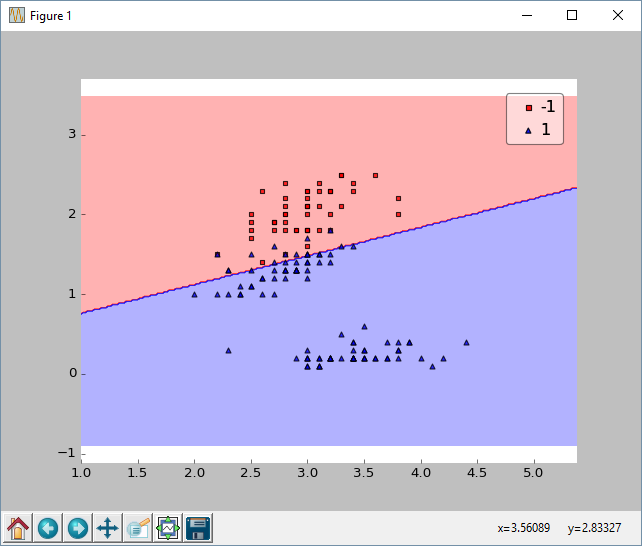
perfeitas de todos os exemplos é interessante para entender o

funcionamento dos algoritmos, mas pouco útil na prática. O que gostaríamos

é que os classificadores fossem capazes de generalizar a partir de alguns

exemplos.

i. **Separe 70% dos dados para treinamento e 30% para teste.**



ii. **Refaça o treinamento dos classificadores utilizando apenas esse**

**subconjunto de dados e verifique qual o desempenho no subconjunto**

**de teste.**

iii. **Quando devemos parar de treinar?**

Quando não ocorrer variação significativa na linha de corte, basicamente quando nao ocorrer mais classificações erradas no grupo de teste.

iv. **Já que o K-NN (K vizinhos mais próximos) não precisa de treinamento,**

**vamos testá-lo agora. Classifique os dados de teste usando KNN com**

**os dados de treinamento como referência.**

v. **A escolha dos dados de treinamento e teste afeta os resultados**

**obtidos com cada um dos métodos?**

Sim, os dados de treinamento influencia no aprendizado do algoritmo. Utilizando somente uma flor para treinamento, o algoritmo acaba tendo m resultado totalmente diferente para a classificação das outras flores.

