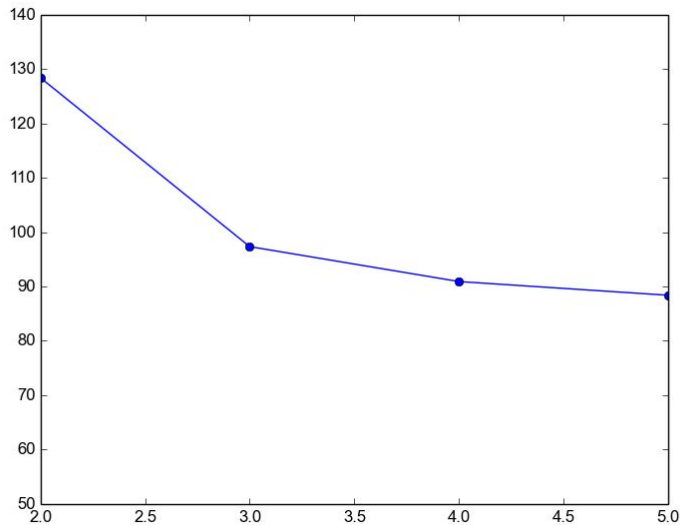


Paulo Bruno de Sousa Serafim - 388149

Aprendizagem Automática - Trabalho 6 - K-médias e PCA

Questão 1.

Parte 1:



O número de agrupamentos para este problema, de acordo com a heurística apresentada em aula.

A heurística, de forma geral, resume-se a verificar se o aumento no número de grupos implica em uma melhora da classificação. Para tanto, utiliza-se o gráfico “Quantidade de grupos” x “Erro”. Se a inclinação do gráfico de uma certa quantidade para uma unidade a mais de grupos for alta, então o aumento na quantidade valeu a pena. Caso a inclinação não seja tão acentuada, deve-se manter o número anterior de grupos.

Comente sobre o número de classes obtido.

Como podemos observar no gráfico acima, a diminuição do erro de 2 para 3 classes é bastante acentuada, portanto faz sentido usar 3 grupos. Já para a mudança de 3 para 4, o erro diminui muito pouco, portanto devemos considerar a existência de 3 agrupamentos.

Parte 2:

Execute o K-médias para o número de agrupamentos obtidos.

Foi executado o K-médias com 3 grupos.

Compare o resultado com o valor real das classes. Comente sobre o resultado obtido.

62 membros no grupo 0; 50 membros no grupo 1; 38 membros no grupo 2.

Podemos observar que 12 membros foram classificados erroneamente, ou seja, 8% do total. Por esse resultado, vemos que o K-médias funcionou de forma satisfatória.

Questão 2.

Parte 1:

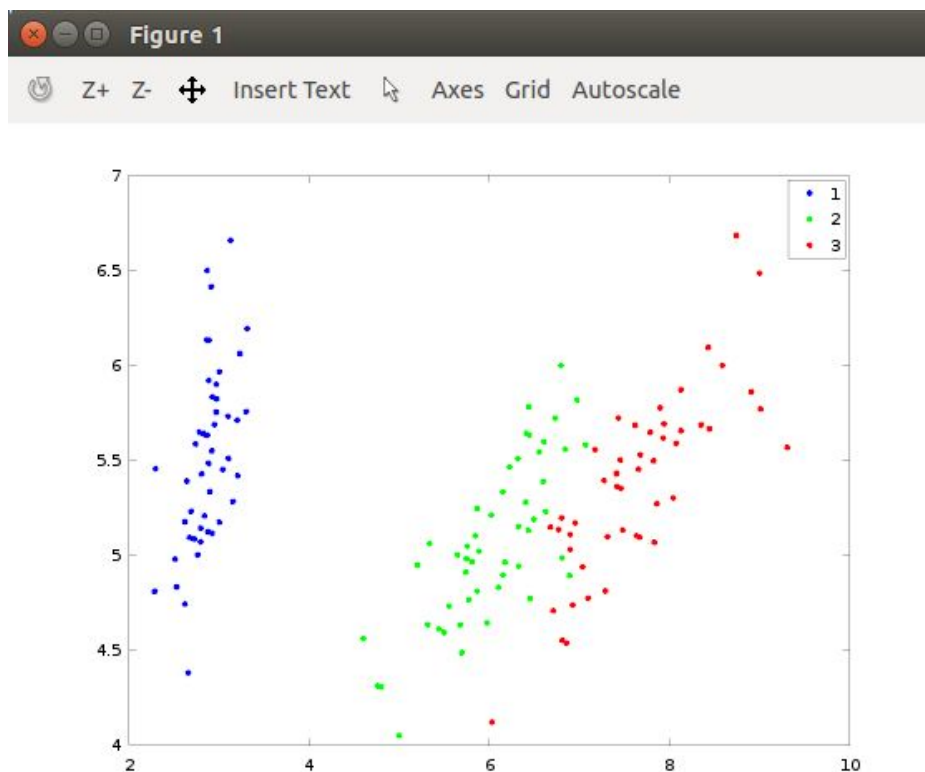
Número de atributos: 3

Comente sobre como foi obtido este número de atributos.

Autovalores: 4.19668
0.24063
0.07800
0.02353

Os autovalores representam as variâncias individuais absolutas de cada atributo. Para preservar 99% da variância, é necessário usar **3 atributos**, pois a variância é **99,482%** $([4.19668 + 0.24063 + 0.07800] / \text{soma})$. Já com 2 atributos a variância é 97,763% $([4.19668 + 0.24063] / \text{soma})$.

Parte 2:



Sabendo que uma classe é linearmente separável e as outras duas não são, verifique se este comportamento é mantido para o conjunto de dados com 2 dimensões.

Sim. Como podemos observar, a classe em azul é linearmente separável das outras duas. Já as classes verde e vermelha não são linearmente separáveis entre si.