

Aprendizaje Automático - Trabajo Práctico 4

Gonzalo Castiglione - 49138

May 27, 2012

Objetivo: Aprender a agrupar datos mediante análisis de clusters

1 Determinación de clases. Clustering.

1. (TO DO)
2. Lirios de Fisher
 - (a) Clasificación de las mediciones según el algoritmo *kmeans*

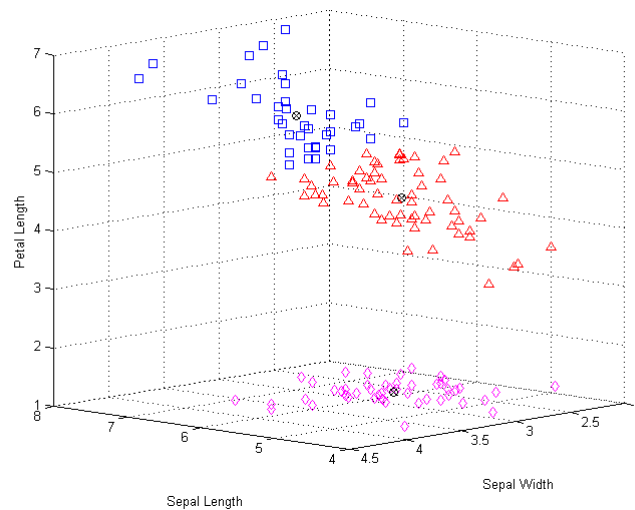


Figure 1: Agrupamiento de los lirios de Fisher

Matriz de confusión

	Clase Predicha			
		Setosa	Versicolor	Virginica
	Setosa	50	0	0
	Versicolor	0	42	8
Clase Real	Virginica	0	14	36

- (b) Calsificación

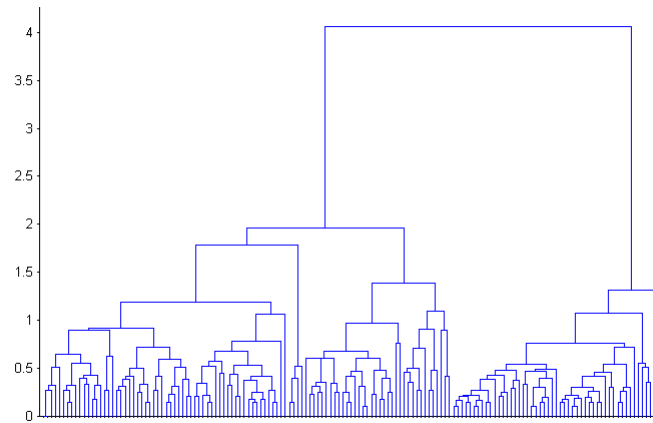


Figure 2: Agrupamiento de los lirios de Fisher

Matriz de confusión

	Clase Predicha			
		Setosa	Versicolor	Virginica
Clase Real	Setosa	50	0	0
	Versicolor	0	49	1
	Virginica	0	15	35

2 Código

1. (TO DO)
2. Lirios de Fisher

```
// a)
ptsymb = {'bs','r^','md','go','c+'};
load fisheriris;
k = 3;
[cidx,cmeans,sumd] = kmeans(meas,k);
for i = 1:k
    clust = find(cidx==i);
    plot3(meas(clust,1), meas(clust,2), meas(clust,3), ptsymb{i});
    hold on
end
plot3(cmeans(:,1), cmeans(:,2), cmeans(:,3),'ko');
plot3(cmeans(:,1), cmeans(:,2), cmeans(:,3),'kx');
hold off
xlabel('Sepal Length');
ylabel('Sepal Width');
zlabel('Petal Length');
view(-137,10);
grid on
types = [ones(1,50)*2 ones(1,50) ones(1,50)*3];
cMat = confusionmat(types,cidx)
// b)
eucD = pdist(meas, 'euclidean');
clustTreeEuc = linkage(eucD, 'average');
[h, nodes] = dendrogram(clustTreeEuc, 0);
set(gca, 'TickDir', 'out', 'TickLength', [.002 0], 'XTickLabel', []);
dist = pdist( data );
linkage_matrix = linkage( dist, 'ward' );
cluster_labels = cluster( linkage_matrix, 'maxclust', 3 );
confusion_matrix = crosstab( cluster_labels, class )
```