Si utilizzi lo spazio sottostante la domanda per rispondere e si utilizzi il comando set.seed(123) per avere risultati replicabili.

**Nome e Cognome:** Fare clic o toccare qui per immettere il testo.

**Matricola:** Fare clic o toccare qui per immettere il testo.

1. Si descriva dettagliatamente il metodo del clustering gerarchico.

Fare clic o toccare qui per immettere il testo.

2. Che cos’è la potatura di un albero decisionale e come si attua?

Fare clic o toccare qui per immettere il testo.

3. Si utilizzi il dataset `Funghi', si costruisca un albero decisionale e si incolli nello spazio sottostante.



Utilizzando l'albero decisionale ottenuto si risponda alle seguenti domande:

a. Quale variabile può essere considerata la più rilevante nella classificazione della commestibilità di un fungo e perché? Esistono delle variabili che non aiutano a classificare la commestibilità? Quali sono e perché?

L’odore è la piu importante.  
non sono importanti perché non si torvano nell’albero forma cappello, lividi, dim lamelle,

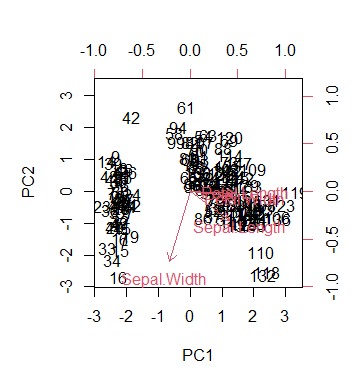
b. Si pensi di aver raccolto dei funghi. Sulla base delle loro caratteristiche e utilizzando l'albero decisionale si classifichino come commestibili o non commestibili:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Forma del cappello** | **Colore del cappello** | **Lividi** | **Odore** | **Dimensione lamelle** | **Radice del gambo** | **Colore del gambo** | **Habitat** | **Classificazione** |
| A campana | Bianco | Sì | Mandorla | Larghe | Gonfia | Grigio | Bosco | Commestibile |
| Concavo | Marrone | Sì | Anice | Larghe | Gonfia | Marrone | Nei prati | Non commestibile |
| A campana | Rosso | No | Pungente | Strette | Regolare | Camoscio | Tra l’erba | Non commestibile |
| A campana | Camoscio | No | Nessuno | Larghe | Regolare | Marrone | In zone incolte | Commestibile |

c. Qual è l’accuratezza dell’albero di classificazione su tutto il campione? Se dovessi dividere il campione in due (validation set), mi aspetto che questo errore di classificazione cresca o diminuisca?

Fare clic o toccare qui per immettere il testo.

4. Si utilizzi il dataset “iris” disponibile su R, si applichi il metodo dell’analisi delle componenti principali e si incolli il biplot nello spazio sottostante.



Utilizzando il biplot appena creato si risponda alle seguenti domande:

a. Quali fiori possono essere considerate degli outlier?

42, 61, 110, 118, 132

b. Esistono variabili che tendono a dare la stessa informazione? Se sì, quali?

Petal.lenght e petal.width