

Relazione Finale: Gli animali e gli alberi

Chiara Camerota

16 gennaio 2018

Esame di Multivariate Analysis & Statistical Learning, Prof. A. Gottard
Università degli studi di Firenze

Abstract

Il problema affrontato nelle prossime pagine, è di classificazione multinomiale. In particolare, il testo si divide in due parti: la prima chiamata gli animali, in cui viene spiegato brevemente il dataset utilizzato; la seconda chiamata gli alberi, in cui vengono spiegati velocemente i metodi con le relative applicazioni. I metodi utilizzati sono tutti basati su alberi decisionali: CART, Random forest e stochastic gradient boosting.

1 Gli animali

1.1 Presentazione dei dati

I dati considerati sono la raccolta di informazioni riguardo 101 animali presenti in uno zoo. Il dataset è formato da una variabile di risposta multinomiale, con 7 gruppi possibili, e un insieme di predittori per lo più booleani. L'obiettivo è quello di classificare gli animali nelle seguenti classi:

- mammiferi
- uccelli
- rettili
- pesci
- anfibi
- insetti
- invertebrati.

I predittori contengono caratteristiche fisiche e caratteriali degli animali, in particolare sono:

- hair** (*Booleano*): indica se l'animale presenta peli;
- feathers** (*Booleano*) : indica se ha un piumaggio;
- eggs** (*Booleano*) : indica se è oviparo;
- milk** (*Booleano*) : indica se allatta;
- airborne** (*Booleano*) : indica se è un volatile;
- aquatic** (*Booleano*) : indica se è acquatico;
- predator** (*Booleano*) : indica se è predatore;
- toothed** (*Booleano*): indica se ha i denti;
- [backbone]** (*Booleano*): indicase se ha una spina dorsale;
- breathes** (*Booleano*) : indica se respira;
- venomous** (*Booleano*): indicase se è velenoso;
- fins** (*Booleano*) : indicase se ha le pinne;
- legs** (*insieme di valori "0,2,4,5,6,8"*): indica il numero di gambe;
- tail** (*Booleano*) : indica se ha una coda;
- domestic** (*Booleano*) : indica se è domestico;
- catsize** (*Booleano*) : indica se ha le dimensioni di un gatto.

1.2 Analisi Preliminari

Molto spesso si tende a confondere i problemi di classificazione, con quelli di clustering, quindi è opportuno specificare cosa si intende per i primi, ovvero: l'insieme di metodologie statistiche che aspirano ad assegnare una classe ad un'osservazione di classe sconosciuta, sulla base di informazioni fornite da un campione di classe invece nota.

Avendo a disposizione il numero di classi e le assegnazioni di ognuna, risulta inopportuno usare metodi di clustering (ovvero: insieme di tecniche che hanno come scopo l'individuazione di gruppi).

Una volta capito il tipo di problema che stiamo affrontando, è importante capire la forma dei dati, un metodo veloce e intuitivo è quello di plottare la densità dei diversi gruppi con le diverse variabili. Nella seguente figura si può notare come la presenza di una classe rispetto ad un'altra sia sbilanciata.

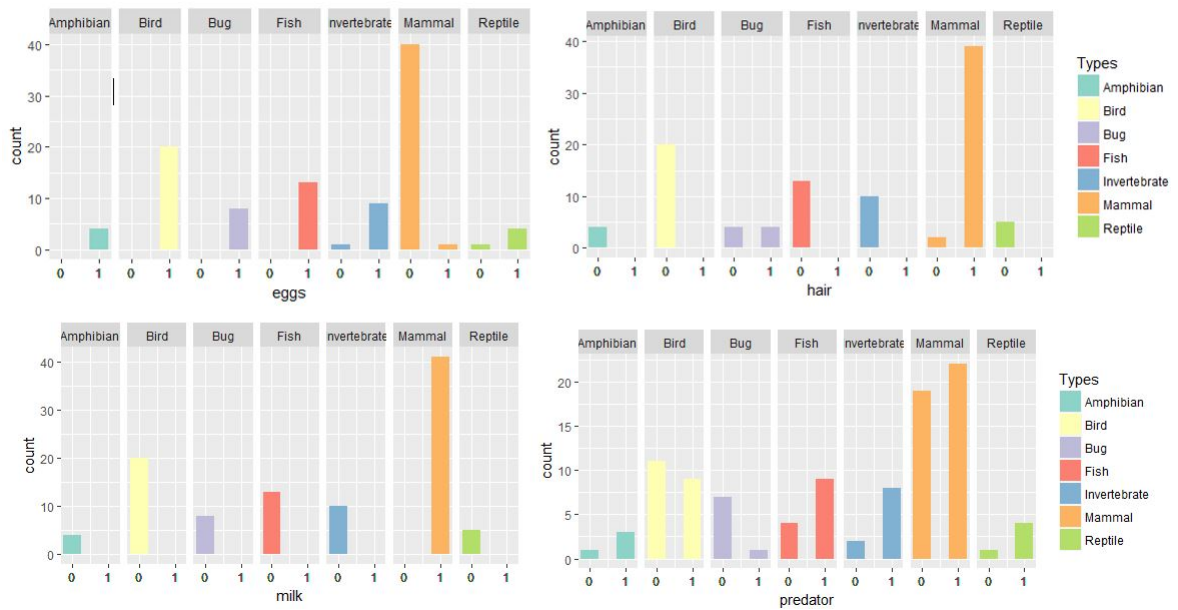


Figure 1: Istogrammi di alcuni predittori per ogni classe

Dall'immagine è intuibile come alcune variabili, tendono a caratterizzare completamente la classe degli animali, per esempio: il fatto che un animale sia oviparo, esclude la sua assegnazione alla classe dei mammiferi (fatta eccezione per i monotremi).

Un'altra importante relazione da esplorare è la correlazione tra variabili, questa ci darà un'idea completa del legame tra variabili. Il seguente grafico rappresenta le correlazioni, calcolate con l'indice di Spearman, sia graficamente che numericamente.

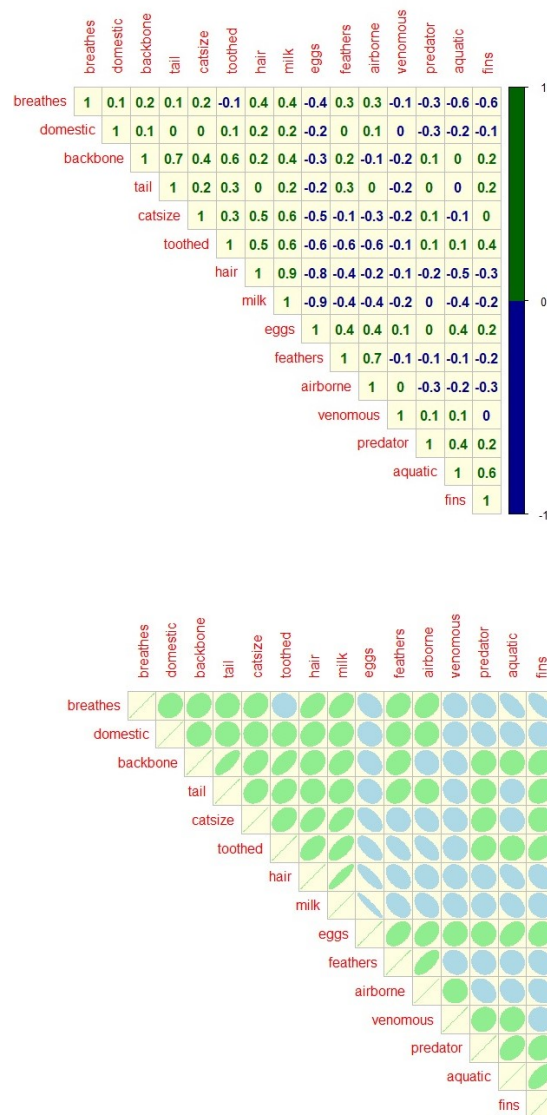


Figure 2: Grafici di correlazione tra i predittori

Possiamo notare che i predittori sono ampiamente e diversamente correlati, per lo più negativamente.

Per applicare i diversi metodi sui dati, è creare un train set, che in questo caso contiene il settanta per cento delle osservazioni, e un test set, con le restanti osservazioni.

Attraverso una summary possiamo avere un'idea di come è costituito il train e il test set.

```
> summary(data[dta == 1,])
```

| hair | feathers | eggs | milk | airborne | aquatic |
|----------------|----------------|----------------|----------------|----------------|----------------|
| Min. :0.0000 | Min. :0.0000 | Min. :0.0000 | Min. :0.0000 | Min. :0.0000 | Min. :0.0000 |
| 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 |
| Median :0.0000 | Median :0.0000 | Median :1.0000 | Median :0.0000 | Median :0.0000 | Median :0.0000 |
| Mean :0.4133 | Mean :0.2133 | Mean :0.5867 | Mean :0.3867 | Mean :0.2533 | Mean :0.3467 |
| 3rd Qu.:1.0000 | 3rd Qu.:0.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:0.5000 | 3rd Qu.:1.0000 |
| Max. :1.0000 | Max. :1.0000 | Max. :1.0000 | Max. :1.0000 | Max. :1.0000 | Max. :1.0000 |

| predator | toothed | backbone | breathes | venomous | fins |
|----------------|----------------|----------------|----------------|-----------------|----------------|
| Min. :0.0000 | Min. :0.0000 | Min. :0.0000 | Min. :0.0000 | Min. :0.00000 | Min. :0.0000 |
| 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:1.0000 | 1st Qu.:1.0000 | 1st Qu.:0.00000 | 1st Qu.:0.0000 |
| Median :1.0000 | Median :1.0000 | Median :1.0000 | Median :1.0000 | Median :0.00000 | Median :0.0000 |
| Mean :0.5067 | Mean :0.5733 | Mean :0.7867 | Mean :0.7733 | Mean :0.06667 | Mean :0.1467 |
| 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:0.00000 | 3rd Qu.:0.0000 |
| Max. :1.0000 | Max. :1.0000 | Max. :1.0000 | Max. :1.0000 | Max. :1.00000 | Max. :1.0000 |

| legs | tail | domestic | catsize | Types |
|---------------|--------------|----------------|----------------|-----------------|
| Min. :0.0000 | Min. :0.00 | Min. :0.0000 | Min. :0.0000 | Amphibian : 3 |
| 1st Qu.:2.000 | 1st Qu.:0.00 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | Bird :16 |
| Median :2.000 | Median :1.00 | Median :0.0000 | Median :0.0000 | Bug : 6 |
| Mean :2.867 | Mean :0.72 | Mean :0.1067 | Mean :0.4133 | Fish : 9 |
| 3rd Qu.:4.000 | 3rd Qu.:1.00 | 3rd Qu.:0.0000 | 3rd Qu.:1.0000 | Invertebrate:10 |
| Max. :8.000 | Max. :1.00 | Max. :1.0000 | Max. :1.0000 | Mammal :29 |
| | | | | Reptile : 2 |

Figure 3: Summary del train set

| ytest | hair | feathers | eggs | milk |
|-----------------|----------------|----------------|----------------|----------------|
| Amphibian : 1 | Min. :0.0000 | Min. :0.0000 | Min. :0.0000 | Min. :0.0000 |
| Bird : 5 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 |
| Bug : 4 | Median :0.0000 | Median :0.0000 | Median :1.0000 | Median :0.0000 |
| Fish : 7 | Mean :0.4545 | Mean :0.1515 | Mean :0.5758 | Mean :0.4242 |
| Invertebrate: 1 | 3rd Qu.:1.0000 | 3rd Qu.:0.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 |
| Mammal :14 | Max. :1.0000 | Max. :1.0000 | Max. :1.0000 | Max. :1.0000 |
| Reptile : 1 | | | | |

| airborne | aquatic | predator | toothed | backbone |
|----------------|---------------|----------------|---------------|----------------|
| Min. :0.0000 | Min. :0.000 | Min. :0.0000 | Min. :0.000 | Min. :0.0000 |
| 1st Qu.:0.0000 | 1st Qu.:0.000 | 1st Qu.:0.0000 | 1st Qu.:0.000 | 1st Qu.:1.0000 |
| Median :0.0000 | Median :0.000 | Median :1.0000 | Median :1.000 | Median :1.0000 |
| Mean :0.1818 | Mean :0.303 | Mean :0.5152 | Mean :0.697 | Mean :0.8485 |
| 3rd Qu.:0.0000 | 3rd Qu.:1.000 | 3rd Qu.:1.0000 | 3rd Qu.:1.000 | 3rd Qu.:1.0000 |
| Max. :1.0000 | Max. :1.000 | Max. :1.0000 | Max. :1.000 | Max. :1.0000 |

| breathes | venomous | fins | legs | tail |
|----------------|----------------|----------------|---------------|----------------|
| Min. :0.0000 | Min. :0.0000 | Min. :0.0000 | Min. :0.000 | Min. :0.0000 |
| 1st Qu.:1.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:2.000 | 1st Qu.:1.0000 |
| Median :1.0000 | Median :0.0000 | Median :0.0000 | Median :4.000 | Median :1.0000 |
| Mean :0.7576 | Mean :0.1212 | Mean :0.2121 | Mean :2.879 | Mean :0.7576 |
| 3rd Qu.:1.0000 | 3rd Qu.:0.0000 | 3rd Qu.:0.0000 | 3rd Qu.:4.000 | 3rd Qu.:1.0000 |
| Max. :1.0000 | Max. :1.0000 | Max. :1.0000 | Max. :6.000 | Max. :1.0000 |

| domestic | catsize |
|----------------|----------------|
| Min. :0.0000 | Min. :0.0000 |
| 1st Qu.:0.0000 | 1st Qu.:0.0000 |
| Median :0.0000 | Median :0.0000 |
| Mean :0.0303 | Mean :0.4848 |
| 3rd Qu.:0.0000 | 3rd Qu.:1.0000 |
| Max. :1.0000 | Max. :1.0000 |

Figure 4: Summary del test set

2 Gli alberi

Gli alberi di decisione, o di classificazione, sono analoghi agli alberi di regressione, infatti usano, non solo, le stesse operazioni sugli alberi, ma anche

la stessa logica.

I dati originari vengono divisi ricorsivamente in due parti rispetto ad uno, o più generici attributi, la suddivisione produce una gerarchia ad albero, dove i sottoinsiemi (di record) vengono chiamati nodi e, quelli finali (o terminali), foglie.

In particolare, i nodi sono etichettati con il nome degli attributi, gli archi sono etichettati con i possibili valori dell'attributo soprastante, mentre le foglie dell'albero sono etichettate con i differenti valori dell'attributo target (valori che descrivono le classi di appartenenza).

Un oggetto è classificato seguendo un percorso lungo l'albero che porti dalla radice ad una foglia. I percorsi rappresentano le regole di classificazione (o regole produttive).

Un'importante differenza tra alberi di classificazione e di regressione è il diverso metodo di valutazione degli split. Nel caso di classificazione non è possibile calcolare l'RSS, per cui si ricorre ad altri metodi, i più comuni sono:

- **Classification error rate:** proporzione di osservazioni, del train set, nella regione, che non appartengono alla classe più comune;

$$E = 1 - \max_k(\hat{p}_{mk})$$

dove \hat{p}_{mk} indica la proporzione di osservazioni del train set nella regione m-sima che appartiene alla classe k.

Questo indice non è ottimale se si vuole un albero parsimonioso, infatti non risente della grandezza di questo.

- **Indice di Gini:** misura la variabilità all'interno della classe k.

$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk})$$

Questo indice può indicare anche il grado di purezza di un nodo, infatti se assume bassi valori numerici nella regione di riferimento predomina una classe.

- **Cross-entropy:** metodo alternativo al precedente.

$$D = - \sum_{k=1}^K \hat{p}_{mk} \log \hat{p}_{mk}.$$

Negli ultimi anni, questi metodi sono sempre più utilizzati, infatti sono di semplice interpretazione per gli inesperti, in alcuni casi è possibile visualizzarli e non necessitano della creazione di variabili dummy in caso di variabili qualitative.

2.1 CART

La tipologia di alberi decisionali più semplice sono i CART (Classification and Regression Tree), di seguito è riportato l'algoritmo.

1. Scegli la radice (primo nodo).
2. Per ogni predittore X trova il sottoinsieme S che minimizza la somma dell'indice di impurità dei due nodi figli. Scegli la coppia che rende X e S minimo.
3. Quando la regola di uscita è soddisfatta, finisce l'algoritmo. Altrimenti continua ad iterare il punto 2. fino a che non finiscono i nodi figli.

Il precedente algoritmo generalizzato, vale per tutti quei metodi basati sugli alberi decisionali. In particolare applicando particolari funzioni sugli indici di impurità e usando, opportunamente, sottoinsiemi di X e di S , si ottengono diversi metodi, come quelli illustrati dopo.

2.2 Applicazione

Il pacchetto usato per le applicazioni è *Rpart*, questo usa il metodo di cross-validation k-fold con $k=10$ di default, inoltre si ha la possibilità di scegliere se usare l'indice di Gini o l'error rate come regola di split.

Di seguito sono rappresentati i due alberi ottenuti con le due diverse regole per lo split, come si può notare l'albero finale cambia, anche se come vedremo più avanti, la classificazione è pressochè identica.

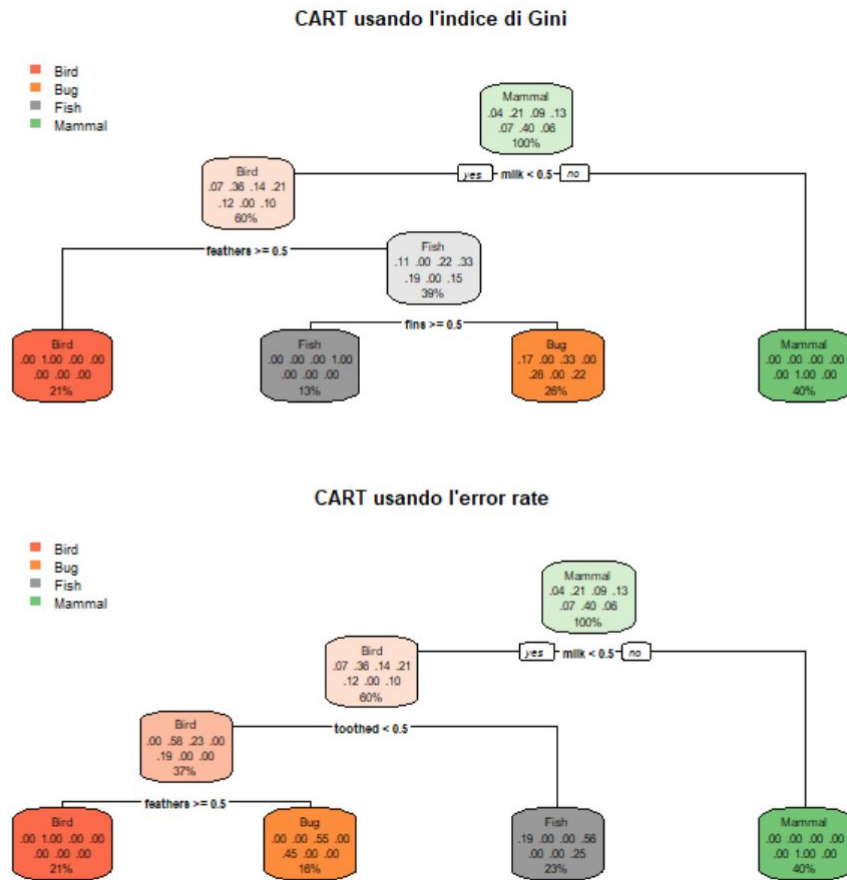


Figure 5: I due diversi alberi finali:il grafico riporta sia le variabili usate per lo split, sia la proporzione di ogni classe presente nel sottoinsieme, sia la percentuale di dati che si trovano in quel gruppo.

Un interessante parametro è quello di complessità (indicato come *cp*) che indica di quanto diminuisce l'errore dopo ogni split, fino ad un minimo fissato. È possibile visualizzare come il *cp* vari a seconda della grandezza dell'albero e dell'errore relativo.

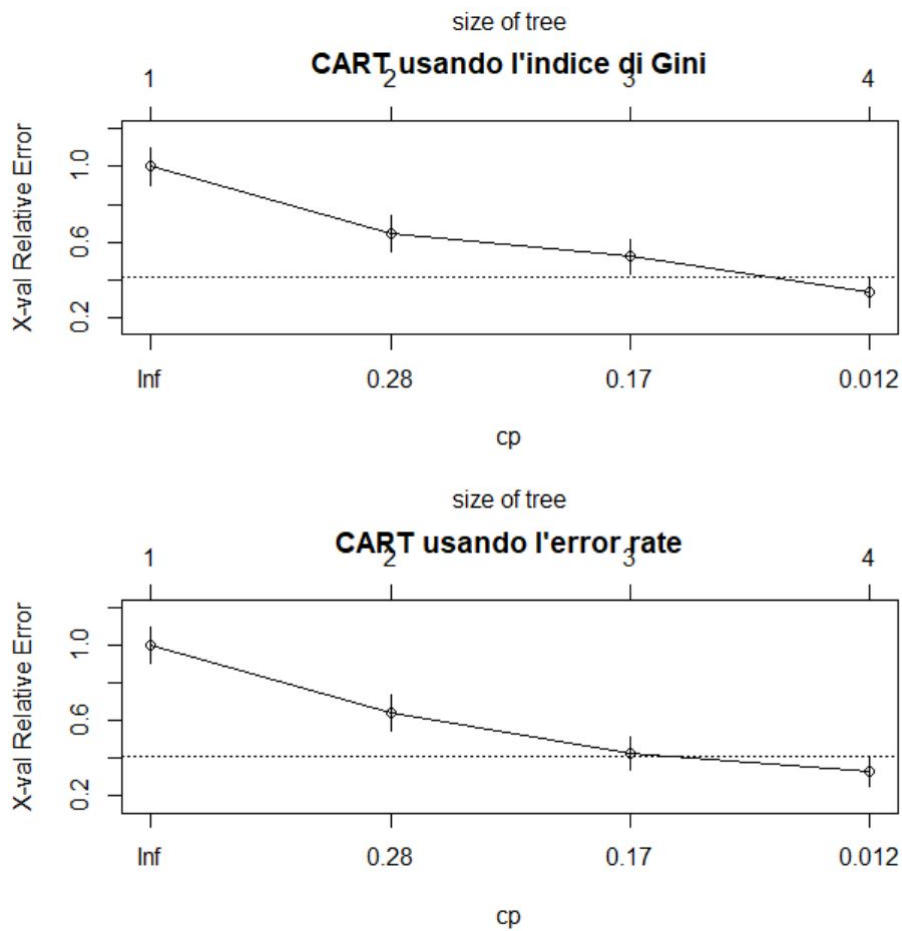


Figure 6: L'andamento del cp

In questo caso la curva dell'albero che usa l'error rate è più rapida dell'altra, questo perchè il terzo split viene fatto su diverse variabili. Provando i due modelli sul test set si nota come entrambi misclassificano le unità che appartengono a classi poco numerose, questo perchè usano solo quattro variabili per la costruzione, ignorando variabili come *backbone* che identifica un'intero gruppo.

| cpredg | ytest | | | | | | |
|--------------|-----------|------|-----|------|--------------|--------|---------|
| | Amphibian | Bird | Bug | Fish | Invertebrate | Mammal | Reptile |
| Amphibian | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bird | 0 | 5 | 0 | 0 | 0 | 0 | 0 |
| Bug | 1 | 0 | 2 | 0 | 5 | 0 | 1 |
| Fish | 0 | 0 | 0 | 4 | 0 | 0 | 0 |
| Invertebrate | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mammal | 0 | 0 | 0 | 0 | 0 | 13 | 0 |
| Reptile | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Figure 7: Matrice di confondimento, usando l'indice di Gini.

| cpredi | ytest | | | | | | |
|--------------|-----------|------|-----|------|--------------|--------|---------|
| | Amphibian | Bird | Bug | Fish | Invertebrate | Mammal | Reptile |
| Amphibian | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bird | 0 | 5 | 0 | 0 | 0 | 0 | 0 |
| Bug | 0 | 0 | 2 | 0 | 5 | 0 | 1 |
| Fish | 1 | 0 | 0 | 4 | 0 | 0 | 0 |
| Invertebrate | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mammal | 0 | 0 | 0 | 0 | 0 | 13 | 0 |
| Reptile | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Figure 8: Matrice di confondimento, usando l'error rate.

2.3 Random Forest

Il *bagging* o *bootstrap aggregation* è una procedura volta a ridurre la variabilità presente nei dati, l'idea è semplice: si divide il data set iniziali in più parti, con il metodo bootstrap, su ognuna delle quali si costruisce un classificatore, infine si combinano tutti questi.

Un problema di questo metodo è che i classificatori sono altamente correlati tra loro, per evitare questo, nasce la Random Forest: essa costruisce un certo numero di alberi, basati su campioni bootstrap, considerando un sottoinsieme di predittori e usando le classiche operazioni sugli alberi.

Una possibilità interessante è che possiamo stilare una graduatoria dei predittori in base alla loro importanza, basata sull'indice di Gini nel nodo corrispondente.

Infine, la Random Forest, al crescere dei campioni bootstrap riduce la probabilità di overfitting.

2.4 Applicazione

Il pacchetto usato è *RandomForest*.

In primo luogo è stata applicata una random forest con 150 alberi, usando due variabili a confronto per ogni split, e l'indice di Gini.

Di seguito è riportato il grafico dell'anadamento dell'errore al variare del numero degli alberi.

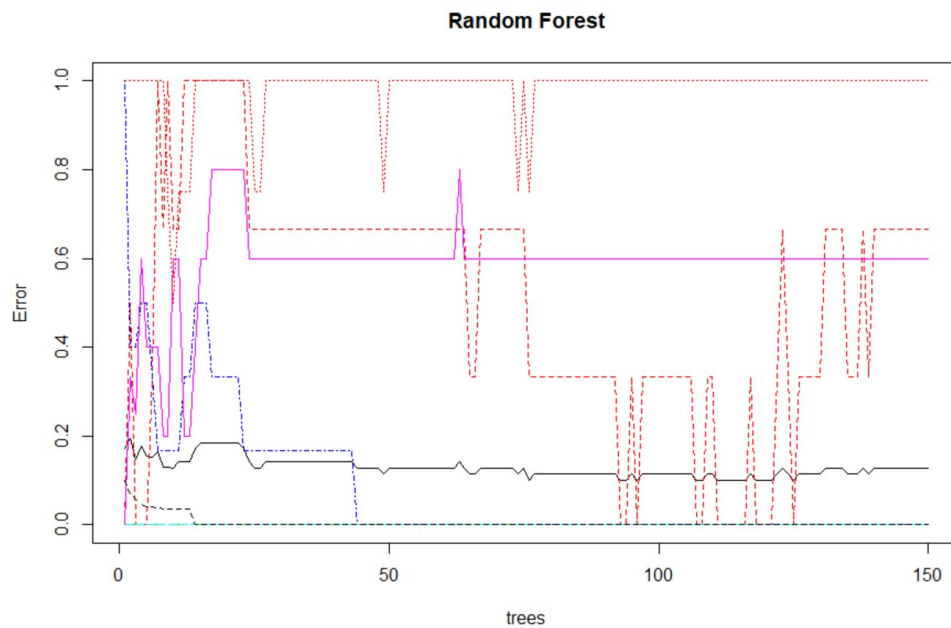


Figure 9: Andamento dell'errore

Avendo un numero parsimonioso di osservazioni, è interessante vedere le variabili più importanti, così da poter provare la performance dello stesso modello su un sottoinsieme di predittori.

Applicando la Random Forest solo sulle suddette variabili, otteniamo un errore identico al precedente, ma osservando le matrici di confondimento si può notare come gli errori di misclassificazione differiscono, sebbene l'errore totale rimanga identico.

| predf | ytest | | | | | | |
|--------------|-----------|------|-----|------|--------------|--------|---------|
| | Amphibian | Bird | Bug | Fish | Invertebrate | Mammal | Reptile |
| Amphibian | 1 | 0 | 0 | 0 | | 0 | 0 |
| Bird | 0 | 5 | 0 | 0 | | 0 | 1 |
| Bug | 0 | 0 | 2 | 0 | | 2 | 0 |
| Fish | 0 | 0 | 0 | 4 | | 0 | 0 |
| Invertebrate | 0 | 0 | 0 | 0 | | 3 | 0 |
| Mammal | 0 | 0 | 0 | 0 | | 0 | 13 |
| Reptile | 0 | 0 | 0 | 0 | | 0 | 0 |

Figure 10: Matrice di confondimento Random Forest

| predfvi | ytest | | | | | | |
|--------------|-----------|------|-----|------|--------------|--------|---------|
| | Amphibian | Bird | Bug | Fish | Invertebrate | Mammal | Reptile |
| Amphibian | 1 | 0 | 0 | 0 | | 0 | 0 |
| Bird | 0 | 5 | 0 | 0 | | 0 | 1 |
| Bug | 0 | 0 | 2 | 0 | | 3 | 0 |
| Fish | 0 | 0 | 0 | 4 | | 0 | 0 |
| Invertebrate | 0 | 0 | 0 | 0 | | 2 | 0 |
| Mammal | 0 | 0 | 0 | 0 | | 0 | 13 |
| Reptile | 0 | 0 | 0 | 0 | | 0 | 0 |

Figure 11: Matrice di confondimento Random Forest con selezione di variabili

2.5 Stochastic Gradient Boosting

Il metodo gradient boosting produce un modello predittivo come combinazione di modelli predittivi deboli, in questo caso alberi.

Il modello viene costruito in maniera simile a quello dei metodi di boosting, ma viene generalizzato permettendo l'ottimizzazione di una funzione di perdita differenziabile arbitraria. In particolare, con questa tecnica gli alberi crescono sequenzialmente: ogni albero cresce sfruttando le informazioni di quello precedente.

Una variante molto interessante è la versione stocastica, introdotta da Friedman, motivata dal metodo di bagging di Breiman.

Per ogni iterazione dell'algoritmo, un base learner viene applicato ad un campione casuale senza reintroduzione del train set; in questo modo l'accuratezza del metodo cresce notevolmente.

Il parametro di campionamento f è una frazione del train set, se è pari ad 1, l'algoritmo è deterministico, bassi valori introducono casualità e preven-
gono l'overfitting, viene usato anche come parametro di regolarizzazione. Tipicamente è impostato a 0.5.

2.6 Applicazione

Il pacchetto usato in questo caso è *caret*, che dispone di un'ampia scelta di modelli e di opzioni.

L'output descrive eusativamente il modello.

```
Stochastic Gradient Boosting

68 samples
16 predictors
7 classes: 'Amphibian', 'Bird', 'Bug', 'Fish', 'Invertebrate', 'Mammal', 'Reptile'

No pre-processing
Resampling: Cross-Validated (5 fold, repeated 5 times)
Summary of sample sizes: 54, 54, 54, 54, 56, 53, ...
Resampling results across tuning parameters:

  interaction.depth  n.trees  Accuracy  Kappa
1                   50      0.8756410  0.8350528
1                   100      0.8990989  0.8673443
1                   150      0.9050330  0.8749537
2                    50      0.8659267  0.8217592
2                   100      0.8904322  0.8560520
2                   150      0.8871648  0.8515514
3                    50      0.8839560  0.8460873
3                   100      0.8959267  0.8623446
3                   150      0.8990696  0.8673871

Tuning parameter 'shrinkage' was held constant at a value of 0.1
Tuning
parameter 'n.minobsinnode' was held constant at a value of 10
Accuracy was used to select the optimal model using the largest value.
The final values used for the model were n.trees = 150, interaction.depth =
1, shrinkage = 0.1 and n.minobsinnode = 10.
```

Figure 12: Descrizione del modello

È da notare che la cross validation ripetuta aiuta l'accuratezza della classificazione aumentando la variabilità dei sottocampioni.

Il metodo permette anche di vedere l'importanza delle variabili, che ci permette di creare un nuovo modello più parsimonioso.

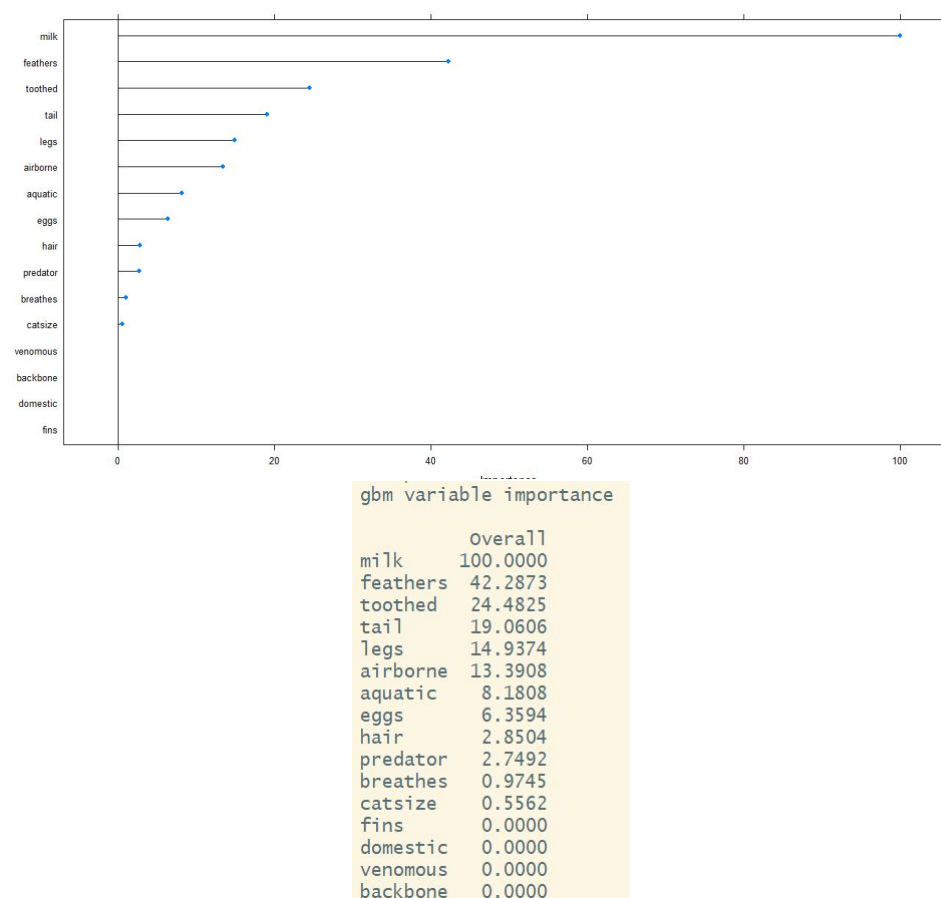


Figure 13: Variabili e relativa importanza

Sebbene riducendo i predittori nel modello l'accuratezza non cambi, cambia invece la curva dell'errore.

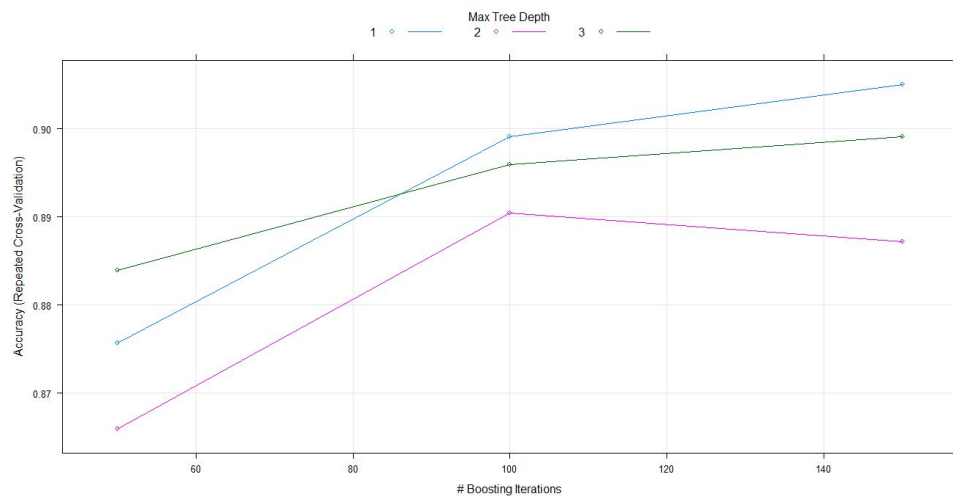


Figure 14: Accuratezza del modello senza selezione di variabili

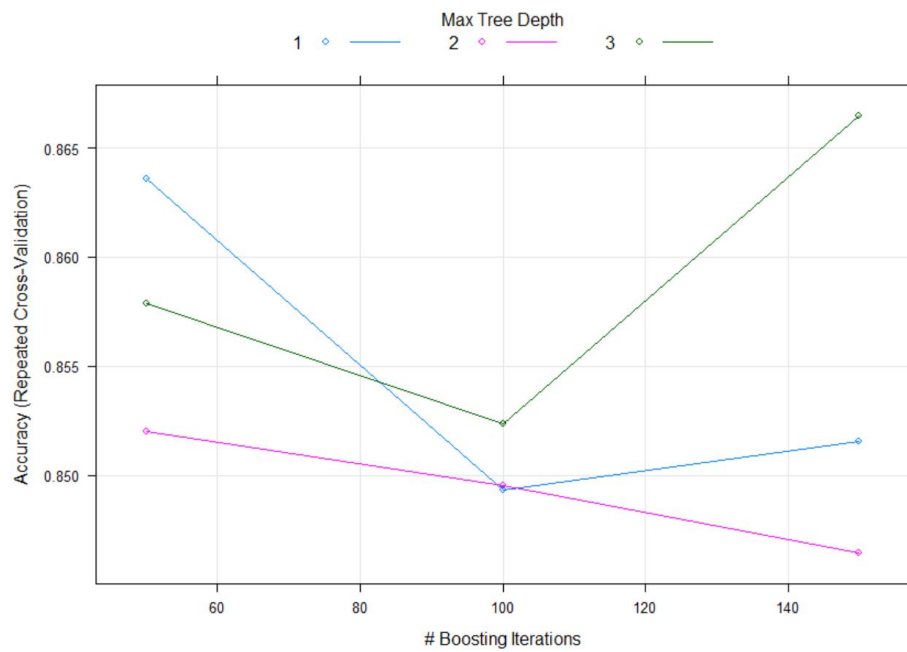


Figure 15: Accuratezza del modello con selezione di variabili

La matrice di confondimento è identica per entrambi i modelli.

| Prediction | Reference | | | | | | |
|--------------|-----------|------|-----|------|--------------|--------|---------|
| | Amphibian | Bird | Bug | Fish | Invertebrate | Mammal | Reptile |
| Amphibian | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bird | 0 | 5 | 0 | 0 | 0 | 0 | 0 |
| Bug | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| Fish | 0 | 0 | 0 | 7 | 0 | 0 | 0 |
| Invertebrate | 0 | 0 | 2 | 0 | 1 | 0 | 0 |
| Mammal | 0 | 0 | 0 | 0 | 0 | 14 | 0 |
| Reptile | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Figure 16: Matrice di confondimento

3 Conclusioni

Questa ultima sezione riassume le prestazioni dei modelli attraverso l'errore di missclassificazione sul test set.

| Classificatore | Accuratezza |
|---|-------------|
| CART (entrambi) | 0.81 |
| Random Forest (senza selezione di variabili) | 0.90 |
| Random Forest (con selezione di variabili) | 0.93 |
| Stochastic GB (entrambi) | 0.93 |

Sulla base dei test, il modello migliore è la random forest con la selezione di variabili, in quanto meno costoso in termini di parametri e di covariate, ma è più laborioso da ottenere. Mentre lo stochastic gradient boosting dà ottimi risultati, e non prevede analisi preliminari.

A mio avviso il metodo migliore è lo stochastic gradient boosting in quanto è più robusto in tutte e due le sue versioni e meno sensibile alla conformazione del test set.

References

- [1] Johnson, R.A., & Wichern, D.W., *Applied multivariate statistical analysis*, Pearson 2014
- [2] T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997
- [3] T. Hastie, R. Tibshirani, & J. Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction. 2nd edition*, Springer, 2009
- [4] Frank, Eibe, et al. *Using model trees for classification*, *Machine Learning* 32.1 1998: 63-76
- [5] Frank, E., Wang, Y., Inglis, S., Holmes, G., & Witten, I. H. *Using model trees for classification. Machine Learning*, 32(1) 1998: 63-76
- [6] FRANK, Eibe, et al. *Using model trees for classification. Machine Learning*, 1998, 32.1: 63-76
- [7] Ho, Tin Kam *Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC*, 14-16 August 1995. pp. 278-282
- [8] Ho, Tin Kam "The Random Subspace Method for Constructing Decision Forests"; *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 20 (8): 832-844
- [9] Friedman, J. H. , *Greedy Function Approximation: A Gradient Boosting Machine.*, 1999
- [10] Friedman, J. H., *Stochastic Gradient Boosting* , 1999
- [11] Mason, L.; Baxter, J.; Bartlett, P. L.; Frean, Marcus "Boosting Algorithms as Gradient Descent"; In *S.A. Solla and T.K. Leen and K. Müller.*, *Advances in Neural Information Processing Systems 12*. MIT Press. pp. 512-518.
- [12] Hastie, T.; Tibshirani, R.; Friedman, J. H. *The Elements of Statistical Learning (2nd ed.)*, New York: Springer, 2009.