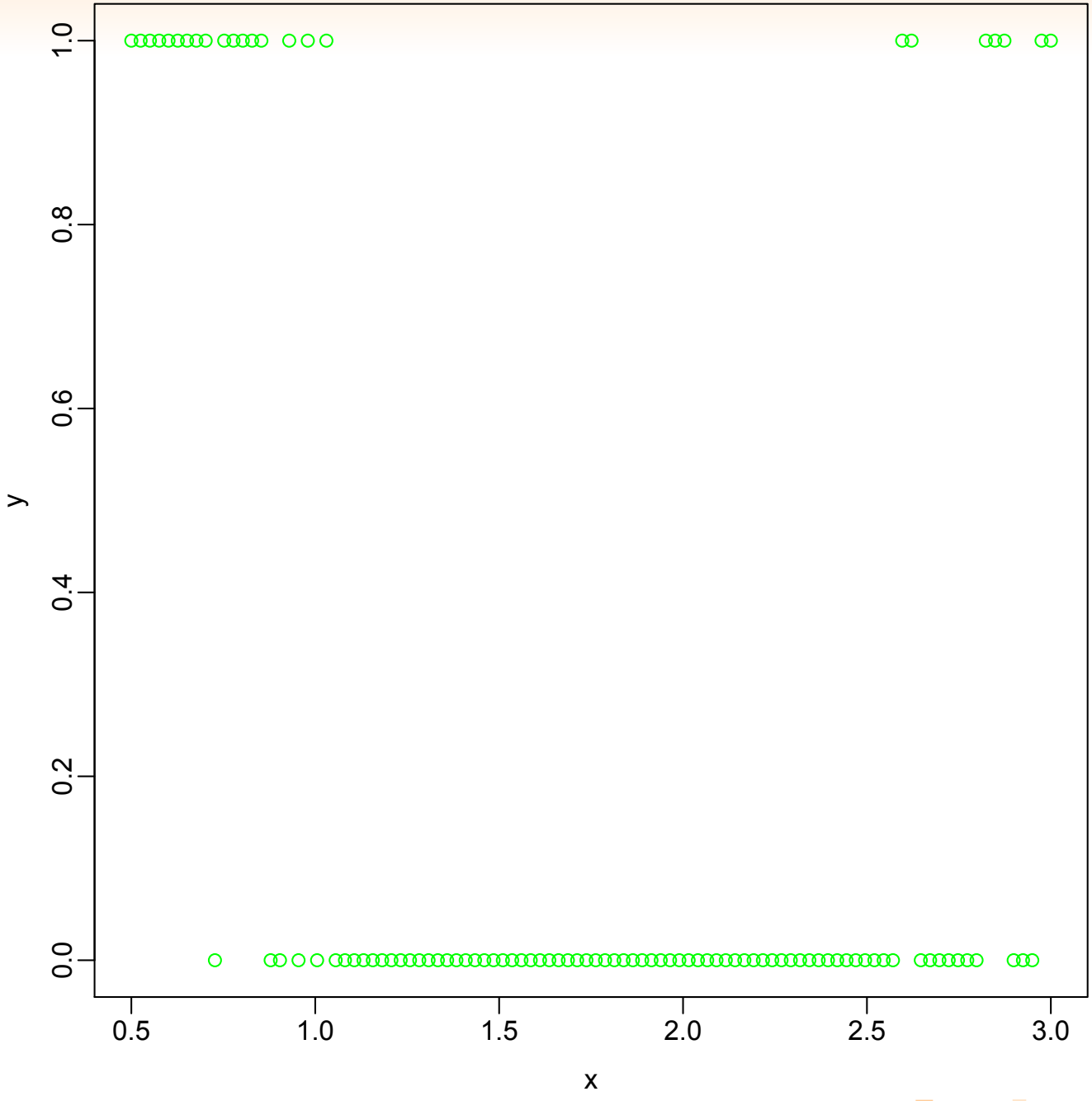
Alberi di classificazione

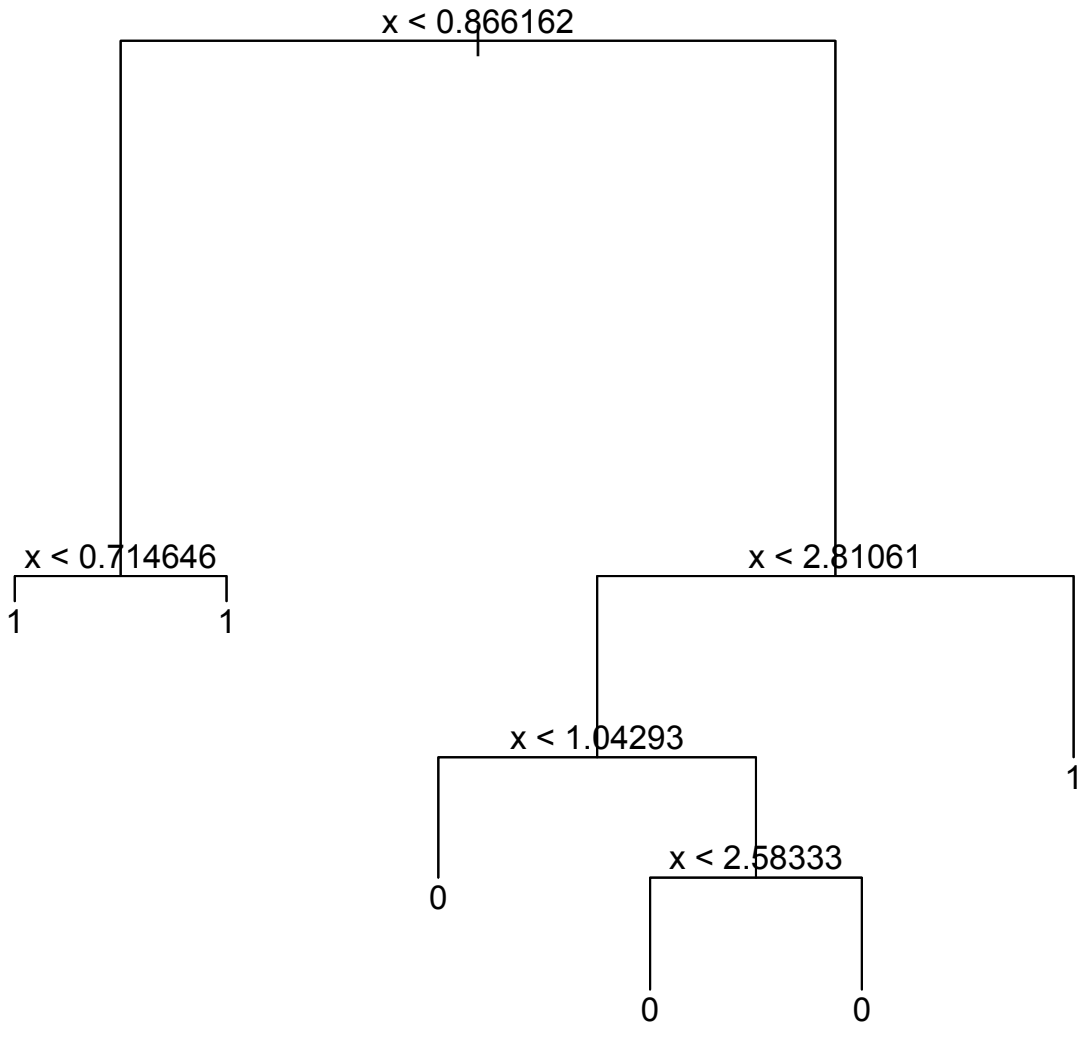

1.0
0.8
0.6
0.4
0.2
0.0
y
0.5
1.0
1.5
2.0
2.5
3.0
x

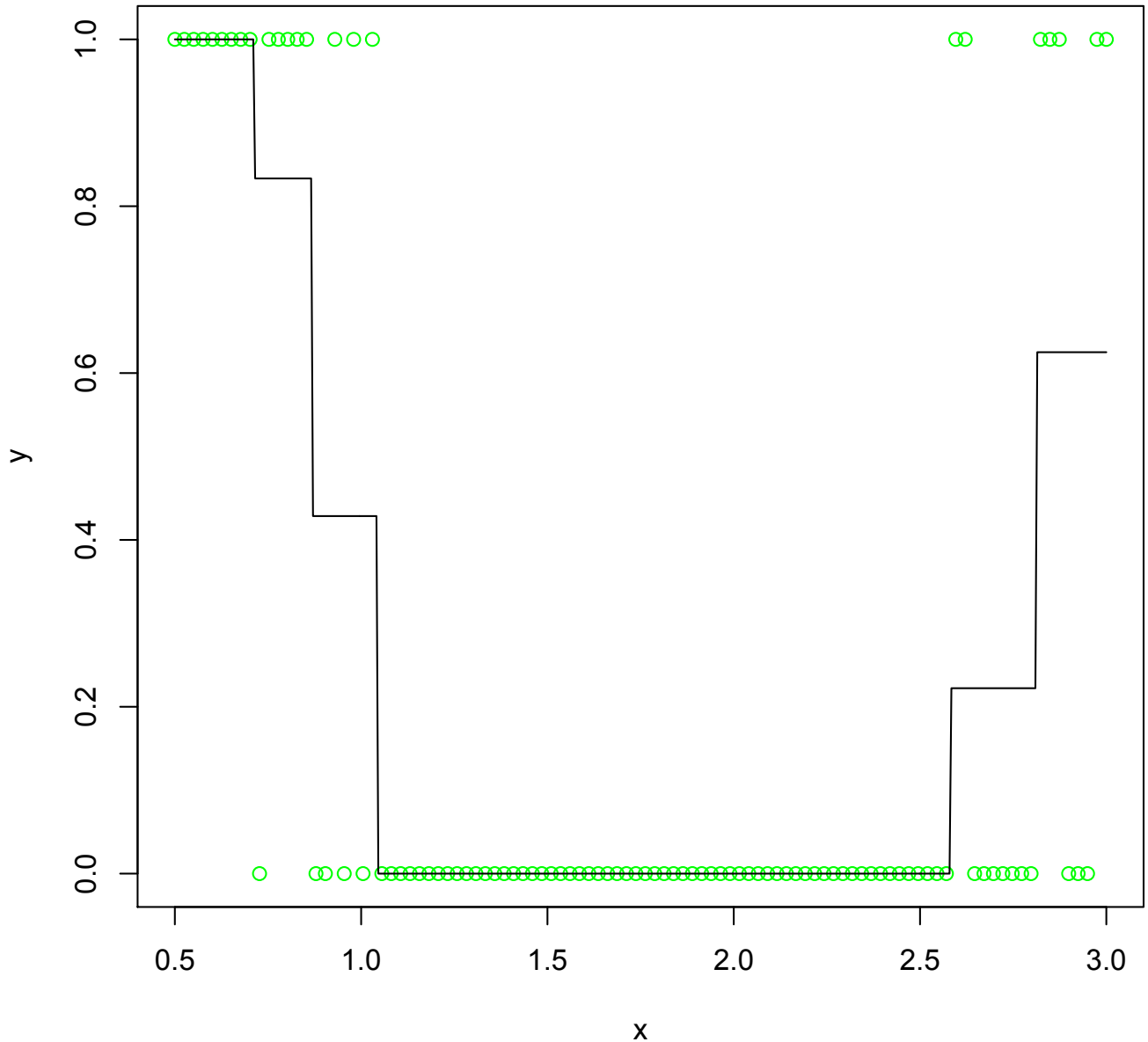
Se si indicano le due classi con 0 e 1 si potrá scrivere:

$$p(x) = P(y = 1|x)$$

$$\hat{p}(x) = \sum_{j=1}^{J} P_j I(x \in R_j)$$

con P_J che rappresenta la probabilitá che $y = 1$ per la regione considerata.


x < 0.866162
x < 0.714646
x < 2.81061
1
1
x < 1.04293
1
x < 2.58333
0
0
0


y
1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x

In questo caso il valore associato alle foglie sará 0 o 1. Quando si fa cadere un'osservazione lungo l'albero e raggiunge una foglia con la probabilitá associata, questa verrá allocata alla classe $C(\hat{p}(x))$ con $C(p) = 0$ se $p \leq 1/2$ e $C(p) = 1$ altrimenti.

Per stimare i P_j si usa la media aritmetica:

$$\hat{P}_j = M(y_i : x_i \ in R_j) = \frac{1}{n_j} \sum_{i \in R_j} I(y_i = 1)$$

che rappresenta la frequenza relativa nella regione considerata.

In questo caso il valore associato alle foglie sará 0 o 1. Quando si fa cadere un'osservazione lungo l'albero e raggiunge una foglia con la probabilitá associata, questa verrá allocata alla classe $C(\hat{p}(x))$ con $C(p) = 0$ se $p \leq 1/2$ e $C(p) = 1$ altrimenti.

Per stimare i P_j si usa la media aritmetica:

$$\hat{P}_j = M(y_i : x_i \ in R_j) = \frac{1}{n_j} \sum_{i \in R_j} I(y_i = 1)$$

che rappresenta la frequenza relativa nella regione considerata.

Ovviamente ora non si considera piú la classica devianza ma si utilizza la devianza connessa alla distribuzione binomiale.

$$D = -2\sum_{i=1}^{n}(y_i log\hat{p}_i + (1-y_i)log(1-\hat{p}_i))$$

o anche

$$D = \sum_{i=1}^{J} -2n_j[\hat{P}_j log\hat{P}_j + (1-\hat{P}_j)log(1-\hat{P}_j)] = \sum_j D_j$$

Si dimostra inoltre che

$$D = 2n\sum_j (n_j/n)Q(\hat{P}_j)$$

che, a meno della costante 2n, é la media delle entropie pesate con la numerositá delle foglie:

$$Q(P_j) = -\sum_{k=0,1} P_{jk} logP_{jk} = -[P_{j1} logP_{j1} + P_{j0} logP_{j0}$$

$$= -[P_{j1} logP_{j1} + (1-P_{j1})log(1-P_{j1})]$$

I termini Q rappresentano una misura di impurtiá, cioé un indicatore del fatto che gli elementi di una certa foglia sono disomogenei rispetto alla variabile di risposta. $Q(p) = 0$ infatti se $p = 0$ o $p = 1$, mentre aumenta quanto piú ci sia avvicina al valore $1/2$.

La misura di entropia si puó quindi sostituire con altre misure di impuritá, ad esempio con l'indice di Gini.

$$Q(P_j) = \sum_{k=0,1} P_{jk}(1 - P_{jk})$$

Con questo adattamento, l'algoritmo di crescita é analogo a quello degli alberi di regressione. Lo stesso vale per la procedura di potatura. In alcuni casi si utilizza anche il semplice tasso di errata classificazione come misura di discrepanza.

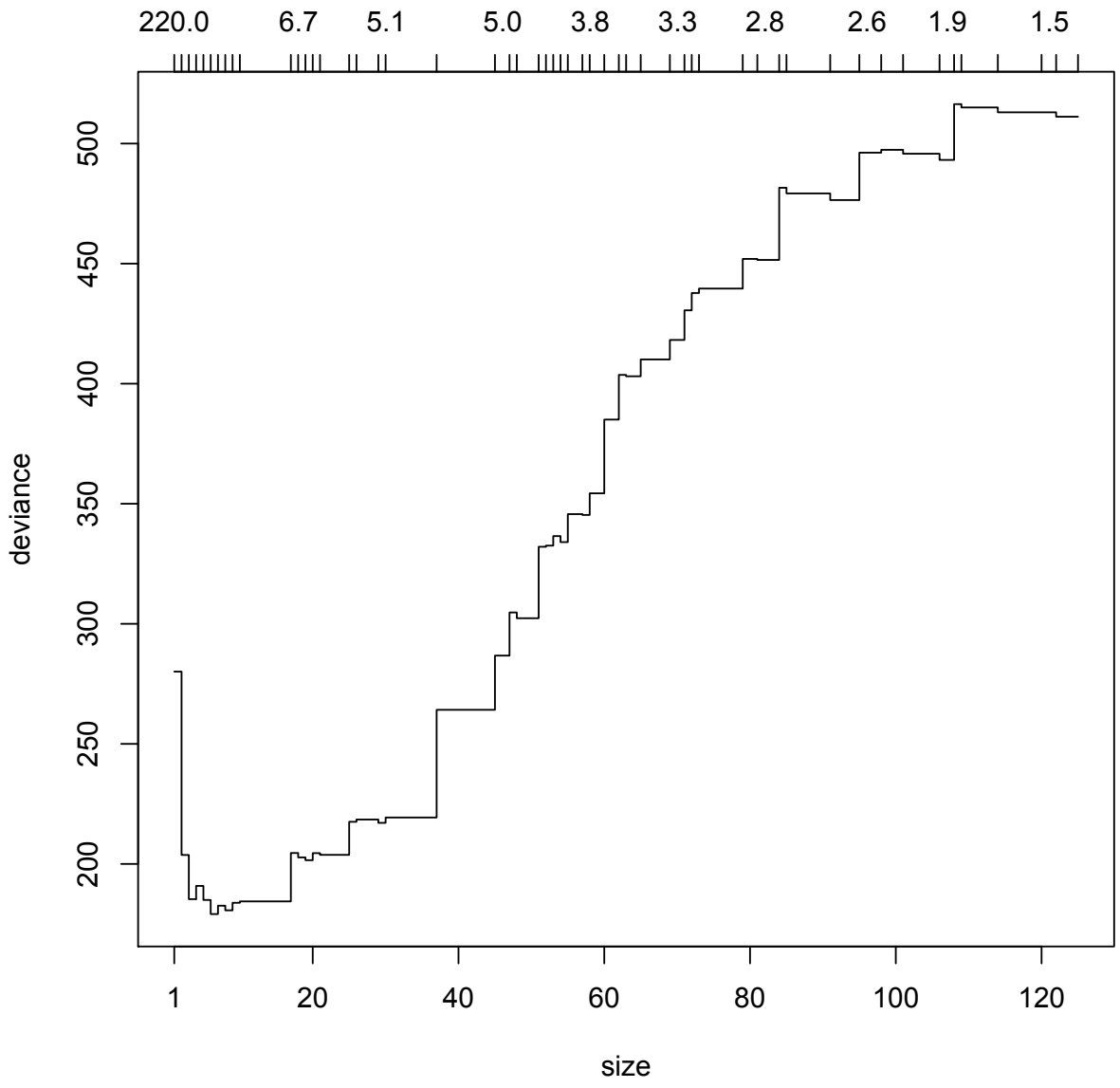
I termini Q rappresentano una misura di impurtiá, cioé un indicatore del fatto che gli elementi di una certa foglia sono disomogenei rispetto alla variabile di risposta. $Q(p) = 0$ infatti se $p = 0$ o $p = 1$, mentre aumenta quanto piú ci sia avvicina al valore $1/2$.

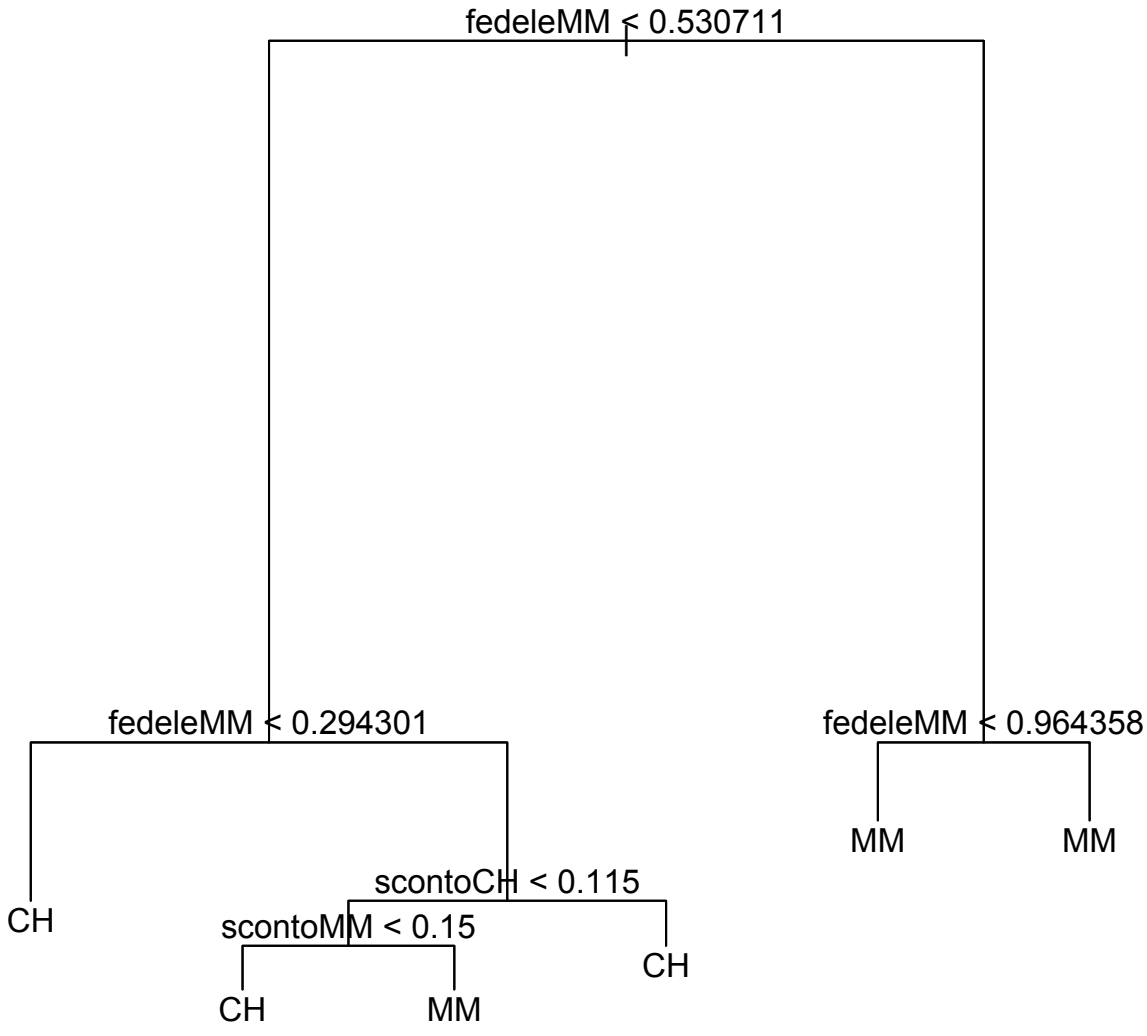
La misura di entropia si puó quindi sostituire con altre misure di impuritá, ad esempio con l'indice di Gini.

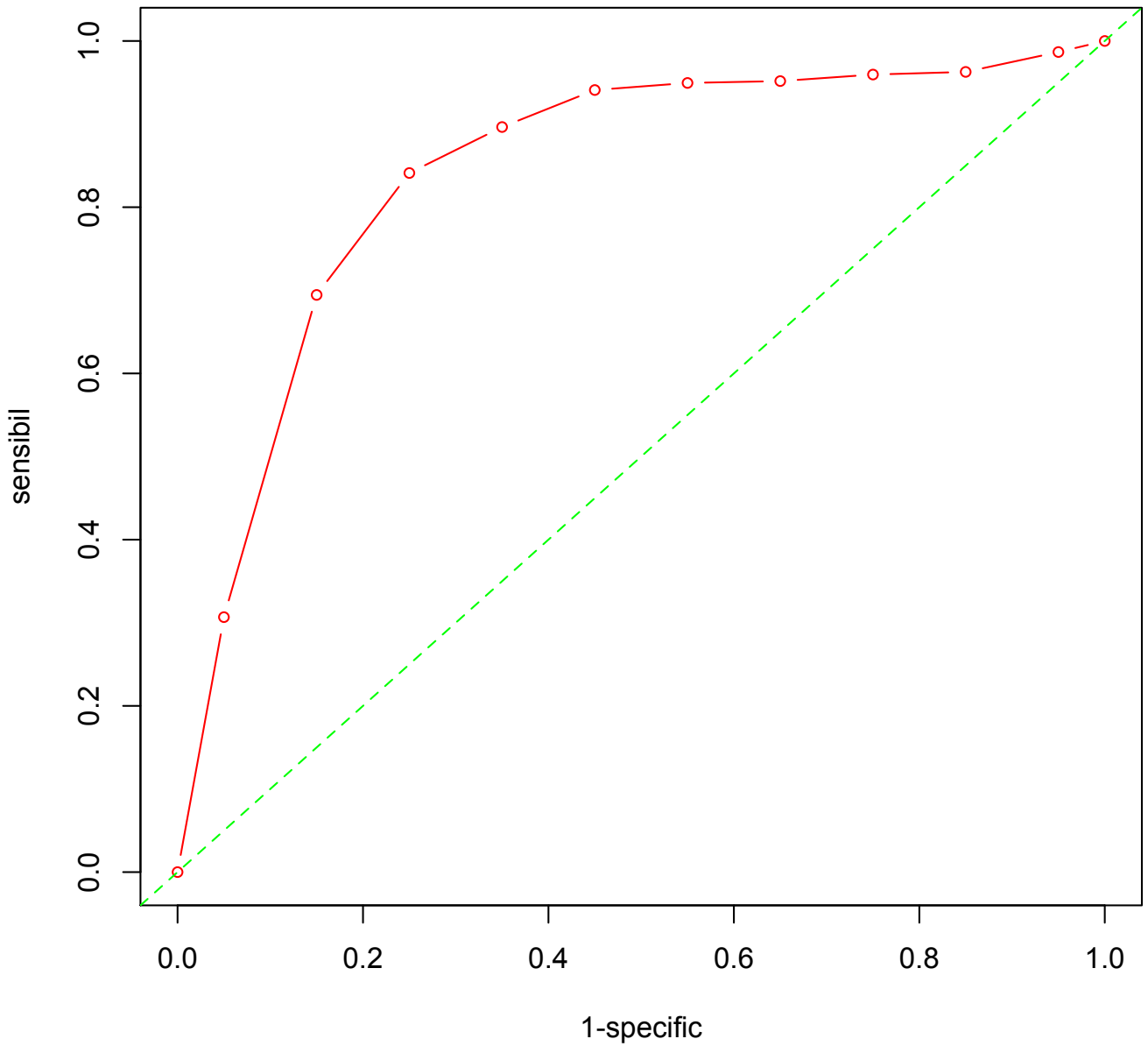
$$Q(P_j) = \sum_{k=0,1} P_{jk}(1 - P_{jk})$$

Con questo adattamento, l'algoritmo di crescita é analogo a quello degli alberi di regressione. Lo stesso vale per la procedura di potatura. In alcuni casi si utilizza anche il semplice tasso di errata classificazione come misura di discrepanza.

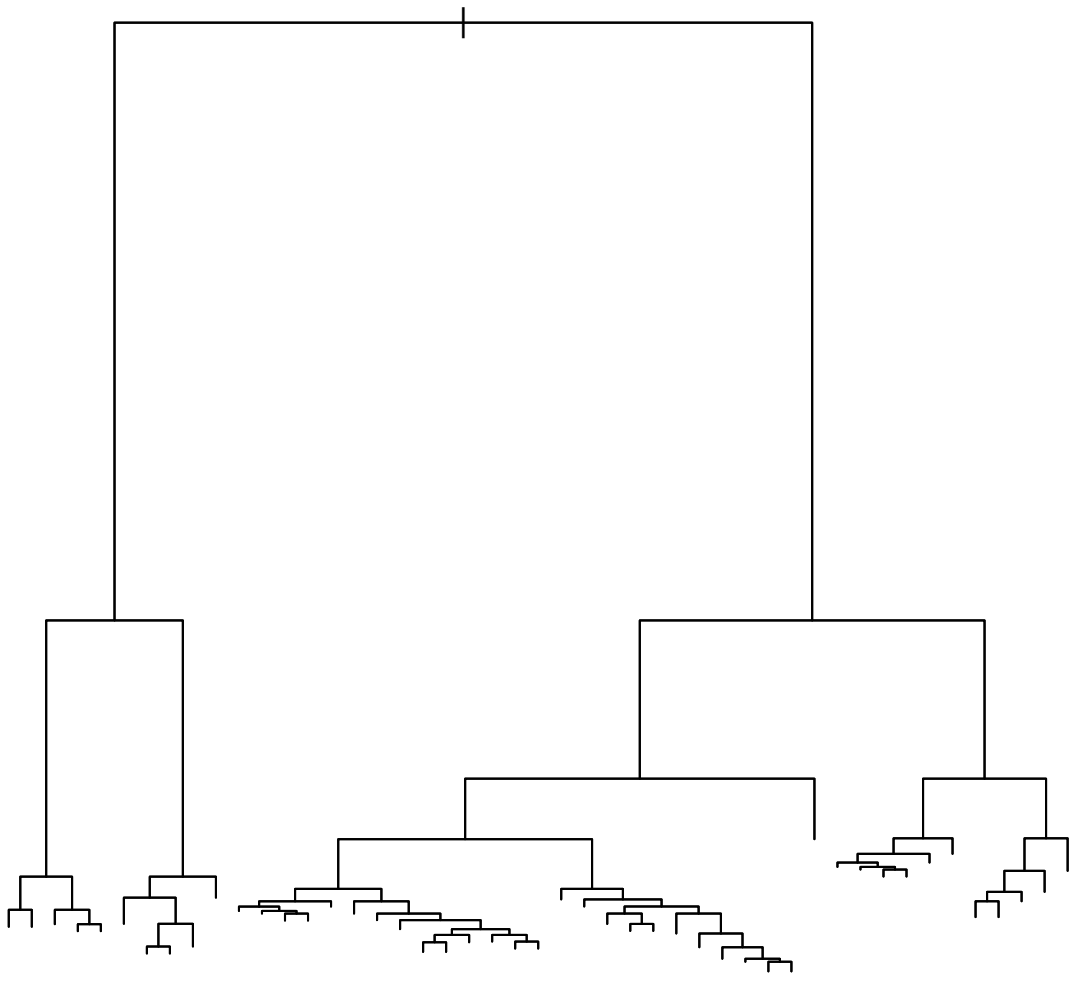
Succhi di frutta: 600 osservazioni per la stima e 202 per la potatura.

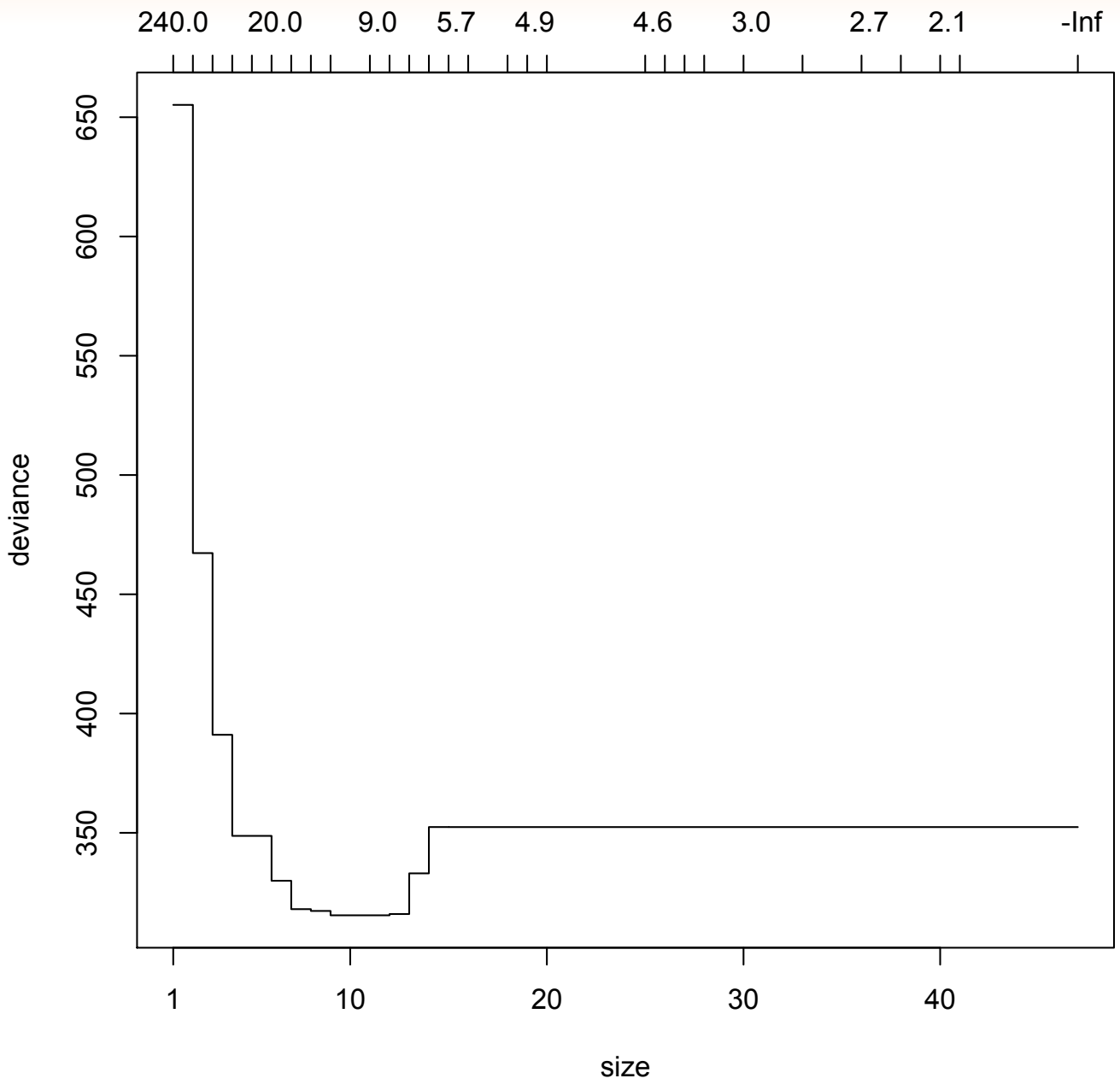


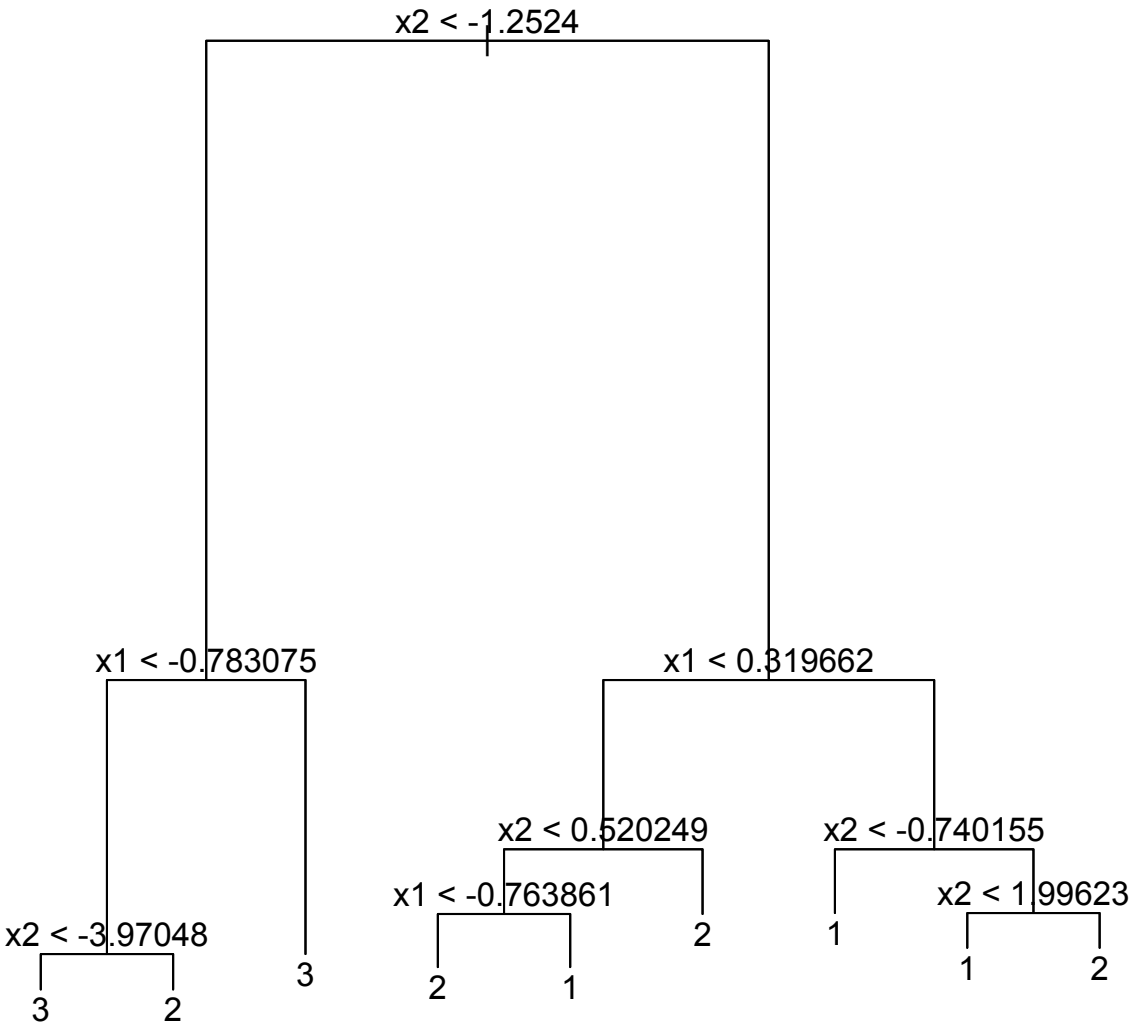

fedeleMM < 0.530711
fedeleMM < 0.294301
fedeleMM < 0.964358
CH
scontoCH < 0.115
scontoMM < 0.15
CH
CH
MM
MM
MM

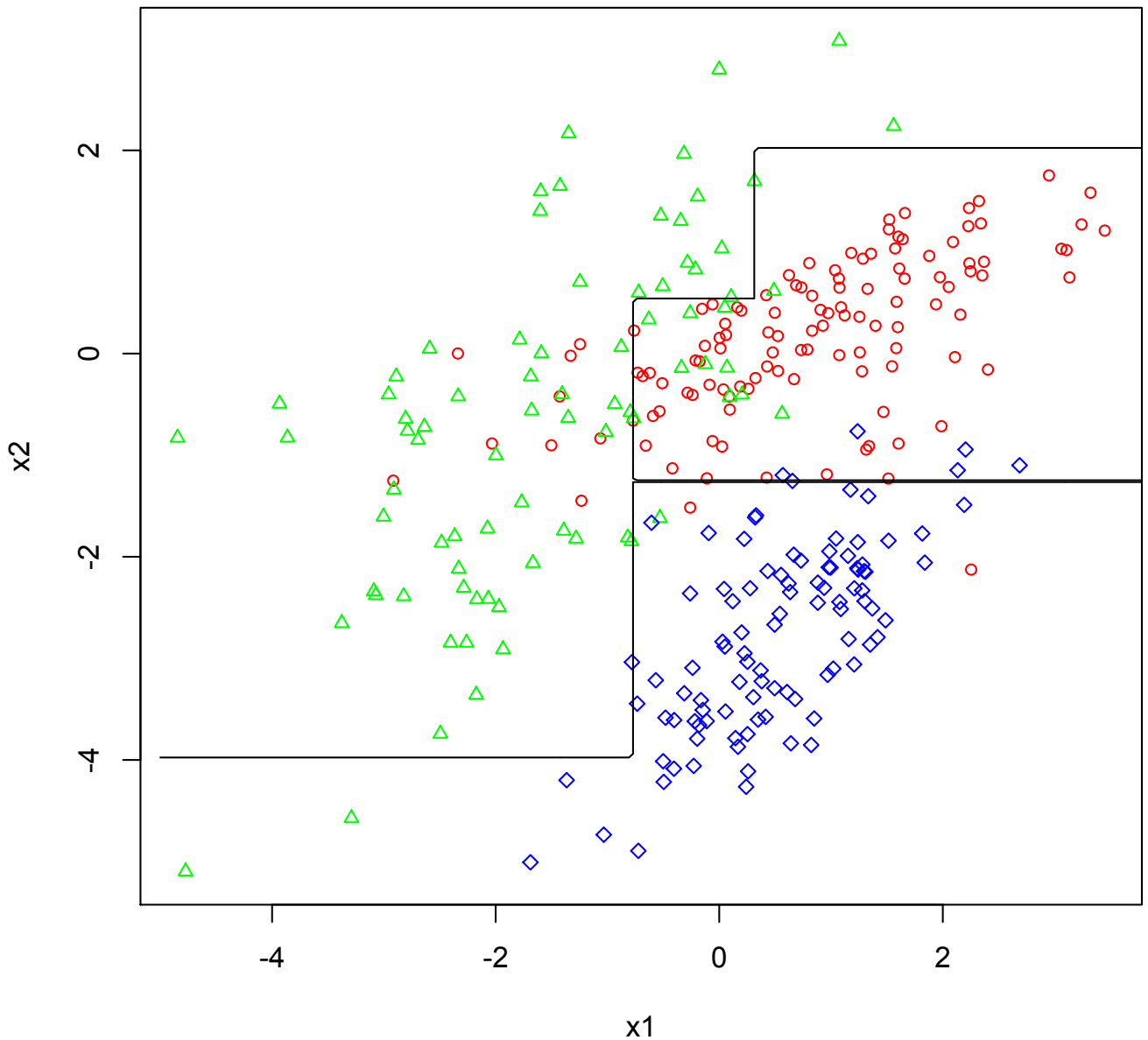

sensibil
1-specific
0.0
0.2
0.4
0.6
0.8
1.0

Se K é maggiore di 2 bisogna considerare degli adattamenti, ma il meccanismo sostanzialmente non cambia:




240.0
20.0
9.0
5.7
4.9
4.6
3.0
2.7
2.1
-Inf
deviance
650
600
550
500
450
400
350
1
10
20
30
40
size


x2 < -1.2524
x1 < -0.783075
x1 < 0.319662
x2 < -3.97048
3
x2 < 0.520249
x2 < -0.740155
x1 < -0.763861
2
1
x2 < 1.99623
3
2
2
1
1
1
2

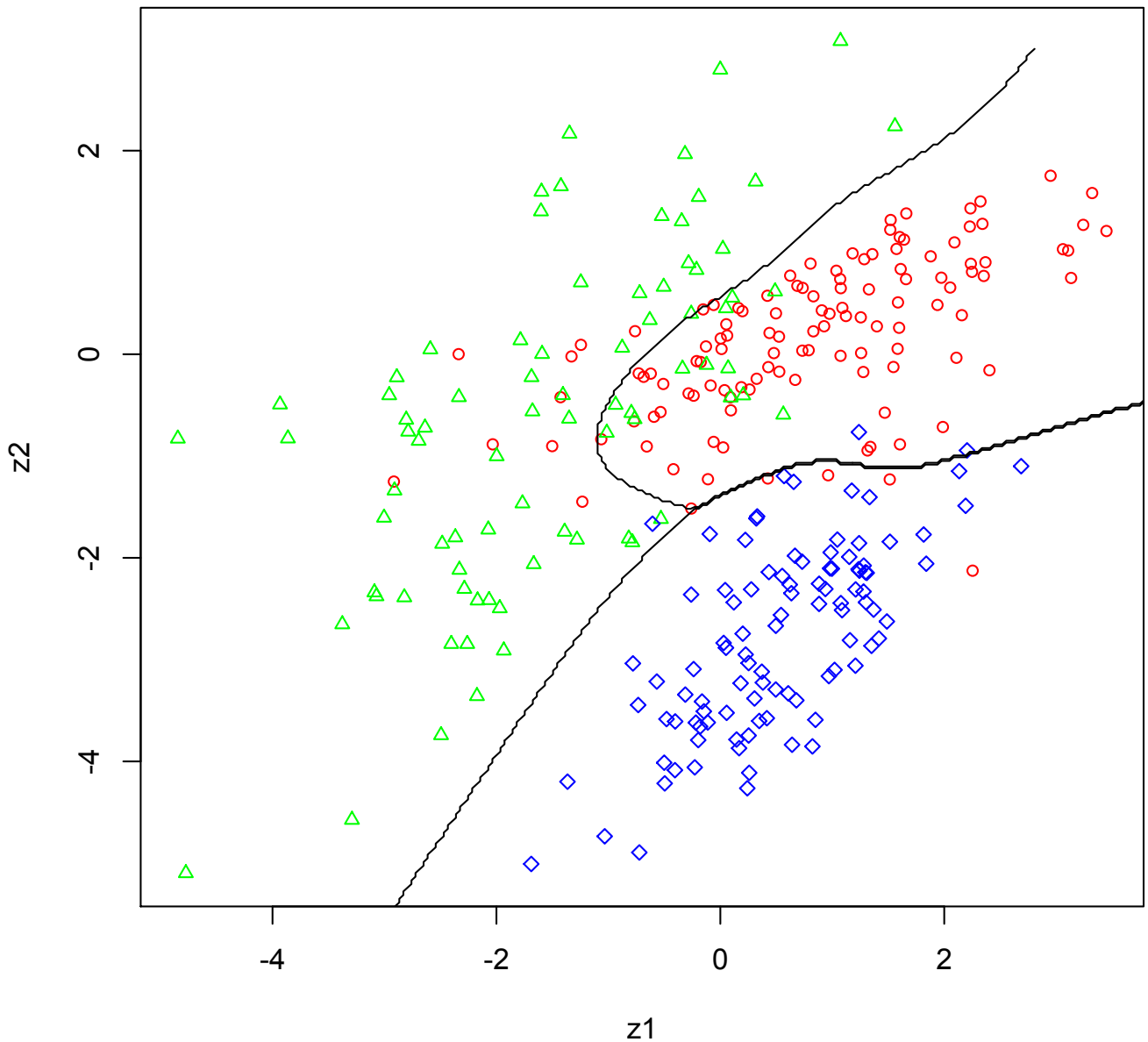

x2
2
0
-2
-4
-4
-2
0
2
x1

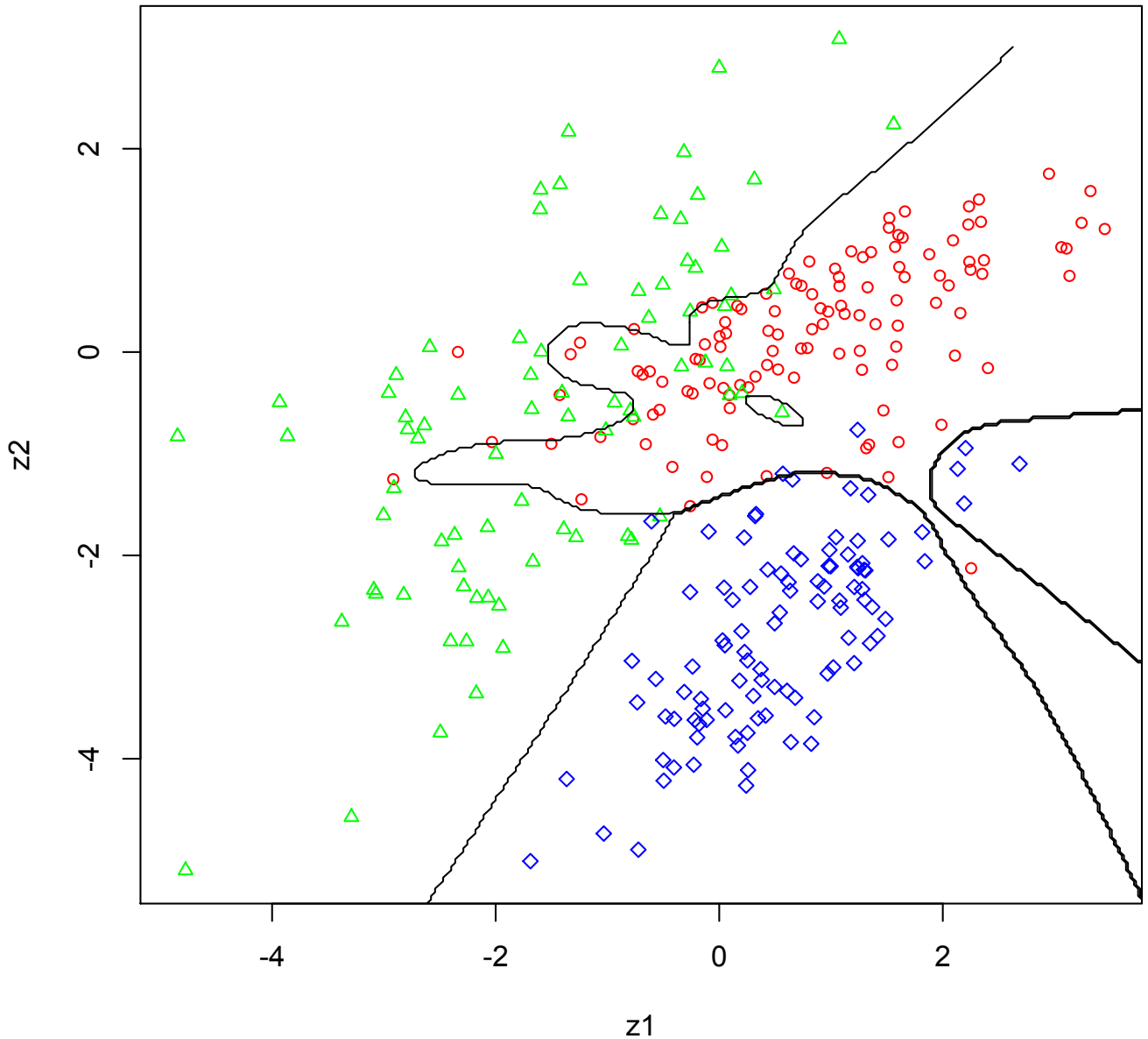
Reti neurali

Anche in questo caso si puó estendere la procedura proposta per le variabili quantitative. L'adattamento da introdurre é nella funzione di attivazione f_1 che deve avere come codominio l'intervallo $0, 1$. Per tali motivi in genere la piú utilizzata é la funzione logistica. In alcuni casi si possono configurare le due classi con -1 e 1 e la funzione utilizzata sará:

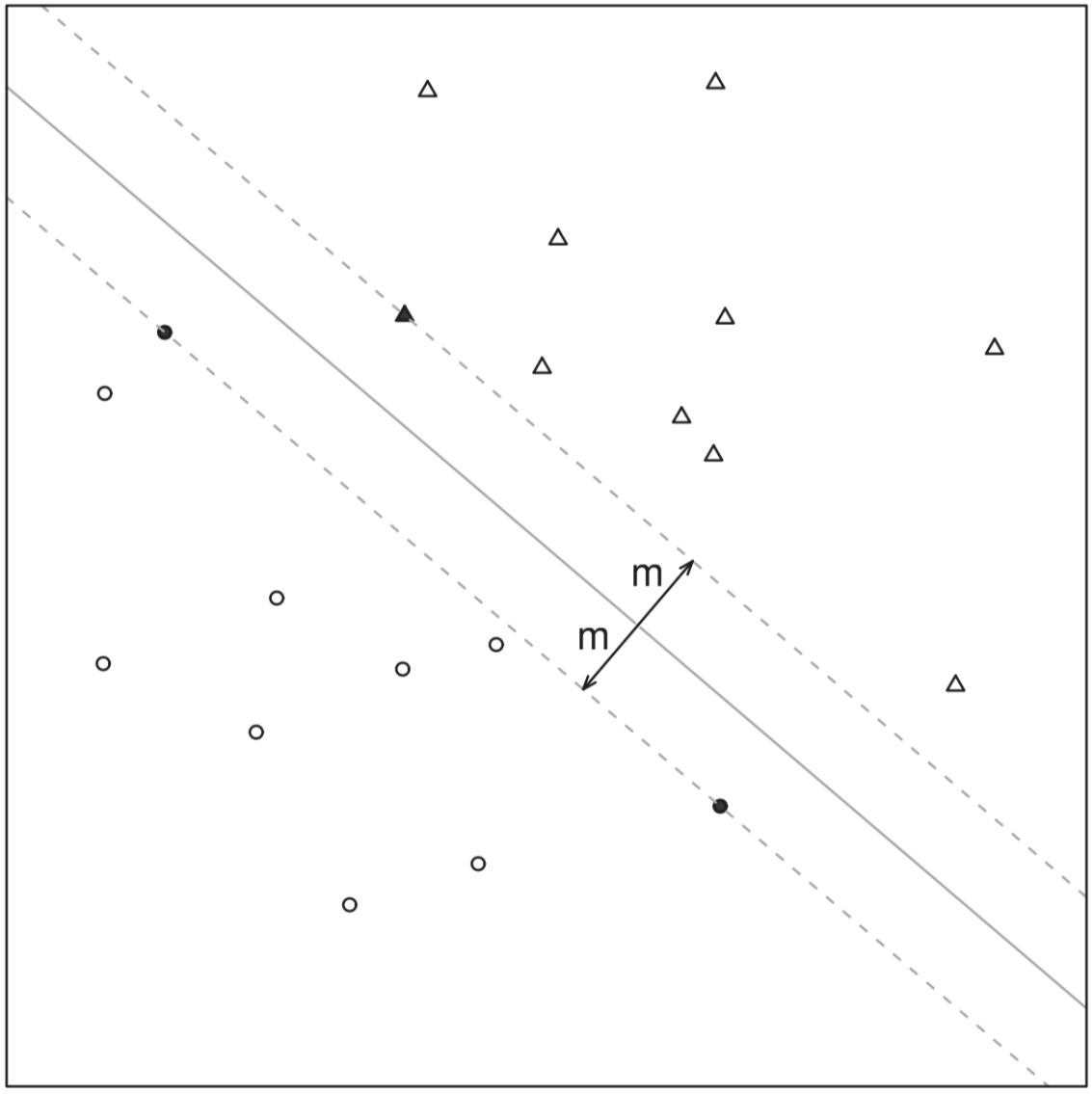
$$2I(x) - 1 = \frac{e^x - 1}{e^x + 1} = tanh(x/2)$$

Se le classi sono piú di due si creano K variabili risposta di tipo $0 - 1$. Inoltre il termine che entra nella funzione obiettivo non é piú la distanza euclidea, ma l'entropia.


z2
2
0
-2
-4
-4
-2
0
2
z1


z2
2
0
-2
-4
-4
-2
0
2
z1

Support vector machines



L'equazione

$$\beta_0 + x'\beta = 0$$

individua un generico iperpiano candidato a separare le due classi, alle quali vengono convenzionalmente assegnati i valori -1 e 1.
L'iperpiano ottimale sará ovviamente quello che massimizza la distanza m.
Il problema di ottimizzazione puó essere formalizzato come segue:

$$\max_{\beta_0, \beta} m$$

con i seguenti vincoli:

$$||\beta|| = 1$$

$$y_i(\beta_0 + \tilde{x}_i'\beta) \geq \ m \quad (i = 1, \ldots, n)$$

dove $(\tilde{x}, y)$ rappresenta una generica unitá da classificare.

Per semplificare la condizione $||\beta = 1||$ i vincoli vengono riscritti nella seguente maniera

$$y_i(\beta_0 + \tilde{x}_i'\beta) \geq m||\beta||$$

e il problema diventa:

$$\min_{\beta_0,\beta} \frac{1}{2}||\beta||^2$$

con vincoli

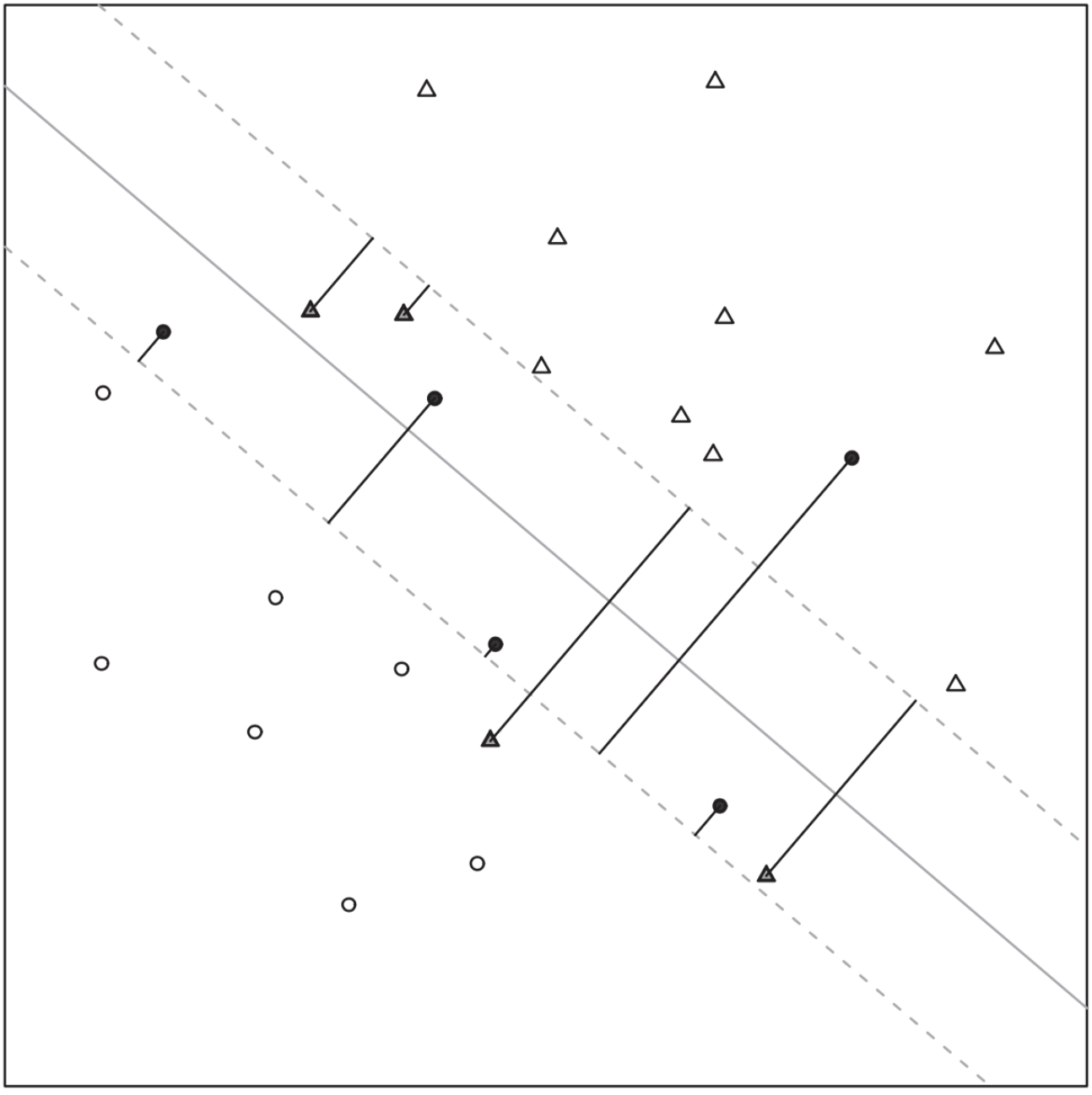
$$y_i(\beta_0 + \tilde{x}_i'\beta) \geq 1$$

Per la soluzione del problema si utilizzano elementi di geometria, ma ovviamente, in situazioni reali, é impossibile trovare una retta che separi perfettamente le due classi. É necessario quindi introdurre un requisito meno stringente, tollerando che alcuni punti saranno comunque mal classificati.

A tal fine vengono introdotte delle variabili ausiliarie a componenti non negative $\xi = (\xi_1, \ldots, \xi_n)$ che esprimono di quanto i vari punti stanno al di la' della linea di margine della loro classe.

Per la soluzione del problema si utilizzano elementi di geometria, ma ovviamente, in situazioni reali, é impossibile trovare una retta che separi perfettamente le due classi. É necessario quindi introdurre un requisito meno stringente, tollerando che alcuni punti saranno comunque mal classificati.

A tal fine vengono introdotte delle variabili ausiliarie a componenti non negative $\xi = (\xi_1, \ldots, \xi_n)$ che esprimono di quanto i vari punti stanno al di la' della linea di margine della loro classe.



In questo caso il problema di ottimizzazione viene adattato sostituendo i vincoli precedenti con la seguente forma

$$y_i(\beta_0 + \tilde{x}_i'\beta) \geq m(1 - \xi)$$

e l'ottimizzazione diventa:

$$\min_{\beta_0,\beta} \frac{1}{2}||\beta||^2 + \gamma \sum_{i=1}^{n} \xi_i$$

con vincoli

$$y_i(\beta_0 + \tilde{x}_i'\beta) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

con γ costante positiva che é un parametro di regolazione e rappresenta il costo di violazione delle barriere.

Si dimostra che:

$$\hat{\beta} = \sum_{i=1}^{n} a_i y_i \tilde{x}_i$$

dove solo alcuni a_i sono non nulli, quindi la soluzione é esprimibile solo con alcune osservazioni, dette support vectors.

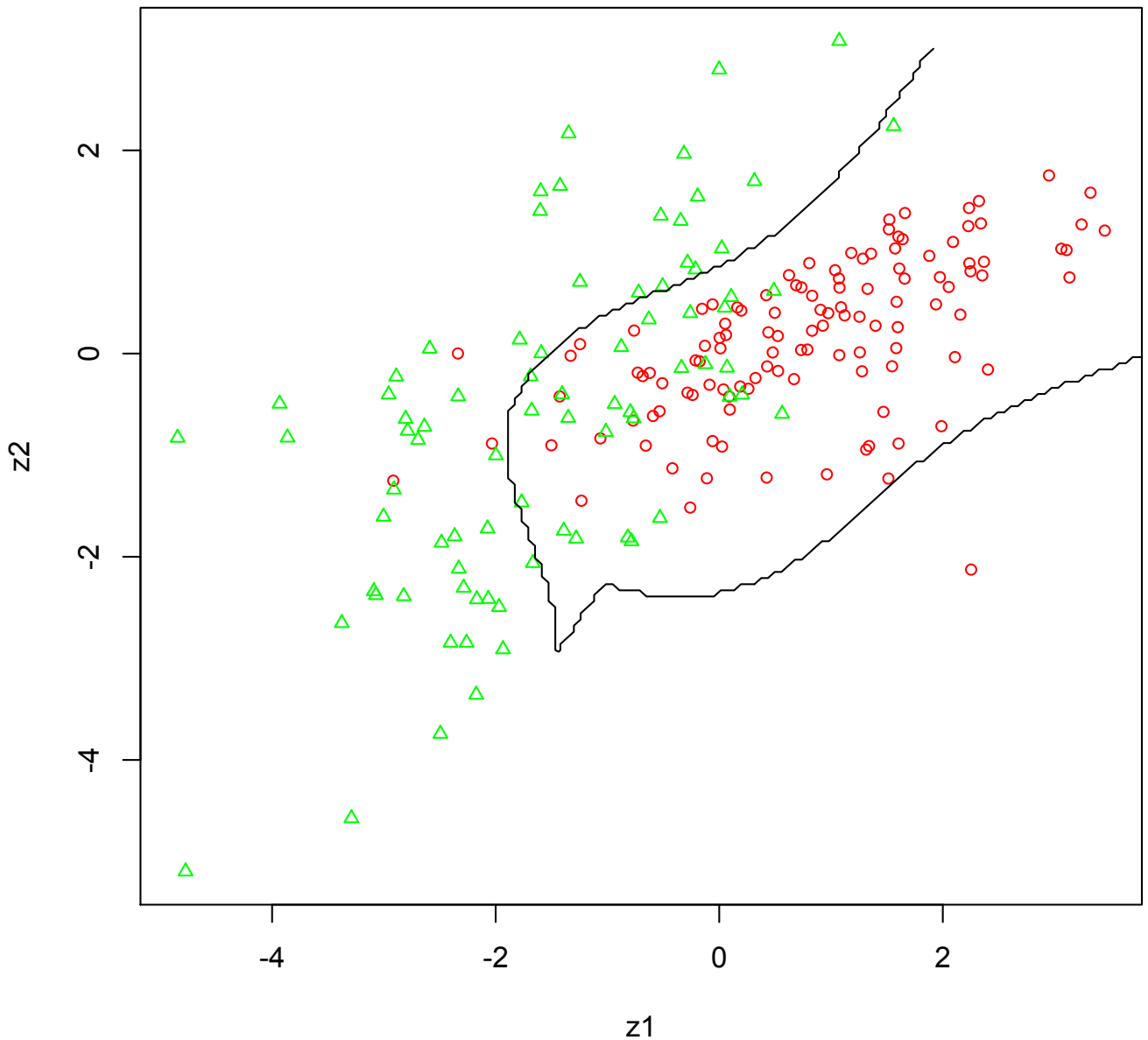
In alcuni casi é opportuno considerare delle trasformazioni delle variabili esplicative, sfruttando diverse funzioni nucleo che portano a risultati differenti.

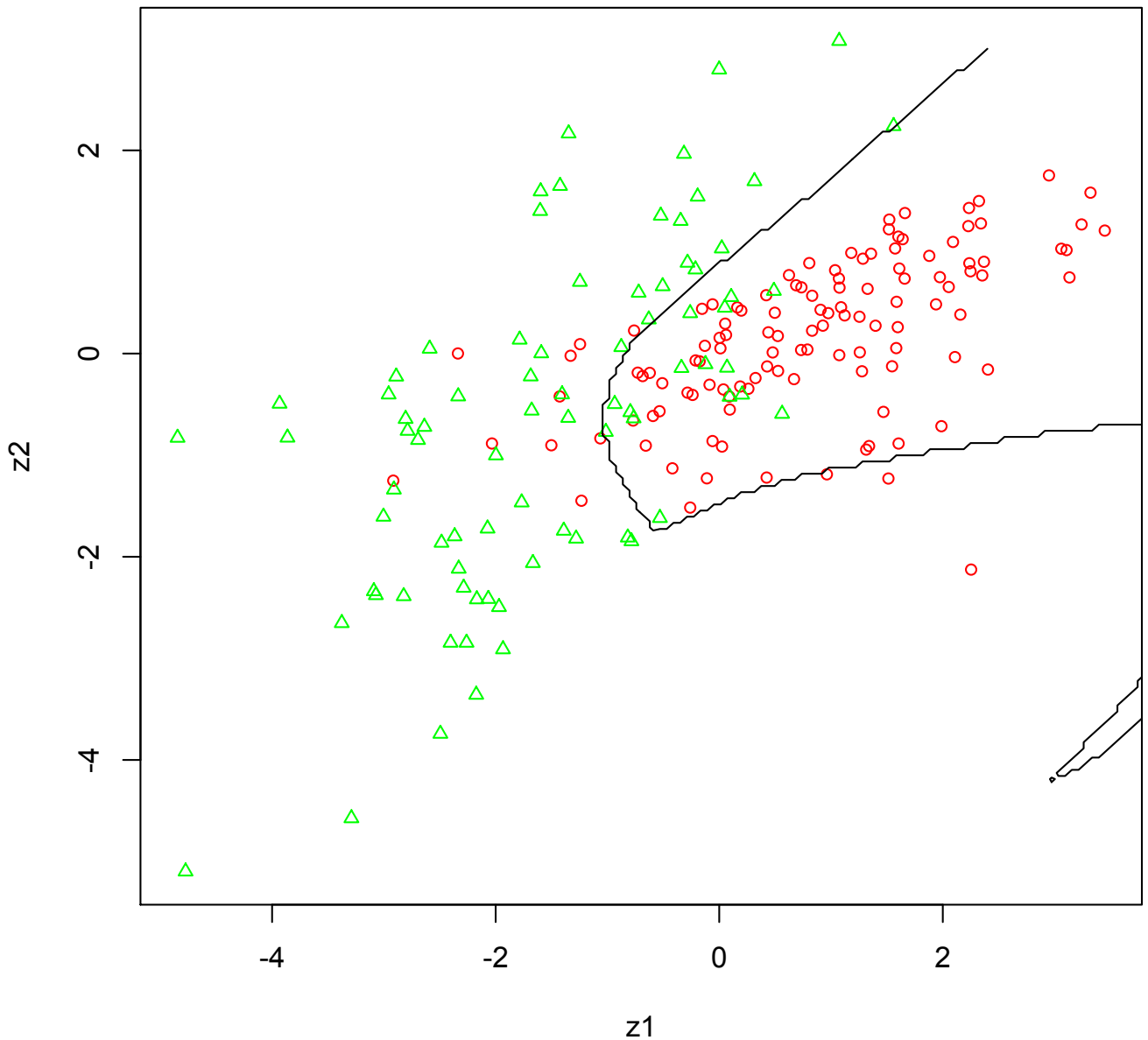
Si dimostra che:

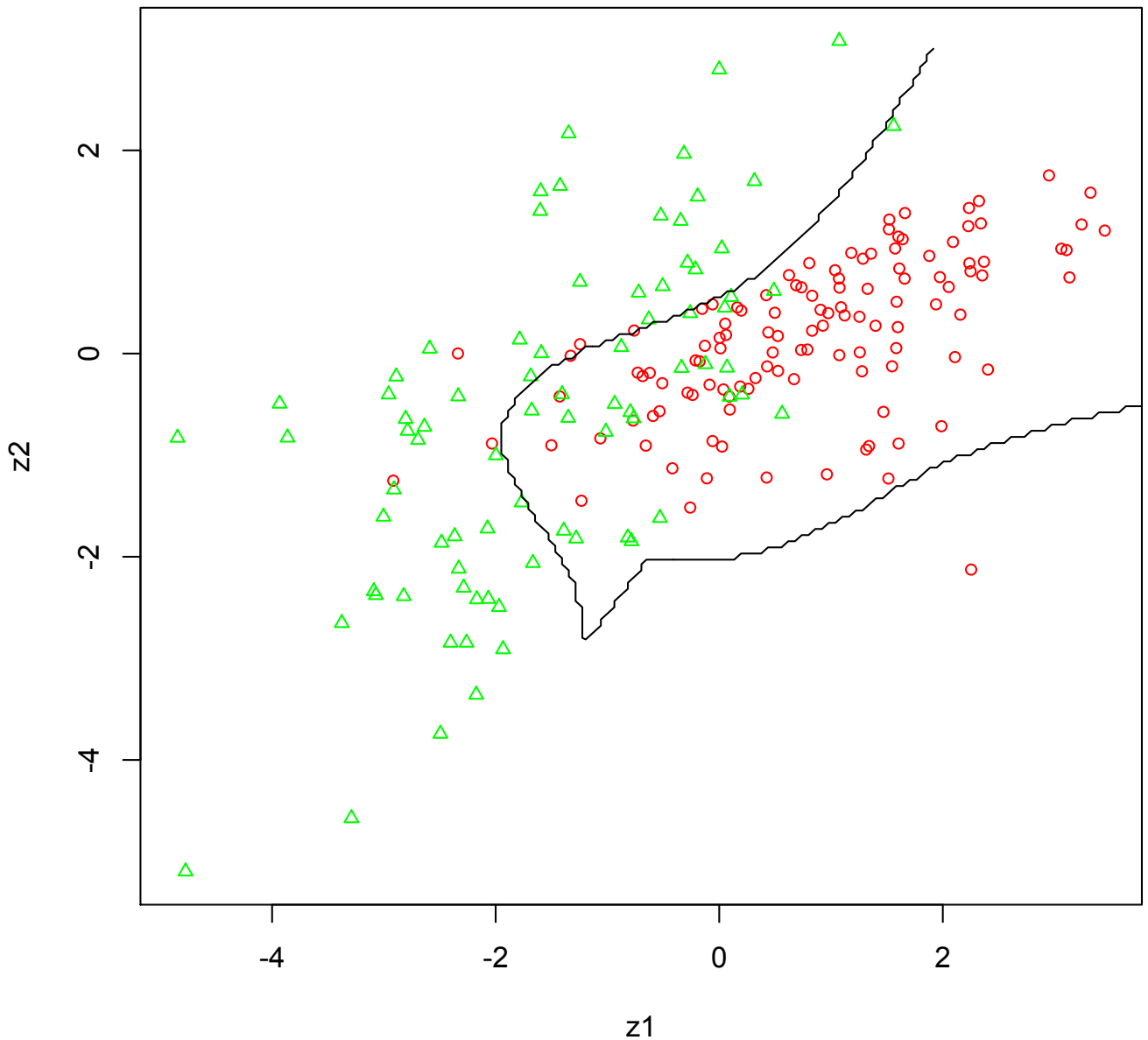
$$\hat{\beta} = \sum_{i=1}^{n} a_i y_i \tilde{x}_i$$

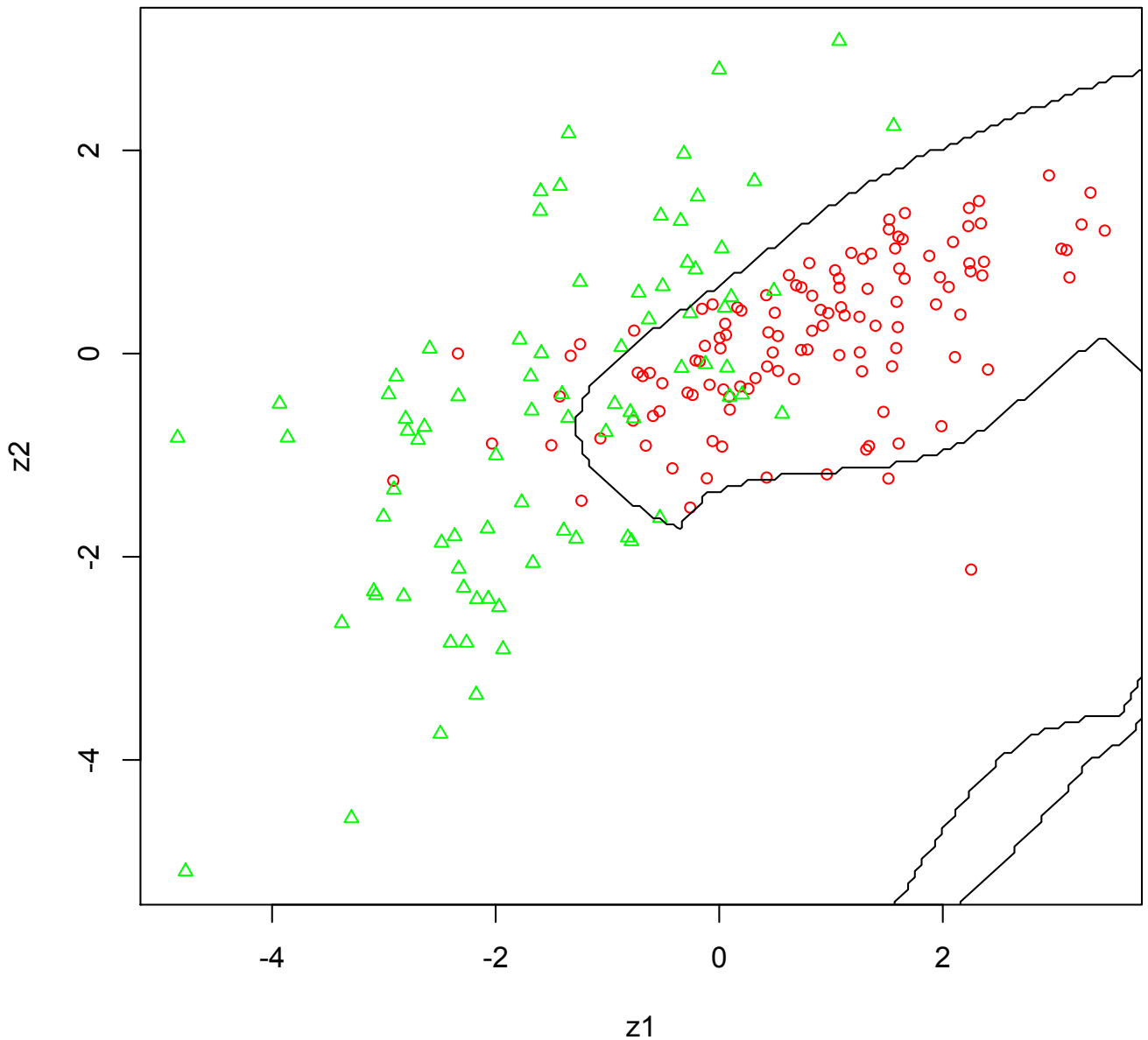
dove solo alcuni a_i sono non nulli, quindi la soluzione é esprimibile solo con alcune osservazioni, dette support vectors.

In alcuni casi é opportuno considerare delle trasformazioni delle variabili esplicative, sfruttando diverse funzioni nucleo che portano a risultati differenti.


z2
2
0
-2
-4
-4
-2
0
2
z1


z2
2
0
-2
-4
-4
-2
0
2
z1


z2
2
0
-2
-4
-4
-2
0
2
z1


z2
2
0
-2
-4
-4
-2
0
2
z1

Combinazioni di classificatori

In molte situazioni diversi modelli sembrano adattarsi bene ai dati e non sembra esserci uno preferito agli altri.
Ad esempio, se abbiamo 30 variabili esplicative e costruiamo un modello di regressione logistica con 5 variabili, ci saranno circa 140mila possibili gruppi che si possono formare.

Andando a calcolare l'errore di previsione su un insieme di verifica, sará di poco diverso considerando tutti i modelli costruiti. Quindi i modelli saranno praticamente equivalenti per effettuare la previsione, ma ognuno ci spiegherá in maniera diversa il problema, a seconda del tipo di variabili che entreranno in gioco.

Combinazioni di classificatori

In molte situazioni diversi modelli sembrano adattarsi bene ai dati e non sembra esserci uno preferito agli altri.
Ad esempio, se abbiamo 30 variabili esplicative e costruiamo un modello di regressione logistica con 5 variabili, ci saranno circa 140mila possibili gruppi che si possono formare.

Andando a calcolare l'errore di previsione su un insieme di verifica, sará di poco diverso considerando tutti i modelli costruiti. Quindi i modelli saranno praticamente equivalenti per effettuare la previsione, ma ognuno ci spiegherá in maniera diversa il problema, a seconda del tipo di variabili che entreranno in gioco.

Se invece utilizziamo per esempio le reti neurali o gli alberi di classificazione, la stima sará fortemente influenzata dall'insieme di dati usato per la stima. Si é provato per esempio che se si prende un insieme di stima e poi si elimina una piccola percentuale di osservazioni $(2 - 3\%)$ il modello puó cambiare completamente anche se non varia l'errore di previsione.

Per superare tali inconvenienti, una possibilitá é quella di combinare le previsioni ottenute da metodi diversi. Sono state in realtá seguite diverse strade, ma ciascuna cerca di raccogliere le qualitá delle singole componenti portando spesso a previsioni piú accurate.

Se invece utilizziamo per esempio le reti neurali o gli alberi di classificazione, la stima sará fortemente influenzata dall'insieme di dati usato per la stima. Si é provato per esempio che se si prende un insieme di stima e poi si elimina una piccola percentuale di osservazioni $(2 - 3\%)$ il modello puó cambiare completamente anche se non varia l'errore di previsione.

Per superare tali inconvenienti, una possibilitá é quella di combinare le previsioni ottenute da metodi diversi. Sono state in realtá seguite diverse strade, ma ciascuna cerca di raccogliere le qualitá delle singole componenti portando spesso a previsioni piú accurate.

Bagging

- Si indica con $C(x)$ un classificatore determinato con uno dei metodi introdotti finora su un insieme di stima $Z = \{(x_1, y_1), \ldots, (x_n, y_n)\}$
- Si utilizza poi la procedura bootstrap, considerando il campione Z_1^* ottenuto estraendo n volte con ripetizione gli elementi dell'insieme di stima Z.
- Utilizzando questo nuovo campione si ottiene un nuovo classificatore $C_1^*(x)$ che sará probabilmente diverso dal precedente.

- Si ripete la procedura B volte, ottenendo i campioni $Z_b^*(b = 1, \ldots, B)$ e da questa i nuovi classificatori $C_b^*(x)$.
- Si introduce quindi un nuovo classificatore che é una sintesi dei risultati ottenuti sui singoli campioni. In genere si utilizza la media aritmetica e si ha

$$C_{bag}(x) = \frac{1}{B} \sum_{b=1}^{B} C_b^*(x)$$

- Si classifica l'unitá nella classe associata al valore 1 se $C_{bag}(x) > 1/2$ e a quella associata al valore 0 altrimenti.

L'errore di classificazione del nuovo modello risulta mediamente inferiore di quello calcolato sui singoli modelli "originali".
Il modello viene chiamato di Bootstrap AGGregatING, da cui BAGGING.

La stessa procedura si puó utilizzare per variabili quantitative, dove non si avrá piú il classificatore ma quello che viene definito il previsore, cioÃ¨ il valore assunto dalla variabile di risposta. In questo caso il nuovo modello avrá generalmente varianza inferiore rispetto al modello originario.

La combinazione dei risultati comporta peró la perdita di ogni semplice struttura nei modelli originari, da cui deriva la difficoltá di interpretazione dei risultati.

Una procedura alternativa é il bumping che individua come classificatore il modello che ha il minor errore di previsione fra tutti i modelli ottenuti con i ricampionamenti boostrap.

L'utilizzo di campioni casuali permette di utilizzare anche la procedura out-of-bag. La parte di dati che viene esclusa dal campione bootstrap viene praticamente usata come insieme di verifica

La stessa procedura si puó utilizzare per variabili quantitative, dove non si avrá piú il classificatore ma quello che viene definito il previsore, cioÃ¨ il valore assunto dalla variabile di risposta. In questo caso il nuovo modello avrá generalmente varianza inferiore rispetto al modello originario.

La combinazione dei risultati comporta peró la perdita di ogni semplice struttura nei modelli originari, da cui deriva la difficoltá di interpretazione dei risultati.

Una procedura alternativa é il bumping che individua come classificatore il modello che ha il minor errore di previsione fra tutti i modelli ottenuti con i ricampionamenti boostrap.

L'utilizzo di campioni casuali permette di utilizzare anche la procedura out-of-bag. La parte di dati che viene esclusa dal campione bootstrap viene praticamente usata come insieme di verifica

La stessa procedura si puó utilizzare per variabili quantitative, dove non si avrá piú il classificatore ma quello che viene definito il previsore, cioÃ¨ il valore assunto dalla variabile di risposta. In questo caso il nuovo modello avrá generalmente varianza inferiore rispetto al modello originario.

La combinazione dei risultati comporta peró la perdita di ogni semplice struttura nei modelli originari, da cui deriva la difficoltá di interpretazione dei risultati.

Una procedura alternativa é il bumping che individua come classificatore il modello che ha il minor errore di previsione fra tutti i modelli ottenuti con i ricampionamenti boostrap.

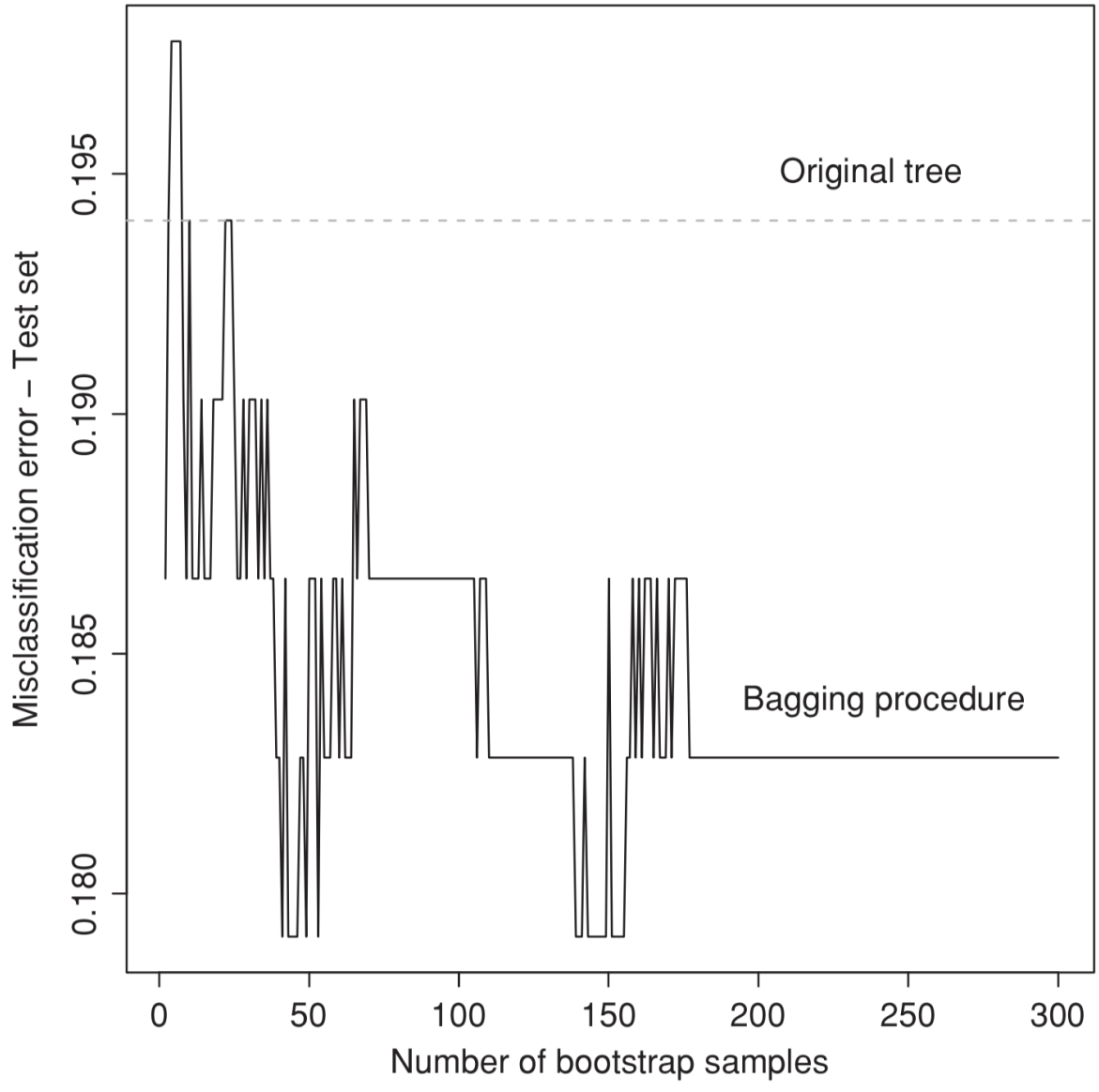
L'utilizzo di campioni casuali permette di utilizzare anche la procedura out-of-bag. La parte di dati che viene esclusa dal campione bootstrap viene praticamente usata come insieme di verifica

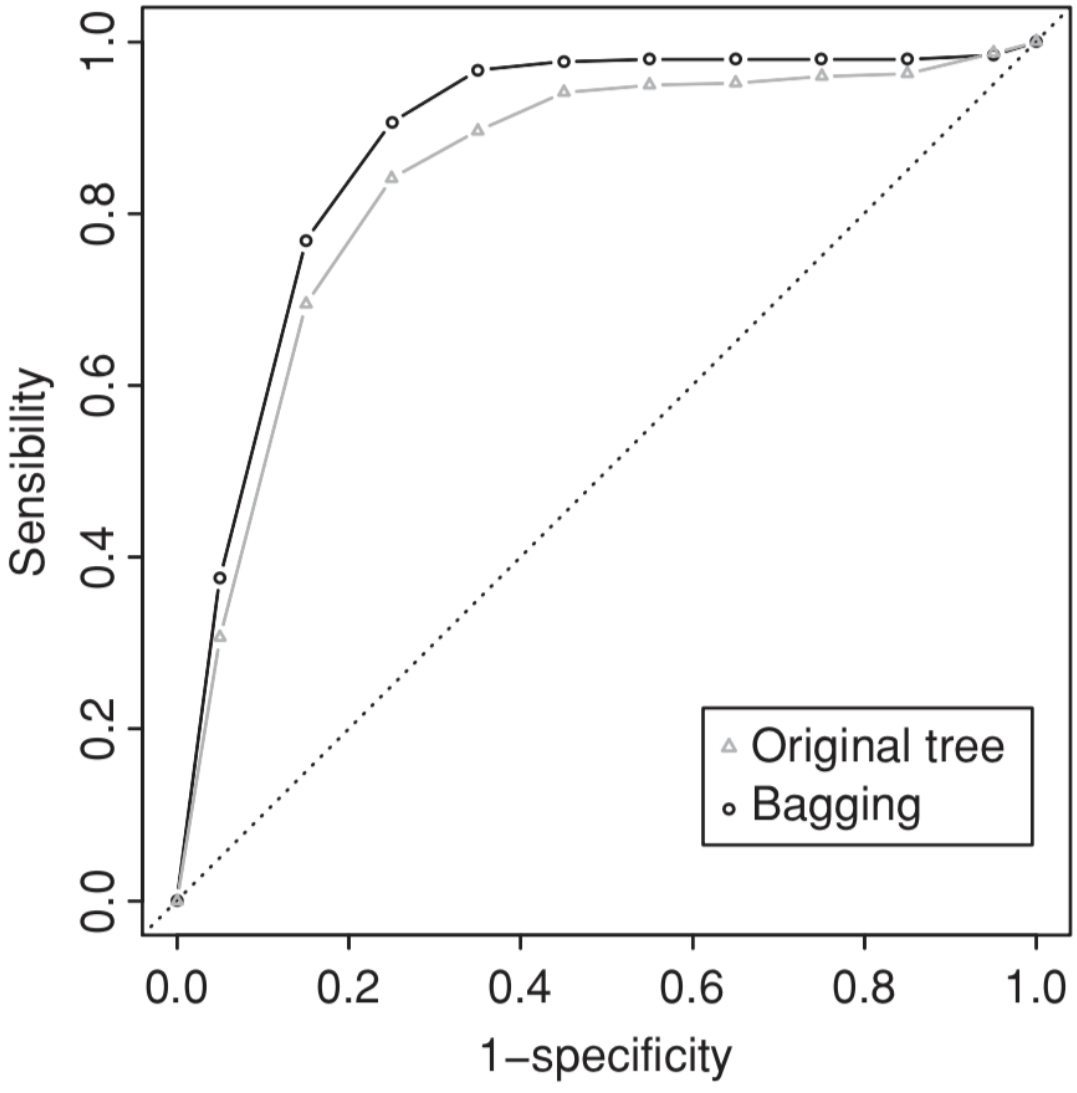
La stessa procedura si puó utilizzare per variabili quantitative, dove non si avrá piú il classificatore ma quello che viene definito il previsore, cioÃ¨ il valore assunto dalla variabile di risposta. In questo caso il nuovo modello avrá generalmente varianza inferiore rispetto al modello originario.

La combinazione dei risultati comporta peró la perdita di ogni semplice struttura nei modelli originari, da cui deriva la difficoltá di interpretazione dei risultati.

Una procedura alternativa é il bumping che individua come classificatore il modello che ha il minor errore di previsione fra tutti i modelli ottenuti con i ricampionamenti boostrap.

L'utilizzo di campioni casuali permette di utilizzare anche la procedura out-of-bag. La parte di dati che viene esclusa dal campione bootstrap viene praticamente usata come insieme di verifica


Misclassification error – Test set
0.180
0.185
0.190
0.195
Original tree
Bagging procedure
0
50
100
150
200
250
300
Number of bootstrap samples


Sensibility
0.0
0.2
0.4
0.6
0.8
1.0
1−specificity
Original tree
Bagging

Boosting

Anche il boosting consiste nel combinare i risultati di un modello adattato usando diversi insiemi di dati, ma selezionando le unitá da inserire nel campione attraverso l'assegnazione a ciascuna unitá di una diversa probabilitá di entrata.

In particolare si assegna maggior peso alle osservazioni che in precedenza sono state classificate peggio: in questo modo si cerca di migliorare la prestazione del nuovo modello, agendo proprio in quelle aree dove il vecchio classificatore presentava maggiori difficoltá.

Boosting

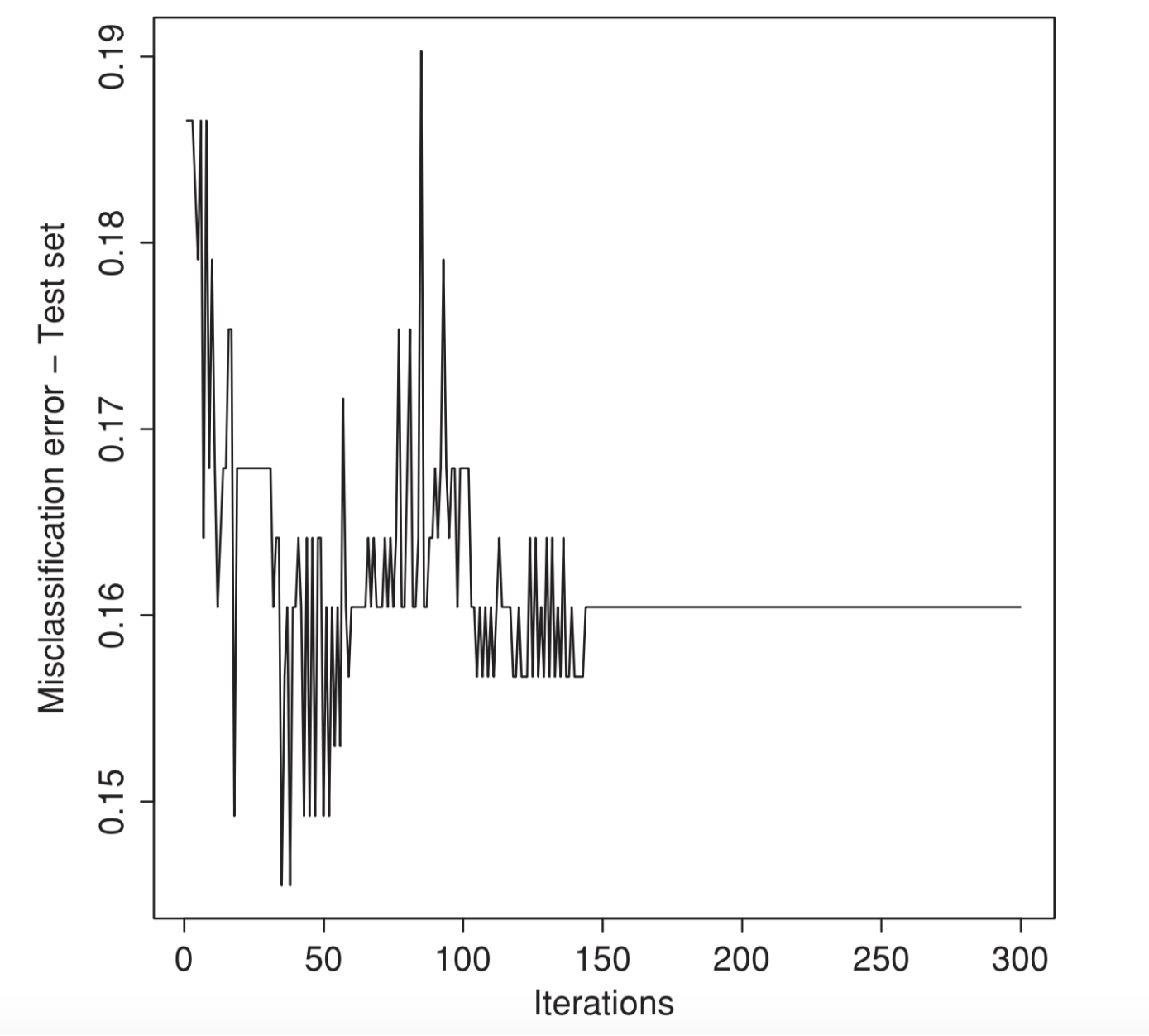
Anche il boosting consiste nel combinare i risultati di un modello adattato usando diversi insiemi di dati, ma selezionando le unitá da inserire nel campione attraverso l'assegnazione a ciascuna unitá di una diversa probabilitá di entrata.

In particolare si assegna maggior peso alle osservazioni che in precedenza sono state classificate peggio: in questo modo si cerca di migliorare la prestazione del nuovo modello, agendo proprio in quelle aree dove il vecchio classificatore presentava maggiori difficoltá.

Si tratta di una procedura iterativa:

- classificatore iniziale costruito dando ad ogni osservazione lo stesso peso;
- si utilizza tale classificatore (definito debole) per la prima classificazione;
- l'insieme dei pesi viene aggiornato ad ogni iterazione;
- Alla fine della procedura si identifica il nuovo classificatore tenendo conto di tutti i modelli costruiti al variare delle iterazioni.

Tale logica é stata concretizzata in diversi modi, ma l'algoritmo piú utilizzato é quello definito AdaBoost.


Misclassification error – Test set
0.15
0.16
0.17
0.18
0.19
0
50
100
150
200
250
300
Iterations

Random forest (Foreste Casuali)

Bagging e boosting fanno variare ad ogni iterazione le unitá statistiche che entrano in gioco nel modello, usando tutte le variabili esplicative disponibili. Le foreste casuali invece considerano diversi sottoinsiemi di variabili esplicative. Tali tecniche sono tipiche degli alberi di classificazione, ma usate anche per altri tipi di classificatori.

La procedura consiste nel selezionare in modo casuale, ad ogni nodo dell'albero, un piccolo gruppo di variabili esplicative che verranno ispezionate per individuare il punto di suddivisione ottimale.

In questo modo non si esplorerebbero tutti i punti di suddivisione, ma solo quelli relativi alle variabili scelte casualmente.

Random forest (Foreste Casuali)

Bagging e boosting fanno variare ad ogni iterazione le unitá statistiche che entrano in gioco nel modello, usando tutte le variabili esplicative disponibili. Le foreste casuali invece considerano diversi sottoinsiemi di variabili esplicative. Tali tecniche sono tipiche degli alberi di classificazione, ma usate anche per altri tipi di classificatori.

La procedura consiste nel selezionare in modo casuale, ad ogni nodo dell'albero, un piccolo gruppo di variabili esplicative che verranno ispezionate per individuare il punto di suddivisione ottimale.

In questo modo non si esplorerebbero tutti i punti di suddivisione, ma solo quelli relativi alle variabili scelte casualmente.

Random forest (Foreste Casuali)

Bagging e boosting fanno variare ad ogni iterazione le unitá statistiche che entrano in gioco nel modello, usando tutte le variabili esplicative disponibili. Le foreste casuali invece considerano diversi sottoinsiemi di variabili esplicative. Tali tecniche sono tipiche degli alberi di classificazione, ma usate anche per altri tipi di classificatori.

La procedura consiste nel selezionare in modo casuale, ad ogni nodo dell'albero, un piccolo gruppo di variabili esplicative che verranno ispezionate per individuare il punto di suddivisione ottimale.

In questo modo non si esplorerebbero tutti i punti di suddivisione, ma solo quelli relativi alle variabili scelte casualmente.

L'albero viene fatto crescere fino alla sua massima dimensione e non viene peró potato. Sará infatti la combinazione di diversi alberi che permetterá di evitare i problemi di sovra-adattamento.

Il numero di variabili da selezionare é un parametro di regolazione da determinare e in genere viene mantenuto costante su tutti i nodi.

Spesso questo viene scelto considerando foreste costruite con un numero diverso di variabili esplicative e determinando quel valore che minimizza l'errore su un insieme di verifica.

Un altro parametro di regolazione da determinare é il numero di alberi che costituiscono la foresta. Si puó mostrare che l'errore globale converge a una soglia inferiore al crescere del numero di alberi.

L'albero viene fatto crescere fino alla sua massima dimensione e non viene peró potato. Sará infatti la combinazione di diversi alberi che permetterá di evitare i problemi di sovra-adattamento.

Il numero di variabili da selezionare é un parametro di regolazione da determinare e in genere viene mantenuto costante su tutti i nodi.

Spesso questo viene scelto considerando foreste costruite con un numero diverso di variabili esplicative e determinando quel valore che minimizza l'errore su un insieme di verifica.

Un altro parametro di regolazione da determinare é il numero di alberi che costituiscono la foresta. Si puó mostrare che l'errore globale converge a una soglia inferiore al crescere del numero di alberi.

L'albero viene fatto crescere fino alla sua massima dimensione e non viene peró potato. Sará infatti la combinazione di diversi alberi che permetterá di evitare i problemi di sovra-adattamento.

Il numero di variabili da selezionare é un parametro di regolazione da determinare e in genere viene mantenuto costante su tutti i nodi.

Spesso questo viene scelto considerando foreste costruite con un numero diverso di variabili esplicative e determinando quel valore che minimizza l'errore su un insieme di verifica.

Un altro parametro di regolazione da determinare é il numero di alberi che costituiscono la foresta. Si puó mostrare che l'errore globale converge a una soglia inferiore al crescere del numero di alberi.

L'albero viene fatto crescere fino alla sua massima dimensione e non viene peró potato. Sará infatti la combinazione di diversi alberi che permetterá di evitare i problemi di sovra-adattamento.

Il numero di variabili da selezionare é un parametro di regolazione da determinare e in genere viene mantenuto costante su tutti i nodi.

Spesso questo viene scelto considerando foreste costruite con un numero diverso di variabili esplicative e determinando quel valore che minimizza l'errore su un insieme di verifica.

Un altro parametro di regolazione da determinare é il numero di alberi che costituiscono la foresta. Si puó mostrare che l'errore globale converge a una soglia inferiore al crescere del numero di alberi.

In genere nella costruzione di una foresta si associa alla selezione casuale delle variabili anche una procedura di bagging.
Praticamente ciascun albero viene fatto crescere su un campione bootstrap e utilizzando per ciascun nodo un numero diverso di variabili.

Tale combinazione permette ancora una volta di utilizzare la procedura out-of-bag, utilizzata anche per valutare l'importanza di ogni variabile esplicativa.

In genere nella costruzione di una foresta si associa alla selezione casuale delle variabili anche una procedura di bagging.
Praticamente ciascun albero viene fatto crescere su un campione bootstrap e utilizzando per ciascun nodo un numero diverso di variabili.

Tale combinazione permette ancora una volta di utilizzare la procedura out-of-bag, utilizzata anche per valutare l'importanza di ogni variabile esplicativa.

Procedura:

- Si costruiscono gli alberi e si effettua la previsione sull'insieme out-of-bag e sullo stesso insieme con i valori della j-esima variabile esplicativa.
- Si misura la differenza tra l'errore di previsione nei due casi e al termine della procedura si considera la media delle differenze tra i vari alberi divisa per l'errore standard.
- Questo valore fornisce una misura di quanto la variabile influisce sulle previsioni.

L'accuratezza della previsione é simile a quella ottenuta col boosting ma soprattutto piú veloce perché ogni albero si basa su poche variabili e quindi la complessitá computazionale é inferiore.

É inoltre facile ottenere un algoritmo di stima che sfrutta il calcolo parallelo che accelera ulteriormente la procedura.

Procedura:

- Si costruiscono gli alberi e si effettua la previsione sull'insieme out-of-bag e sullo stesso insieme con i valori della j-esima variabile esplicativa.
- Si misura la differenza tra l'errore di previsione nei due casi e al termine della procedura si considera la media delle differenze tra i vari alberi divisa per l'errore standard.
- Questo valore fornisce una misura di quanto la variabile influisce sulle previsioni.

L'accuratezza della previsione é simile a quella ottenuta col boosting ma soprattutto piú veloce perché ogni albero si basa su poche variabili e quindi la complessitá computazionale é inferiore.

É inoltre facile ottenere un algoritmo di stima che sfrutta il calcolo parallelo che accelera ulteriormente la procedura.

Procedura:

- Si costruiscono gli alberi e si effettua la previsione sull'insieme out-of-bag e sullo stesso insieme con i valori della j-esima variabile esplicativa.
- Si misura la differenza tra l'errore di previsione nei due casi e al termine della procedura si considera la media delle differenze tra i vari alberi divisa per l'errore standard.
- Questo valore fornisce una misura di quanto la variabile influisce sulle previsioni.

L'accuratezza della previsione é simile a quella ottenuta col boosting ma soprattutto piú veloce perché ogni albero si basa su poche variabili e quindi la complessitá computazionale é inferiore.

É inoltre facile ottenere un algoritmo di stima che sfrutta il calcolo parallelo che accelera ulteriormente la procedura.