

# Árboles (CART), Random Forests y Kernels

Universidad de los Andes y Quantil

Febrero de 2023

# Contenido

- 1 Árboles de clasificación y regresión (CART)
- 2 Árboles de regresión
- 3 Árboles de clasificación
- 4 Random Forests
- 5 Métodos de Kernels: Aprendizaje no supervisado

# Árboles de clasificación y regresión (CART)

- En este caso el mejor clasificador (regresor) se expresa de la forma:

$$\hat{f}(x) = \sum_{m=1}^M c_m I\{x \in R_m\} \quad (1)$$

donde  $R_m$  son diferentes regiones en las cuales la función se va aproximar por una contante  $c_m$ .

- Algunas regiones son difíciles de describir (panel izquierdo arriba siguiente gráfica).
- Una alternativa es encontrar regiones haciendo separaciones binarias secuenciales.

# Árboles de clasificación y regresión (CART)

- En este caso el mejor clasificador (regresor) se expresa de la forma:

$$\hat{f}(x) = \sum_{m=1}^M c_m I\{x \in R_m\} \quad (1)$$

donde  $R_m$  son diferentes regiones en las cuales la función se va aproximar por una contante  $c_m$ .

- Algunas regiones son difíciles de describir (panel izquierdo arriba siguiente gráfica).
- Una alternativa es encontrar regiones haciendo separaciones binarias secuenciales.

# Árboles de clasificación y regresión (CART)

- En este caso el mejor clasificador (regresor) se expresa de la forma:

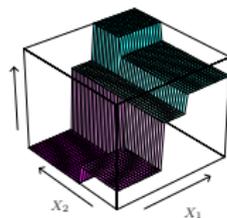
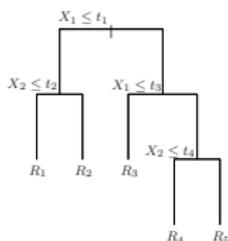
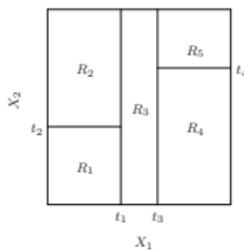
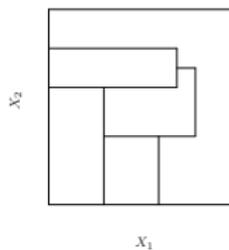
$$\hat{f}(x) = \sum_{m=1}^M c_m I\{x \in R_m\} \quad (1)$$

donde  $R_m$  son diferentes regiones en las cuales la función se va aproximar por una contante  $c_m$ .

- Algunas regiones son difíciles de describir (panel izquierdo arriba siguiente gráfica).
- Una alternativa es encontrar regiones haciendo separaciones binarias secuenciales.

# Árboles de clasificación y regresión (CART)

- Los paneles 2,3,4 representan regiones obtenidas mediante separaciones binarias secuenciales.



# Contenido

- 1 Árboles de clasificación y regresión (CART)
- 2 Árboles de regresión
- 3 Árboles de clasificación
- 4 Random Forests
- 5 Métodos de Kernels: Aprendizaje no supervisado

# Árboles de regresión

- Consideramos primero el caso en el que la variable objetivo es continua (problema de regresión).
- Si la función de pérdida es cuadrática las mejores constantes en cada región son el promedio.

$$\hat{c}_m = \text{media}\{y_i \mid x_i \in R_m\} \quad (2)$$

# Árboles de regresión

- Consideramos primero el caso en el que la variable objetivo es continua (problema de regresión).
- Si la función de pérdida es cuadrática las mejores constantes en cada región son el promedio.

$$\hat{c}_m = \text{media}\{y_i \mid x_i \in R_m\} \quad (2)$$

- Para definir las regiones hacemos lo siguiente. Sea  $j$  una variable de separación y  $s$  el punto de corte. Definimos las dos regiones:

$$R_1(j, s) = \{X \mid x_j \leq s\} \quad (3)$$

$$R_2(j, s) = \{X \mid x_j > s\} \quad (4)$$

- Ahora resolvemos el siguiente problema:

$$\min_{j,s}(\min_{c_1} \sum_{x \in R_1(j,s)} (y_j - c_1)^2 + \min_{c_2} \sum_{x \in R_2(j,s)} (y_j - c_2)^2) \quad (5)$$

- Para cada elección de  $j, s$ , los valores  $c_1, c_2$  que minimizan el error son los promedios en cada región  $R_1(j, s), R_2(j, s)$ , respectivamente.
- Una vez resuelto este problema, con  $j, s, c_1, c_2$ , se repite el proceso en cada region  $R_1(j, s), R_2(j, s)$ .
- Ahora la pregunta fundamental es qué tanto crecemos el árbol? El tamaño del árbol es una medida de la complejidad del mismo.

- Ahora resolvemos el siguiente problema:

$$\min_{j,s} (\min_{c_1} \sum_{x \in R_1(j,s)} (y_j - c_1)^2 + \min_{c_2} \sum_{x \in R_2(j,s)} (y_j - c_2)^2) \quad (5)$$

- Para cada elección de  $j, s$ , los valores  $c_1, c_2$  que minimizan el error son los promedios en cada región  $R_1(j, s), R_2(j, s)$ , respectivamente.
- Una vez resuelto este problema, con  $j, s, c_1, c_2$ , se repite el proceso en cada region  $R_1(j, s), R_2(j, s)$ .
- Ahora la pregunta fundamental es qué tanto crecemos el árbol? El tamaño del árbol es una medida de la complejidad del mismo.

- Ahora resolvemos el siguiente problema:

$$\min_{j,s} (\min_{c_1} \sum_{x \in R_1(j,s)} (y_j - c_1)^2 + \min_{c_2} \sum_{x \in R_2(j,s)} (y_j - c_2)^2) \quad (5)$$

- Para cada elección de  $j, s$ , los valores  $c_1, c_2$  que minimizan el error son los promedios en cada región  $R_1(j, s), R_2(j, s)$ , respectivamente.
- Una vez resuelto este problema, con  $j, s, c_1, c_2$ , se repite el proceso en cada region  $R_1(j, s), R_2(j, s)$ .
- Ahora la pregunta fundamental es qué tanto crecemos el árbol? El tamaño del árbol es una medida de la complejidad del mismo.

- Ahora resolvemos el siguiente problema:

$$\min_{j,s}(\min_{c_1} \sum_{x \in R_1(j,s)} (y_j - c_1)^2 + \min_{c_2} \sum_{x \in R_2(j,s)} (y_j - c_2)^2) \quad (5)$$

- Para cada elección de  $j, s$ , los valores  $c_1, c_2$  que minimizan el error son los promedios en cada región  $R_1(j, s), R_2(j, s)$ , respectivamente.
- Una vez resuelto este problema, con  $j, s, c_1, c_2$ , se repite el proceso en cada region  $R_1(j, s), R_2(j, s)$ .
- Ahora la pregunta fundamental es qué tanto crecemos el árbol? El tamaño del árbol es una medida de la complejidad del mismo.

# Árboles de regresión

- Una alternativa sería solo permitir ramificaciones cuando la reducción en el error supera cierto umbral. Sin embargo esta estrategia puede no ser muy buena en la medida que ramificaciones posteriores pueden reducir sustancialmente el error.
- Otra alternativa es crecer el árbol hasta que cada nodo terminal tenga menos de cierto umbral de ejemplos de la muestra.
- Podar:
  - 1 Cortar ramas y cambiar hojas con las observaciones de todos los nodos terminales que se cortan.
  - 2 El tamaño del árbol se determina podándolo con el objetivo de minimizar una medida del costo de la complejidad (*cost complexity pruning*).

# Árboles de regresión

- Una alternativa sería solo permitir ramificaciones cuando la reducción en el error supera cierto umbral. Sin embargo esta estrategia puede no ser muy buena en la medida que ramificaciones posteriores pueden reducir sustancialmente el error.
- Otra alternativa es crecer el árbol hasta que cada nodo terminal tenga menos de cierto umbral de ejemplos de la muestra.
- Podar:
  - 1 Cortar ramas y cambiar hojas con las observaciones de todos los nodos terminales que se cortan.
  - 2 El tamaño del árbol se determina podándolo con el objetivo de minimizar una medida del costo de la complejidad (*cost complexity pruning*).

# Árboles de regresión

- Una alternativa sería solo permitir ramificaciones cuando la reducción en el error supera cierto umbral. Sin embargo esta estrategia puede no ser muy buena en la medida que ramificaciones posteriores pueden reducir sustancialmente el error.
- Otra alternativa es crecer el árbol hasta que cada nodo terminal tenga menos de cierto umbral de ejemplos de la muestra.
- Podar:
  - 1 Cortar ramas y cambiar hojas con las observaciones de todos los nodos terminales que se cortan.
  - 2 El tamaño del árbol se determina podándolo con el objetivo de minimizar una medida del costo de la complejidad (*cost complexity pruning*).

# Árboles de regresión

- Una alternativa sería solo permitir ramificaciones cuando la reducción en el error supera cierto umbral. Sin embargo esta estrategia puede no ser muy buena en la medida que ramificaciones posteriores pueden reducir sustancialmente el error.
- Otra alternativa es crecer el árbol hasta que cada nodo terminal tenga menos de cierto umbral de ejemplos de la muestra.
- Podar:
  - 1 Cortar ramas y cambiar hojas con las observaciones de todos los nodos terminales que se cortan.
  - 2 El tamaño del árbol se determina podándolo con el objetivo de minimizar una medida del costo de la complejidad (*cost complexity pruning*).

# Árboles de regresión

- Sea  $T$  un árbol, indexemos por  $m$  los nodos terminales del árbol y sea  $|T|$  el número de nodos terminales en  $T$ .
- Sea  $N_m$  el número de observaciones tal que  $x \in R_m$ .
- $\hat{c}_m$  el promedio de las variables objetivo en la región  $R_m$ .
- Sea  $Q_m(T)$  el error promedio cuadrático entre la variable objetivo y el promedio en la región.
- El costo de la complejidad se define como:

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha |T| \quad (6)$$

# Árboles de regresión

- Sea  $T$  un árbol, indexemos por  $m$  los nodos terminales del árbol y sea  $|T|$  el número de nodos terminales en  $T$ .
- Sea  $N_m$  el número de observaciones tal que  $x \in R_m$ .
- $\hat{c}_m$  el promedio de las variables objetivo en la región  $R_m$ .
- Sea  $Q_m(T)$  el error promedio cuadrático entre la variable objetivo y el promedio en la región.
- El costo de la complejidad se define como:

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha |T| \quad (6)$$

# Árboles de regresión

- Sea  $T$  un árbol, indexemos por  $m$  los nodos terminales del árbol y sea  $|T|$  el número de nodos terminales en  $T$ .
- Sea  $N_m$  el número de observaciones tal que  $x \in R_m$ .
- $\hat{c}_m$  el promedio de las variables objetivo en la región  $R_m$ .
- Sea  $Q_m(T)$  el error promedio cuadrático entre la variable objetivo y el promedio en la región.
- El costo de la complejidad se define como:

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha |T| \quad (6)$$

# Árboles de regresión

- Sea  $T$  un árbol, indexemos por  $m$  los nodos terminales del árbol y sea  $|T|$  el número de nodos terminales en  $T$ .
- Sea  $N_m$  el número de observaciones tal que  $x \in R_m$ .
- $\hat{c}_m$  el promedio de las variables objetivo en la región  $R_m$ .
- Sea  $Q_m(T)$  el error promedio cuadrático entre la variable objetivo y el promedio en la región.
- El costo de la complejidad se define como:

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha |T| \quad (6)$$

# Árboles de regresión

- Sea  $T$  un árbol, indexemos por  $m$  los nodos terminales del árbol y sea  $|T|$  el número de nodos terminales en  $T$ .
- Sea  $N_m$  el número de observaciones tal que  $x \in R_m$ .
- $\hat{c}_m$  el promedio de las variables objetivo en la región  $R_m$ .
- Sea  $Q_m(T)$  el error promedio cuadrático entre la variable objetivo y el promedio en la región.
- El costo de la complejidad se define como:

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha |T| \quad (6)$$

- El problema que queremos resolver es:

$$\min_{T' \subset T} C_\alpha(T')$$

donde  $T'$  es cualquier sub árbol que se puede obtener de  $T$  podándolo.

- Valores altos de  $\alpha$  resulta en árboles más pequeños. Si es cero el resultado es el árbol  $T$ .
- El parámetro  $\alpha$  se puede determinar usando validación cruzada. Por ejemplo el valor de  $\alpha$  que reduzca el error promedio,  $M$ -fold, de validación cruzada.

- El problema que queremos resolver es:

$$\min_{T' \subset T} C_\alpha(T')$$

donde  $T'$  es cualquier sub árbol que se puede obtener de  $T$  podándolo.

- Valores altos de  $\alpha$  resulta en árboles más pequeños. Si es cero el resultado es el árbol  $T$ .
- El parámetro  $\alpha$  se puede determinar usando validación cruzada. Por ejemplo el valor de  $\alpha$  que reduzca el error promedio,  $M$ -fold, de validación cruzada.

- El problema que queremos resolver es:

$$\min_{T' \subset T} C_\alpha(T')$$

donde  $T'$  es cualquier sub árbol que se puede obtener de  $T$  podándolo.

- Valores altos de  $\alpha$  resulta en árboles más pequeños. Si es cero el resultado es el árbol  $T$ .
- El parámetro  $\alpha$  se puede determinar usando validación cruzada. Por ejemplo el valor de  $\alpha$  que reduzca el error promedio,  $M$ -fold, de validación cruzada.

# Contenido

- 1 Árboles de clasificación y regresión (CART)
- 2 Árboles de regresión
- 3 Árboles de clasificación**
- 4 Random Forests
- 5 Métodos de Kernels: Aprendizaje no supervisado

# Árboles de clasificación

- La única modificación necesaria es definir la función de error en cada ramificación:  $Q_m(T)$  (i.e. medida de impureza).
- Sean  $R_1, \dots, R_M$  un conjunto de regiones.
- Sea  $p(m, k)$  La cantidad de observaciones de la clase  $k$  en la región  $m$  como proporción de la cantidad de observaciones en esa región.
- Sea  $k(m)$  la clase mayoritaria en la región  $m$ .
- Definimos el error de clasificación como  $L(m) = 1 - p(m, k(m))$ .
- Ahora para crecer y podar el árbol se sigue el mismo algoritmo de árboles de regresión.

# Árboles de clasificación

- La única modificación necesaria es definir la función de error en cada ramificación:  $Q_m(T)$  (i.e. medida de impureza).
- Sean  $R_1, \dots, R_M$  un conjunto de regiones.
- Sea  $p(m, k)$  La cantidad de observaciones de la clase  $k$  en la región  $m$  como proporción de la cantidad de observaciones en esa región.
- Sea  $k(m)$  la clase mayoritaria en la región  $m$ .
- Definimos el error de clasificación como  $L(m) = 1 - p(m, k(m))$ .
- Ahora para crecer y podar el árbol se sigue el mismo algoritmo de árboles de regresión.

# Árboles de clasificación

- La única modificación necesaria es definir la función de error en cada ramificación:  $Q_m(T)$  (i.e. medida de impureza).
- Sean  $R_1, \dots, R_M$  un conjunto de regiones.
- Sea  $p(m, k)$  La cantidad de observaciones de la clase  $k$  en la región  $m$  como proporción de la cantidad de observaciones en esa región.
- Sea  $k(m)$  la clase mayoritaria en la región  $m$ .
- Definimos el error de clasificación como  $L(m) = 1 - p(m, k(m))$ .
- Ahora para crecer y podar el árbol se sigue el mismo algoritmo de árboles de regresión.

# Árboles de clasificación

- La única modificación necesaria es definir la función de error en cada ramificación:  $Q_m(T)$  (i.e. medida de impureza).
- Sean  $R_1, \dots, R_M$  un conjunto de regiones.
- Sea  $p(m, k)$  La cantidad de observaciones de la clase  $k$  en la región  $m$  como proporción de la cantidad de observaciones en esa región.
- Sea  $k(m)$  la clase mayoritaria en la región  $m$ .
- Definimos el error de clasificación como  $L(m) = 1 - p(m, k(m))$ .
- Ahora para crecer y podar el árbol se sigue el mismo algoritmo de árboles de regresión.

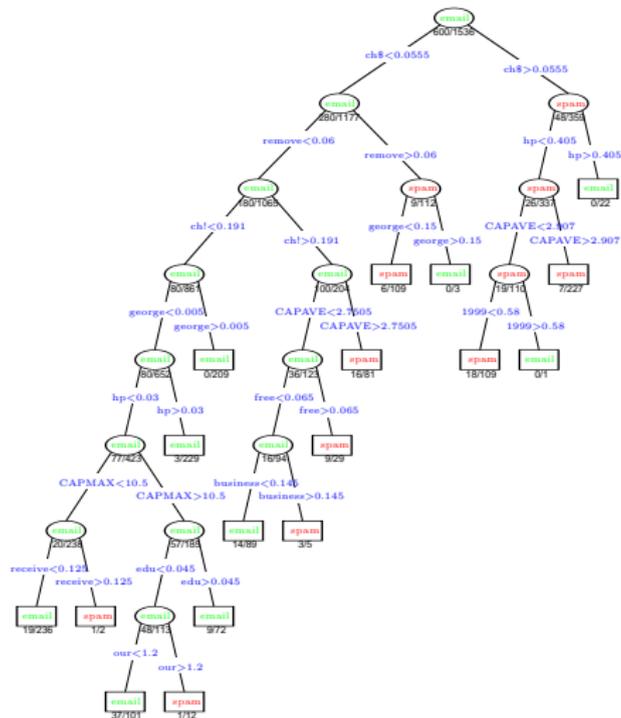
# Árboles de clasificación

- La única modificación necesaria es definir la función de error en cada ramificación:  $Q_m(T)$  (i.e. medida de impureza).
- Sean  $R_1, \dots, R_M$  un conjunto de regiones.
- Sea  $p(m, k)$  La cantidad de observaciones de la clase  $k$  en la región  $m$  como proporción de la cantidad de observaciones en esa región.
- Sea  $k(m)$  la clase mayoritaria en la región  $m$ .
- Definimos el error de clasificación como  $L(m) = 1 - p(m, k(m))$ .
- Ahora para crecer y podar el árbol se sigue el mismo algoritmo de árboles de regresión.

# Árboles de clasificación

- La única modificación necesaria es definir la función de error en cada ramificación:  $Q_m(T)$  (i.e. medida de impureza).
- Sean  $R_1, \dots, R_M$  un conjunto de regiones.
- Sea  $p(m, k)$  La cantidad de observaciones de la clase  $k$  en la región  $m$  como proporción de la cantidad de observaciones en esa región.
- Sea  $k(m)$  la clase mayoritaria en la región  $m$ .
- Definimos el error de clasificación como  
$$L(m) = 1 - p(m, k(m)).$$
- Ahora para crecer y podar el árbol se sigue el mismo algoritmo de árboles de regresión.

# Árboles de clasificación



**FIGURE 9.5.** The pruned tree for the `spam` example. The split variables are shown in blue on the branches, and the classification is shown in every node. The numbers under the terminal nodes indicate misclassification rates on the test data.

# Contenido

- 1 Árboles de clasificación y regresión (CART)
- 2 Árboles de regresión
- 3 Árboles de clasificación
- 4 Random Forests**
- 5 Métodos de Kernels: Aprendizaje no supervisado

# Random Forests

- Es una técnica para reducir varianza construyendo una gran cantidad de árboles no correlacionados para después promediarlos.

# Random Forests: Algoritmo

- Para  $b = 1, \dots, B$
- Muestrear  $Z^*$  del mismo tamaño de la muestra.
- Construir un árbol  $T_b$  usando el siguiente procedimiento hasta que se alcance un número de datos sea menor que  $n_{min}$  en cada nodo.
  - 1 Elegir aleatoriamente  $m$  variables de las  $p$ .
  - 2 Elegir la mejor división en el nodo entre las  $m$ .
  - 3 Repetir los dos pasos anteriores hasta alcanzar el mínimo en cada hoja.
- Promedia los árboles en el caso de un problema de regresión o usar voto mayoritario (cada árbol contribuye con un voto) en el caso de clasificación.

# Random Forests: Propiedades

- Si cada árbol se crece bastante es posible reducir el sesgo pero la varianza aumenta. Al promediar se reduce la varianza.
- Como los árboles se contruyen usando el mismo procedimiento, el sesgo del promedio es similar al sesgo de cada árbol. Potencialmente la ganancia está en en la reducción de varianza.
- La reducción de la varianza se obtiene eficientemente al promediar árboles no correlacionados.
- La elección aleatoria en cada división garantiza que unas pocas variables no dominen la regresión.

# Random Forests: Propiedades

- Si cada árbol se crece bastante es posible reducir el sesgo pero la varianza aumenta. Al promediar se reduce la varianza.
- Como los árboles se contruyen usando el mismo procedimiento, el sesgo del promedio es similar al sesgo de cada árbol. Potencialmente la ganancia está en en la reducción de varianza.
- La reducción de la varianza se obtiene eficientemente al promediar árboles no correlacionados.
- La elección aleatoria en cada división garantiza que unas pocas variables no dominen la regresión.

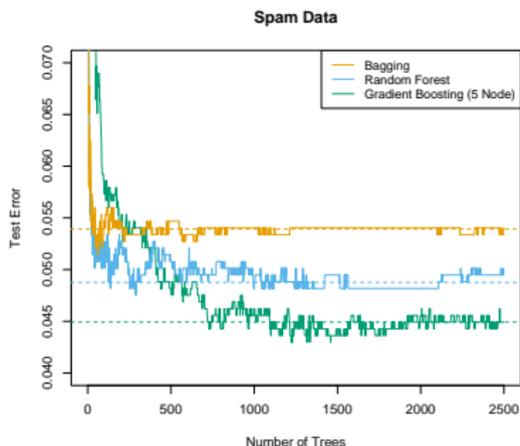
# Random Forests: Propiedades

- Si cada árbol se crece bastante es posible reducir el sesgo pero la varianza aumenta. Al promediar se reduce la varianza.
- Como los árboles se contruyen usando el mismo procedimiento, el sesgo del promedio es similar al sesgo de cada árbol. Potencialmente la ganancia está en en la reducción de varianza.
- La reducción de la varianza se obtiene eficientemente al promediar árboles no correlacionados.
- La elección aleatoria en cada división garantiza que unas pocas variables no dominen la regresión.

# Random Forests: Propiedades

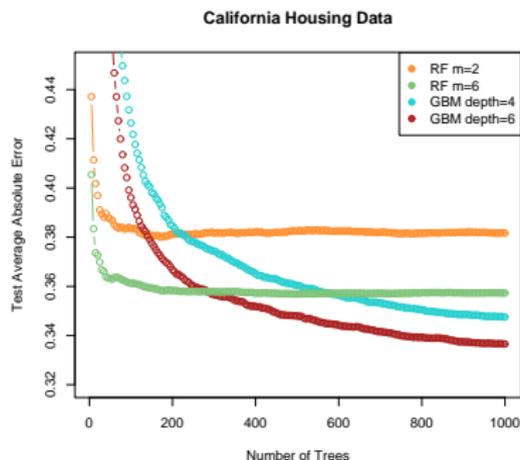
- Si cada árbol se crece bastante es posible reducir el sesgo pero la varianza aumenta. Al promediar se reduce la varianza.
- Como los árboles se contruyen usando el mismo procedimiento, el sesgo del promedio es similar al sesgo de cada árbol. Potencialmente la ganancia está en en la reducción de varianza.
- La reducción de la varianza se obtiene eficientemente al promediar árboles no correlacionados.
- La elección aleatoria en cada división garantiza que unas pocas variables no dominen la regresión.

# Random Forests: Rendimiento



**FIGURE 15.1.** Bagging, random forest, and gradient boosting, applied to the spam data. For boosting, 5-node trees were used, and the number of trees were chosen by 10-fold cross-validation (2500 trees). Each “step” in the figure corresponds to a change in a single misclassification (in a test set of 1536).

# Random Forests: Rendimiento



**FIGURE 15.3.** *Random forests compared to gradient boosting on the California housing data. The curves represent mean absolute error on the test data as a function of the number of trees in the models. Two random forests are shown, with  $m = 2$  and  $m = 6$ . The two gradient boosted models use a shrinkage parameter  $\nu = 0.05$  in (10.41), and have interaction depths of 4 and 6. The boosted models outperform random forests.*

- Para clasificación  $m = \sqrt{p}$  y  $n_{min} = 1$ .
- Para regresión  $m = \frac{p}{3}$  y  $n_{min} = 5$ .
- En la práctica deben ser calibrados. Por ejemplo, en el ejemplo de California, funciona mejor otros valores.

# Random Forests: Importancia relativa

- Dado un nodo  $t$  del árbol, sea  $\nu(t)$  la variable que se usa en ese nodo para dividir en el árbol. Sea  $L(\nu(t))$  la pérdida en ese nodo si no se hiciera la división y  $L_1(t)$  y  $L_2(t)$  la pérdida en cada una de las divisiones que se realizan.
- Definamos la importancia de usar la variable  $l$  en el árbol  $T$  como:

$$I_l^2(T) = \sum_t i_t^2 I(\nu(t) = l)$$

donde  $i_t^2 = (L(\nu(t)) - L(1) - L(2))^2$

- Para un bosque aleatorio se define como:

$$I_l^2 = \frac{1}{B} \sum_b I_l^2(T_b)$$

- Se normaliza la importancia más alta en 100.



# Contenido

- 1 Árboles de clasificación y regresión (CART)
- 2 Árboles de regresión
- 3 Árboles de clasificación
- 4 Random Forests
- 5 Métodos de Kernels: Aprendizaje no supervisado

## Estimación de densidades

- Supongamos que  $x_1, \dots, x_N$  es una muestra de datos tomada con una distribución con densidad  $f_X(x)$ .
- Un primer estimador local es:

$$f_X(x_0) = \frac{\text{num}\{N(x_0)\}}{N\lambda} \quad (7)$$

donde  $N(x_0)$  es una vecindad de tamaño  $\lambda$ .

- Una versión suavizada es:

$$f_X(x_0) = \frac{1}{N\lambda} \sum_{i=1}^N K_\lambda(x_0, x_i) \quad (8)$$

donde por ejemplo  $K_\lambda(x_0, x_i) = \psi\left(\frac{|x-x_0|}{\lambda}\right)$  y  $\psi_\lambda$  es la densidad Gaussiana  $(0, \lambda^2)$ .

## Estimación de densidades

- Supongamos que  $x_1, \dots, x_N$  es una muestra de datos tomada con una distribución con densidad  $f_X(x)$ .
- Un primer estimador local es:

$$f_X(x_0) = \frac{\text{num}\{N(x_0)\}}{N\lambda} \quad (7)$$

donde  $N(x_0)$  es una vecindad de tamaño  $\lambda$ .

- Una versión suavizada es:

$$f_X(x_0) = \frac{1}{N\lambda} \sum_{i=1}^N K_\lambda(x_0, x_i) \quad (8)$$

donde por ejemplo  $K_\lambda(x_0, x_i) = \psi\left(\frac{|x-x_0|}{\lambda}\right)$  y  $\psi_\lambda$  es la densidad Gaussiana  $(0, \lambda^2)$ .

## Estimación de densidades

- Supongamos que  $x_1, \dots, x_N$  es una muestra de datos tomada con una distribución con densidad  $f_X(x)$ .
- Un primer estimador local es:

$$f_X(x_0) = \frac{\text{num}\{N(x_0)\}}{N\lambda} \quad (7)$$

donde  $N(x_0)$  es una vecindad de tamaño  $\lambda$ .

- Una versión suavizada es:

$$f_X(x_0) = \frac{1}{N\lambda} \sum_{i=1}^N K_\lambda(x_0, x_i) \quad (8)$$

donde por ejemplo  $K_\lambda(x_0, x_i) = \psi\left(\frac{|x-x_0|}{\lambda}\right)$  y  $\psi_\lambda$  es la densidad Gaussiana  $(0, \lambda^2)$ .

# Estimación de densidades

- En forma reducida:

$$f_X(x) = \frac{1}{N\lambda} \sum_{i=1}^N \psi_\lambda(x - x_i) \quad (9)$$

- En  $p$  dimensiones:

$$f_X(x) = \frac{1}{N\lambda(2\lambda^2\pi)^{\frac{p}{2}}} \sum_{i=1}^N \exp\left(-\frac{1}{2}\left(\frac{x - x_i}{\lambda}\right)^2\right) \quad (10)$$

## Estimación de densidades

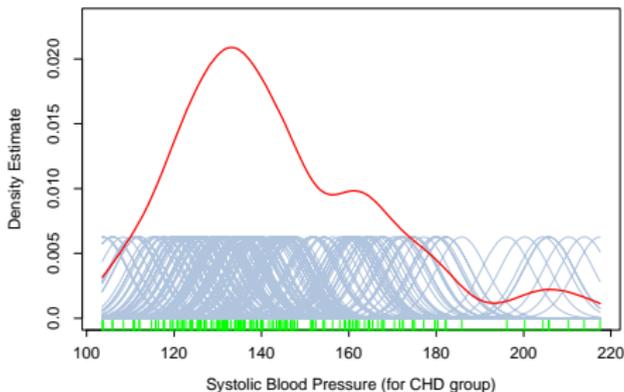
- En forma reducida:

$$f_X(x) = \frac{1}{N\lambda} \sum_{i=1}^N \psi_\lambda(x - x_i) \quad (9)$$

- En  $p$  dimensiones:

$$f_X(x) = \frac{1}{N\lambda(2\lambda^2\pi)^{\frac{p}{2}}} \sum_{i=1}^N \exp\left(-\frac{1}{2}\left(\frac{x - x_i}{\lambda}\right)^2\right) \quad (10)$$

## Estimación de densidades



**FIGURE 6.13.** A kernel density estimate for systolic blood pressure (for the CHD group). The density estimate at each point is the average contribution from each of the kernels at that point. We have scaled the kernels down by a factor of 10 to make the graph readable.