

CART, Bagging, Forests, Boosting

Test de Lectura

Como discutiesemos, estos *tests de lectura* son una serie de preguntas simples que permiten monitorear si estan siguiendo efectivamente las clases. No seran corregidos ni evaluados, los administraran directamente ustedes. Cualquier duda que tengan sera discutida en nuestras sesiones sincronicas, la semana siguiente a publicado el material, luego de que hayan hecho un esfuerzo en trabajarlo, no antes. Es muy importante que trabajen este material y lo discutan entre ustedes.

1. En el ejemplo de 'hitters' del slide 5, en que sentido el modelo CART construido es 'no-lineal'.
2. Como le explicarias a alguien 'no tecnico' como funciona el ejemplo de hitters y en que sentido es no lineal?
3. En el slide 6, porque una vez que elegimos la variable de particion y en que punto partir, la solucion al problema de minimizacion dentro de una region esta dado por el promedio de los Y 's para cada region?
4. Cuanto vale la formula anterior en un arbol que en su nodo terminal tiene tantas ramificaciones como observaciones?
5. En regresion, la complejidad la da la cantidad de regresores. Y en CART?
6. Por que 'podar' el arbol aumenta la impureza en cada paso de la poda?
7. Cual es el proposito de usar 'weakest link pruning'?
8. Lee el paper de Varian (2003) y chequea que entendes el ejemplo del Titanic. Tambien lo discuto con detalle en Big Data.
9. Que tipo de problema resuelven bagging, forests y boosting?
10. Por que si hay un predictor fuerte bagging no produce una autentica reduccion en la varianza del predictor?
11. Como se computa el error de prediccion usando 'out of bag'?
12. Como es que random forest resuelve el problema anterior?
13. En boosting, que mide α_m ?
14. Intuitivamente, como hace boosting para ajustar los ponderadores en cada iteracion?

15. *Para pensar*: vuelve al modelo de regresión simple, con dos variables Y y X . Pensa intuitivamente que haría boosting usando como predictor mínimos cuadrados. Pensa el caso en donde $M = 3$ (hace un dibujo, solo necesito una discusión informal, intuitiva).