Autor: Nikita Polyanskiy

Sesión 04 de Noviembre 2021

Tema 6: Boosting y adaboost

Bagging:

Bootstrap aggregation, también llamado bagging, consiste en entrenar los diferentes modelos con subconjuntos del conjunto de entrenamiento, dando a cada resultado obtenido un voto con el mismo peso. Por defecto, las muestras escogidas para cada aprendiz son escogidas aleatoriamente con reemplazo (lo que significa que una muestra dada puede ser escogida más de una vez en cada bloque de entrenamiento aleatorio)

Boosting:

El boosting consiste en combinar los resultados de varios clasificadores débiles para obtener un clasificador robusto. Cuando se añaden estos clasificadores débiles, se lo hace de modo que estos tengan diferente peso en función de la exactitud de sus predicciones. Luego de que se añade un clasificador débil, los datos cambian su estructura de pesos: los casos que son mal clasificados ganan peso y los que son clasificados correctamente pierden peso. Así, los clasificadores débiles se centran de mayor manera en los casos que fueron mal clasificados por los clasificadores débiles.

Adaboost:

Explicación sencilla de AdaBoost

1. Entrene un clasificador.
2. Use el clasificador.
3. Identifique los casos que fueron mal clasificados.
4. Construya un nuevo clasificador que clasifique mejor los casos mal clasificados del punto anterior.
5. Repita los pasos 2 a 4 varias veces.
6. Asígnele un peso a cada clasificador y júntelos para obtener un clasificador con mejor desempeño.

Explicación detallada de AdaBoost

1. Inicie con un conjunto de entrenamiento (X,Y) con m observaciones denotadas como $(x_1,y_1),\dots,(x_m,y_m)$ de tal manera que $x_i \in R^p$ Los valores de y deben ser -1 o 1 para aplicar el método.
2. Inicie con la distribución discreta $D_1(i)=1/m$ que indica el peso de la observación i en la iteración 1.
3. Para $t=1,\dots,T$

- Construya un clasificador ht definido así: ht:X→{−1,1}.
- Calcule el error asociado ϵt al clasificador $\epsilon_t = \sum_{i=1}^{m} D_t(i) \times \delta_i$
  donde δi=0 si ht(xi)=yi, es decir, si fue correcta la clasificación; caso contrario es δi=1.
- Calcule la nueva distribución Dt+1(i)=Dt(i)×Fi/Zt, donde:
  - Fi=exp(−αt) si la clasificación fue correcta, es decir si ht(xi)=yi.
  - Fi=exp(αt) si la clasificación fue incorrecta, es decir si ht(xi)≠yi.
  - αt= (1/2) * log ((1−ϵt) / ϵt).
  - Zt es una constante de normalización de tal manera que $\sum_{i=1}^{m} D_t(i) = 1$
  - Usualmente es $\sum D_t(i) \times F_i$

3. Construya el clasificador final Hfinal como el promedio ponderado de los t clasificadores ht, usando $H_{final} = sign(\sum_t \alpha_t h_t(x))$