

Boosting

Marzo de 2018

Contenido

1 Boosting

AdaBoost

- Viene de Adaptive Boosting.
- Supongamos que tenemos una muestra $\tau_n = \{(x_1, y_1), \dots, (x_m, y_m)\}$ donde $y_i \in \{-1, 1\}$.
- Defina unos pesos iniciales para cada observación i : $D_1(i) = \frac{1}{m}$. D_t siempre será una distribución sobre las m observaciones.

AdaBoost: Algoritmo

- Para cada $t = 1, \dots, T$
- Construir un clasificador (puede ser débil) h_t que minimice la función de pérdida:
 - 1 Defina el error e_t como:

$$e_t = \sum_{i=1}^m D_t(i) I(y_i \neq h_t(x_i)) \quad (1)$$

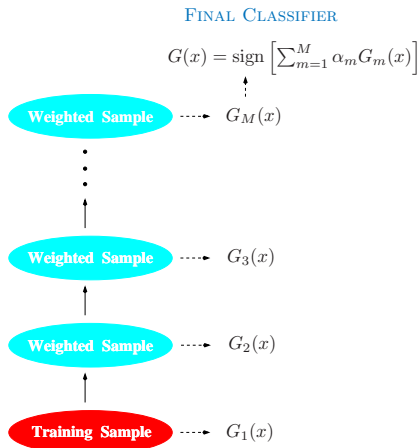
- 2 Calcular $\alpha_t = \frac{1}{2} \log\left(\frac{1-e_t}{e_t}\right)$
- 3 Modificar los pesos:

$$D_{t+1}(i) \rightarrow \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

donde $Z_t = \sum_{i=1}^m D_t(i)$

- $H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$

AdaBoost: Algoritmo



Boosting: Es un clasificador muy potente

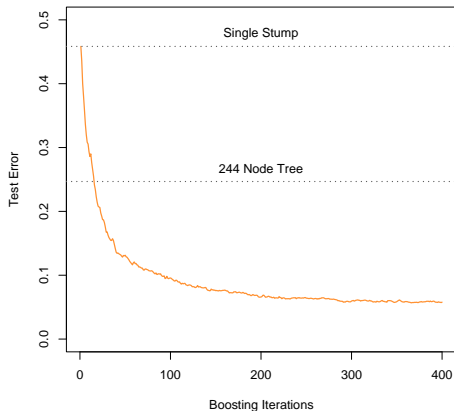


FIGURE 10.2. *Simulated data (10.2): test error rate for boosting with stumps, as a function of the number of iterations. Also shown are the test error rate for a single stump, and a 244-node classification tree.*

Boosting: Observaciones

- Si elegimos un ejemplo de τ de forma aleatoria usando la distribución D_t la probabilidad de que h_t lo clasifique de forma errada es e_t .
- e_t se mide usando la misma distribución D_t con respecto a la cual se entrena la hipótesis h_t .
- No es absolutamente necesario que h_t minimice e_t solamente que sea marginalmente mejor que un clasificador aleatorio (i.e., la hipótesis de clasificadores débiles significa que existe un $\gamma > 0$ tal que $e_t < \frac{1}{2} - \gamma$).
- $\alpha_t > 0$ si $e_t < \frac{1}{2}$ y es mayor entre menor sea el error.
- Si calculamos el error de h_t usando la distribución D_{t+1} este es idéntico a $\frac{1}{2}$.
- Más adelante mostramos que el error de entrenamiento cae exponencialmente como función del número de clasificadores débiles que se usen para combinar.

- Por el momento no hay garantía de que el clasificador tenga un buen error de generalización.
- El error de entrenamiento $h(x)$ lo denotamos por $\widehat{err}(h)$ y es simplemente la frecuencia de ejemplos mal clasificados por h sin ningún tipo de ponderación.

Boosting en acción: clasificadores débiles

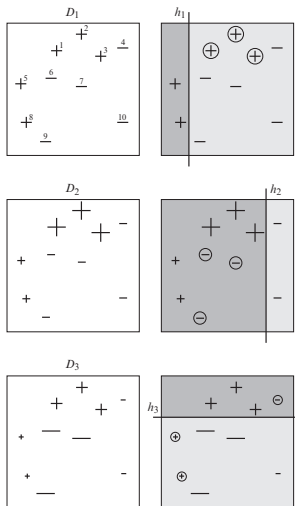


Figure 1.1

An illustration of how AdaBoost behaves on a tiny toy problem with $m = 10$ examples. Each row depicts one round, for $t = 1, 2, 3$. The left box in each row represents the distribution D_t , with the size of each example scaled in proportion to its weight under that distribution. Each box on the right shows the weak hypothesis h_t , where darker shading indicates the region of the domain predicted to be positive. Examples that are misclassified by h_t have been circled.

Boosting en acción: clasificadores débiles

Table 1.1

The numerical calculations corresponding to the toy example in figure 1.1

	1	2	3	4	5	6	7	8	9	10	
$D_1(i)$	<u>0.10</u>	<u>0.10</u>	<u>0.10</u>	0.10	0.10	0.10	0.10	0.10	0.10	0.10	$\epsilon_1 = 0.30, \alpha_1 \approx 0.42$
$e^{-\alpha_1 y_i h_1(x_i)}$	1.53	1.53	1.53	0.65	0.65	0.65	0.65	0.65	0.65	0.65	
$D_1(i) e^{-\alpha_1 y_i h_1(x_i)}$	0.15	0.15	0.15	0.07	0.07	0.07	0.07	0.07	0.07	0.07	$Z_1 \approx 0.92$
$D_2(i)$	0.17	0.17	0.17	0.07	0.07	<u>0.07</u>	<u>0.07</u>	0.07	<u>0.07</u>	0.07	$\epsilon_2 \approx 0.21, \alpha_2 \approx 0.65$
$e^{-\alpha_2 y_i h_2(x_i)}$	0.52	0.52	0.52	0.52	0.52	1.91	1.91	0.52	1.91	0.52	
$D_2(i) e^{-\alpha_2 y_i h_2(x_i)}$	0.09	0.09	0.09	0.04	0.04	0.14	0.14	0.04	0.14	0.04	$Z_2 \approx 0.82$
$D_3(i)$	0.11	0.11	0.11	<u>0.05</u>	<u>0.05</u>	0.17	0.17	<u>0.05</u>	0.17	0.05	$\epsilon_3 \approx 0.14, \alpha_3 \approx 0.92$
$e^{-\alpha_3 y_i h_3(x_i)}$	0.40	0.40	0.40	2.52	2.52	0.40	0.40	2.52	0.40	0.40	
$D_3(i) e^{-\alpha_3 y_i h_3(x_i)}$	0.04	0.04	0.04	0.11	0.11	0.07	0.07	0.11	0.07	0.02	$Z_3 \approx 0.69$

Calculations are shown for the ten examples as numbered in the figure. Examples on which hypothesis h_t makes a mistake are indicated by underlined figures in the rows marked D_t .

Boosting en acción: clasificadores débiles

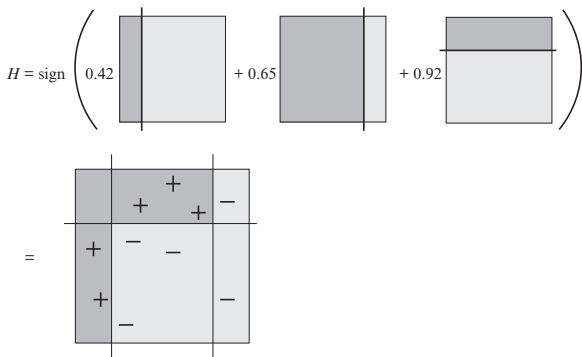


Figure 1.2

The combined classifier for the toy example of figure 1.1 is computed as the sign of the weighted sum of the three weak hypotheses, $\alpha_1 h_1 + \alpha_2 h_2 + \alpha_3 h_3$, as shown at the top. This is equivalent to the classifier shown at the bottom. (As in figure 1.1, the regions that a classifier predicts positive are indicated using darker shading.)

Ejemplo de clasificación de enfermedad cardiaca

Table 1.2

The first six base classifiers found when using AdaBoost on the heart-disease dataset

Round	If	Then Predict	Else Predict
1	thalamus normal	healthy	sick
2	number of major vessels colored by fluoroscopy > 0	sick	healthy
3	chest pain type is asymptomatic	sick	healthy
4	ST depression induced by exercise relative to rest ≥ 0.75	sick	healthy
5	cholesterol ≥ 228.5	sick	healthy
6	resting electrocardiographic results are normal	healthy	sick

Ejemplo de clasificación de enfermedad cardiaca

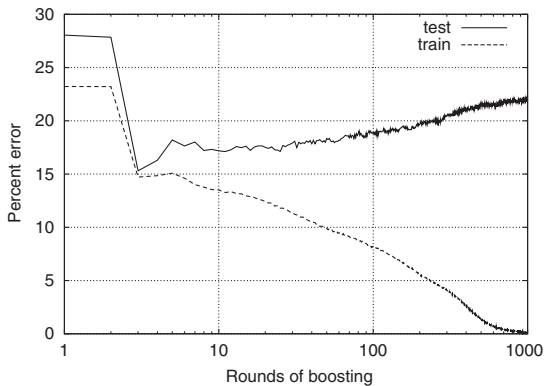


Figure 1.5

The training and test percent error rates obtained using boosting on the heart-disease dataset. Results are averaged over multiple train-test splits of the data.

Non Overfitting

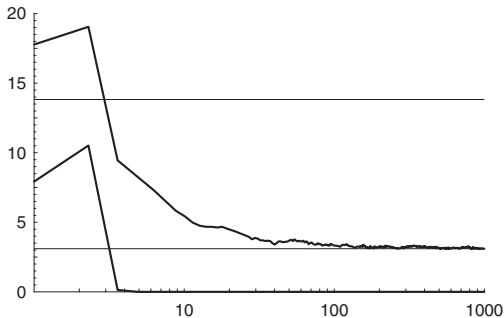


Figure 1.7

The training and test percent error rates obtained using boosting on an OCR dataset with C4.5 as the base learner. The top and bottom curves are test and training error, respectively. The top horizontal line shows the test error rate using just C4.5. The bottom line shows the final test error rate of AdaBoost after 1000 rounds. (Reprinted with permission of the Institute of Mathematical Statistics.)

Non Overfitting

- En el anterior ejemplo el error de prueba no empeora a pesar de que la complejidad aumenta.
- Esto es una ilustración de la teoría marginal del boosting. No solo importa el error sino la confianza en la clasificación.
- La confianza se mide como $y_t f(x_t) \in (-1, 1)$ donde $f = \sum_{t=1}^T \alpha_t h_t(x)$ y α se ha normalizado para que sumen 1.
- En este ejemplo la confianza aumenta con la complejidad.