Lab-8

Bafna Gold
Date: Page:

* Implement Boosting ensemble method on a given dataset.

CGPA	Actual Class	Job profile	Initial wt	Updated wt
>=9	Yes	Yes	1/6	0.1249
<9	No	Yes	1/6	0.2501
>=9	Yes	No	1/6	0.2501
<9	No	No	1/6	0.1249
>=9	Yes	Yes	1/6	0.1249
>=9	Yes	Yes	1/6	0.1249

Step-1: Assign initial weight = 1/6

Step-2: a) $\cdot \Sigma_i = \sum_{j=1}^{n} H_i(d_j) \, wt(d_j)$

Weighted error Σ_{CGPA}

$\Sigma_{CGPA} = 2 \times \frac{1}{6} = 0.333$

b) Compute weight of each weak classifier:

$\alpha_{CGPA} = \frac{1}{2} \ln\left(\frac{1-\Sigma_i}{\Sigma_i}\right)$

$= \frac{1}{2} \ln\left(\frac{1-0.33}{0.33}\right)$.

$\alpha_{CGPA} = 0.347$

c) Calculate the normalizing factor

$Z_{CGPA} = \frac{1}{6} \times 4 \times e^{-0.347} + \frac{1}{6} * 2 \times e^{+0.347}$

$Z_{CGPA} = 0.9428$

d) Update the weight of all data instances

$$wt(d_j)_{i+1} = \frac{wt(d_j)_{CGPA\ of\ correct\ instance} * e^{-\alpha_{CGPA}}}{Z_{CGPA}}$$

$$= \frac{1/6 * e^{-0.347}}{0.9428}$$

$$= 0.1249$$

$$= \frac{1/6 * e^{0.347}}{0.9428}$$

$$= 0.2501$$

→ For "income.csv" dataset

Best Accuracy score : 0.86 or 86%.

Confusion matrix :

TN	FP
[6873	489
FN	TP
845	[328]