Chaudhary Machine Learning PS # 6

Ashwin Acharya

May 18, 2016

1 Question 5

1.1 Formulae

Consider K classifiers with error rates ϵ, and assume their errors are all independent. What is the probability that simple majority voting assigns the wrong class?

Assuming two classes for convenience, this corresponds to the probability that more than half of the classifiers assign the wrong class to a sample. Since the book does not specify how ties are dealt with, I will consider them to assign a class randomly, giving them half as much weight as cases with a true majority of misclassifications.

In general, we want to find the sum $P_{bad} = \Sigma_{n=n_0}^{K} p(n \text{ misclassifications}) * w(n)$, where the weight w is $\frac{1}{2}$ for ties and 1 otherwise, and $n_0 = \frac{K}{2}$ for K even or $\frac{K+1}{2}$ for K odd.

In general, $p(n) = \epsilon^n (1-\epsilon)^{K-n} \frac{K!}{n!(K-n)!}$.

Then, for odd K, $P_{bad} = \Sigma_{n=\frac{K+1}{2}}^{K} \epsilon^n (1-\epsilon)^{K-n} \frac{K!}{n!(K-n)!}$.

For even K, $P_{bad} = \frac{1}{2} * (\epsilon * (1-\epsilon))^{K/2} \frac{K!}{\frac{K}{2}!^2} + \Sigma_{n=\frac{K}{2}+1}^{K} \epsilon^n (1-\epsilon)^{K-n} \frac{K!}{n!(K-n)!}$.

1.2 Examples

- $K = 5, \epsilon = 0.1; P_{bad} = 0.00856$
- $K = 5, \epsilon = 0.2; P_{bad} = 0.05792$
- $K = 5, \epsilon = 0.4; P_{bad} = 0.31744$
- $K = 10, \epsilon = 0.1; P_{bad} = 0.00089092$
- $K = 10, \epsilon = 0.2; P_{bad} = 0.01958144$
- $K = 10, \epsilon = 0.4; P_{bad} = 0.26656768$
- $K = 20, \epsilon = 0.1; P_{bad} = 3.92988232713e-06$
- $K = 20, \epsilon = 0.2; P_{bad} = 0.00157912054917$
- $K = 20, \epsilon = 0.4; P_{bad} = 0.186092021415$