Diverse density at a point in the feature space is defined as a probabilistic measure of the density of both positive instances and negative ones [1,2]. The DD attempts to find the optimal value(also known as target concept), h, with the maximum diverse diversity and the maximum likelihood estimator of h, h_{DD}, is defined as:

$$h_{DD} = \arg\max_{h} \prod_{i} \Pr(B_i \mid h) = \arg\min_{h} \sum_{i} -\log \Pr(B_i \mid h) \tag{6}$$

$$\Pr(B_i \mid h) = \exp(-\mathrm{d}) \tag{7}$$

$$\mathrm{d} = s^2 (B_{ij} - h)^2 \tag{8}$$

where B_i^{+}, B_i^{-}, B_{ij} and s, respectively denote positive bags, negative bags, instances and the parameter representing the importance of the feature.

In 2001, Expectation-Maximization (EM) approach was combined with DD by zhang et al to simplify the search step and decrease the computation complexity. First, k (k is assumed to be 10)instances, h_k, as initial values of h are selected randomly from positive bags. Second(E-Step), a set of instances, B_{ij}, are found to represent their bags by maximizing $\Pr(B_i|h_k)$. Third(M-step), according to formula(6), possible target concept, h_{DDk}, can be determined by using gradient descent. Fourth, h_{DDk}, is used as a new initial value and then the last two steps are performed repeatedly until $\Pr(B_i|h)$ is maximized. Finally, the labels of the bags, $B_i^{'}$, in the testing set are estimated using formula(9).

$$\Pr(B_i^{'} \mid h) = \frac{1}{k} \sum_{1}^{k} \Pr(B_i^{'} \mid h_{DDk}) \tag{9}$$

[1] Qi, Zhang, and S. A. Goldman. "EM-DD: An Improved Multiple-Instance Learning Technique (Preliminary Version)." 2001.

[2] Maron, O. & Lozano-Perez, T. A framework for multiple-instance learning. Neural Information Processing Systems 10. Cambridge, MA: MIT Press,1998.