Gini index : (impurity measure)

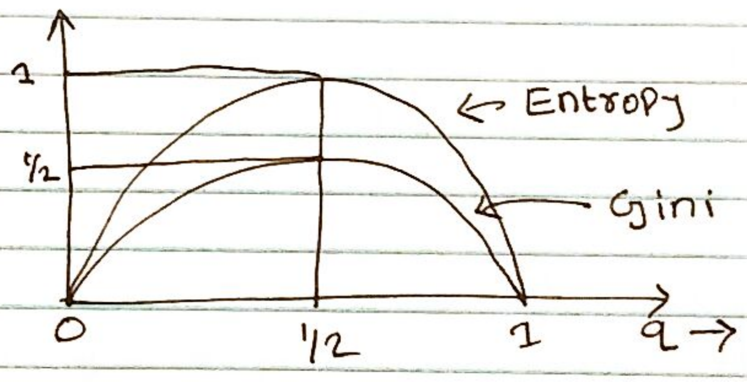
Given ~~a~~ Probability ~~P~~ $(P_1, P_2, \ldots P_k)$ on finite set

$$Gini(P) \triangleq \sum_i P_i(1-P_i) = 1 - \sum_i P_i^2$$

Example :
binary set with $P_1 = q$, $P_2 = 1-q$

$$Gini(P) = 2q(1-q)$$

$$Entropy(P) = -q\log q - (1-q)\log(1-q)$$



Example :
Playing Tennis :

$S = [9+, 5-]$

Wind $\in$ { Weak, Strong }

$S_{weak} = [6+, 2-]$

$S_{strong} = [3+, 3-]$

$$Gini(S) = 1 - \left(\frac{5}{14}\right)^2 - \left(\frac{9}{14}\right)^2$$

$$= 0.459$$

$$Gini(S_{weak}) = 1 - \left(\frac{6}{8}\right)^2 - \left(\frac{2}{8}\right)^2$$

$$= 0.375$$

$$Gini(S_{strong}) = 0.5$$

$$Gain_G(S, wind) = Gini(S)$$

$$- \sum_{v \in \{strong, weak\}} \frac{|S_v|}{|S|} \, Gini(S_v)$$

$$= 0.495 - \frac{8}{14} \times 0.375 - \frac{6}{14} \times 0.5$$

$$= 0.066$$

Gaussian NB as a linear model:

Two classes $\{+1, -1\}$

Given 'n' training data

$(\bar{x}(1), f(\bar{x}(1)), \ldots (\bar{x}(n), f(\bar{x}(n))$

Estimate mean conditioned on class

$\{\hat{\mu}_{j|+1}, \hat{\mu}_{j|-1}\}$

Variances do not depend on class

$\hat{\sigma}_{j|+1} = \hat{\sigma}^2_{j|-1} = \hat{\sigma}_j^2$ (suppose given)

Gaussian NB:

$$p(\bar{x}|c) = \prod_{j=1}^{K} p(\bar{x}_j|c)$$

$$= \prod_{j=1}^{K} \frac{1}{\sqrt{2\pi\hat{\sigma}_j^2}} \exp\left(-\frac{(\bar{x}_j - \hat{\mu}_{j|c})^2}{2\hat{\sigma}_j^2}\right)$$

Decision Rule:

$$P(C=+1)\,P(\bar{x}\mid +1) \underset{-1}{\overset{+1}{\gtrless}} P(C=-1)\,P(\bar{x}\mid -1)$$

$$\Leftrightarrow \log P(C=+1) + \log P(\bar{x}\mid +1) \underset{-1}{\overset{+1}{\gtrless}} \log P(C=-1) + \log P(\bar{x}\mid -1)$$

$$\therefore \log P(C=+1) - \sum_{j=1}^{K} \frac{\bar{x}_j^2 - 2x_j\mu_{j|+1} + \mu_{j|+1}^2}{2\hat{\sigma}_j^2}$$

$$\underset{-1}{\overset{+1}{\gtrless}} \log P(C=-1) - \sum_{j=1}^{K} \left(\quad \right)$$

$$\sum_{j=1}^{K} \bar{x}_j \left(\frac{\hat{\mu}_{j|+1} - \hat{\mu}_{j|-1}}{\hat{\sigma}_j^2} \right)$$

$$+ \log \frac{P(+1)}{P(-1)} - \sum_{j=1}^{K} \frac{\hat{\mu}_{j|+1}^2 - \hat{\mu}_{j|-1}^2}{\hat{\sigma}_j^2}$$

$$\underset{-1}{\overset{+1}{\gtrless}} 0$$