5.15

C

5.16

(1)

$$egin{aligned} H(D) &= -\sum_{i=1}^n p(x_i) log p(x_i) \ &= -rac{4}{9} log rac{4}{9} - rac{5}{9} log rac{5}{9} \ &= rac{4}{9} log rac{9}{4} + rac{5}{9} log rac{9}{5} \end{aligned}$$

(2)

由

$$\begin{split} H(D) &= \frac{4}{9}log\frac{9}{4} + \frac{5}{9}log\frac{9}{5} \\ H(D|a_1) &= \frac{4}{9}(-\frac{3}{4}log\frac{3}{4} - \frac{1}{4}log\frac{1}{4}) + \frac{5}{9}(-\frac{1}{5}log\frac{1}{5} - \frac{4}{5}log\frac{4}{5}) \\ H(D|a_2) &= \frac{5}{9}(-\frac{2}{5}log\frac{2}{5} - \frac{3}{5}log\frac{3}{5}) + \frac{4}{9}(-\frac{2}{4}log\frac{2}{4} - \frac{2}{4}log\frac{2}{4}) \\ H(D|a_3) &= \frac{4}{9}(-\frac{3}{4}log\frac{3}{4} - \frac{1}{4}log\frac{1}{4}) + \frac{5}{9}(-\frac{1}{5}log\frac{1}{5} - \frac{4}{5}log\frac{4}{5}) \end{split}$$

所以 a_1, a_2, a_3 的信息增益分别是

$$H(D) - H(D|a_1)$$

 $H(D) - H(D|a_2)$
 $H(D) - H(D|a_3)$

5.17

B,C,D