

Quiz#8 of ESE 417 FL2024 (11/19, 12:35PM~12:50PM)

Consider the following training data set S : 16

x_1	x_2	x_3	y
+	+	+	yes
+	+	+	yes
+	+	-	yes
+	+	-	yes
+	-	+	yes
+	-	+	yes
+	+	-	yes
-	+	-	yes
+	+	+	yes
+	+	+	yes
+	+	-	yes
-	+	-	yes
+	-	+	no
-	-	+	no
-	-	-	no
-	-	-	no

12

4

Each instance in the data set has three features x_1, x_2 and x_3 and a target value y . The features take values of "+" and "-", and y takes values "yes" and "no". Calculate the information gains of features x_1 and x_3 respectively.

$$\text{Entropy}(S) = -\frac{12}{16} \log_2 \frac{12}{16} - \frac{4}{16} \log_2 \frac{4}{16} = 0.8113$$

For x_1 , $\text{Value}(x_1) = "+", "-"$

$$S_{x_1=+} = \{10 \text{ yes}, 1 \text{ no}\}, S_{x_1=-} = \{2 \text{ yes}, 3 \text{ no}\}$$

$$\text{Entropy}(S_{x_1=+}) = -\frac{10}{11} \log_2 \frac{10}{11} - \frac{1}{11} \log_2 \frac{1}{11} = 0.4395$$

$$\text{Entropy}(S_{x_1=-}) = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.9710$$

$$\begin{aligned} \text{Gain}(S, x_1) &= \text{Entropy}(S) - \frac{|S_{x_1=+}|}{|S|} \text{Entropy}(S_{x_1=+}) - \frac{|S_{x_1=-}|}{|S|} \text{Entropy}(S_{x_1=-}) \\ &= 0.8113 - \frac{11}{16} \times 0.4395 - \frac{5}{16} \times 0.9710 = 0.2057 \end{aligned}$$

For X_3 , $Value(X_3) = "+"$, $"-"$

$$S_{X_3=+} = \{6 \text{ yes}, 2 \text{ no}\}, S_{X_3=-} = \{6 \text{ yes}, 2 \text{ no}\}$$

$$\text{Entropy}(S_{X_3=+}) = -\frac{6}{8} \log_2 \frac{6}{8} - \frac{2}{8} \log_2 \frac{2}{8} = 0.8113$$

$$\text{Entropy}(S_{X_3=-}) = 0.8113$$

$$\begin{aligned} \text{Gain}(S, X_3) &= \text{Entropy}(S) - \frac{|S_{X_3=+}|}{|S|} \text{Entropy}(S_{X_3=+}) - \frac{|S_{X_3=-}|}{|S|} \text{Entropy}(S_{X_3=-}) \\ &= 0.8113 - \frac{8}{16} \times 0.8113 - \frac{8}{16} \times 0.8113 = 0 \end{aligned}$$