Franklin Hu, Sunil Pedapudi
Assignment 3

1. Entropy and Information Gain

   (a)

$$\begin{aligned}
H(S) =& B(\frac{p}{p+n}) \\
=& -\frac{p}{p+n} \cdot log(\frac{p}{p+n}) - (1 - \frac{p}{p+n}) \cdot log(1 - \frac{p}{p+n}) \\
=& -\frac{p}{p+n} \cdot log(\frac{p}{p+n}) - \frac{n}{p+n} \cdot log(\frac{n}{p+n}) \\
=& \frac{1}{p+n} \cdot \left(-p \cdot log(\frac{p}{p+n}) - n \cdot log(\frac{n}{p+n}) \right)
\end{aligned}$$

Since $\frac{p}{p+n} \in [0,1]$, $log(\frac{p}{p+n}) \in (-\infty, 0]$ and$-log(\frac{p}{p+n}) \in [0, \infty)$. Because $log(\frac{p}{p+n}) < 1$, $H(S)$ must be greater than some function were we substitute 1 for log:

$$\begin{aligned}
H(S) \leq& \frac{1}{p+n}(-p \cdot 1 - n \cdot 1) \\
H(S) \leq& \frac{p+n}{p+n} \\
H(S) \leq& 1
\end{aligned}$$

When $p = n$:

$$\begin{aligned}
H(S) =& B(\frac{p}{p+p}) = B(0.5) \\
=& -0.5 \cdot log(0.5) - 0.5 \cdot log(0.5) \\
=& -log(0.5) \\
=& 1
\end{aligned}$$

2. Empirical Loss and Splits
3. Splitting continuous attributes
4. Majority voting

   (a) Suppose we have K predictions X_k which predict some true value μ

with an error $\epsilon_k \sim N(0, \sigma^2)$ so $X_k = \mu + \epsilon_k$. Then,

$$\begin{aligned}\overline{X} =& \frac{1}{k} \cdot \sum_k X_k \\ \text{Var}(\overline{X}) =& \text{Var}(\frac{1}{k} \cdot \sum_k X_k) \\ =& \frac{1}{k^2} \text{Var}(\sum_k X_k)\end{aligned}$$

Since the errors are independent, $\text{Var}(\sum_k X_k) = \sum_k \text{Var}(X_k)$

$$\begin{aligned}\text{Var}(\overline{X}) =& \frac{1}{k^2} \sum_k \text{Var}(X_k) \\ =& \frac{1}{k^2} \cdot k\sigma^2 \\ =& \frac{\sigma^2}{k}\end{aligned}$$

(b) Consider the case when the ensemble learning algorithm uses hypotheses that are entirely dependent. In this case, the error would be equal to ϵ. Thus, the error of ensemble learning where the independence assumption is removed is never greater than ϵ.