# Solution to Homework 5

Shoeb Mohammed and Zhuo Chen

April 17, 2016

## 1 Decision trees, entropy and information gain

### 1.1

$$\begin{aligned} H\left(\frac{p}{p+n}\right) &= \frac{p}{p+n} log \frac{p+n}{p} + \frac{n}{p+n} log \frac{p+n}{n} \\ &\leq log\left(\frac{p}{p+n}.\frac{p+n}{p} + \frac{n}{p+n}.\frac{p+n}{n}\right) \quad \text{since } log(.) \text{ is concave function} \\ &= log 2 \\ &= 1 \end{aligned} \tag{1}$$

When $p = n$, the formula gives $H(S) = \frac{1}{2}log2 + \frac{1}{2}log2 = 1$.

### 1.2

- Misclassification rate for A = $\frac{1}{4}$
- Miscalssificationn rate for B = $\frac{1}{4}$
- Entropy gain model A = $\frac{3}{4}log3 - 1 \sim 0.1887$
- Entropy gain model B = $\frac{3}{2} - \frac{3}{4}log3 \sim 0.3113$.
  This means entropy after split is lower for model B.
- Gini index model A = $\frac{3}{8}$
- Gini index model B = $\frac{1}{9}$

### 1.3

Yes, it is possible. For an example, consider a dataset with 700 examples of class $C_1$ and 100 of class $C_2$. If a feature splits it into two leaves $(200, 200)$ and $(200, 200)$; then misclassification rate is bigger.

# 2 Bagging

## 2.1

*Proof.* Simplifying, we get $\epsilon_{bag}(x) = \frac{1}{L} \sum\limits_{l=1}^{L} \epsilon_l(x)$. Thus,

$$
\begin{aligned}
E_{bag} &= E_X \left[\epsilon_{bag}(x)^2\right] \\
&= \frac{1}{L^2} E_X \left[\left(\sum_{l=1}^{L} \epsilon_l(x)\right)^2\right] \\
&= \frac{1}{L^2} E_X \left[\sum_{l=1}^{L} \epsilon_l^2(x) + \sum_{\substack{1\le i,j\le L \\ i\ne j}} \epsilon_i(x)\epsilon_i(x)\right] \\
&= \frac{1}{L^2} E_X \left[\sum \_l = 1^L \epsilon_l^2(x)\right] + \frac{1}{L^2} E_X \left[\sum_{\substack{1\le i,j\le L \\ i\ne j}} \epsilon_i(x)\epsilon_j(x)\right] \\
&= \frac{1}{L^2} E_X \left[\sum_{l=1}^{L} \epsilon_l^2(x)\right] + \frac{1}{L^2} \sum_{\substack{1\le i,j\le L \\ i\ne j}} E_X \left[\epsilon_i(x)\epsilon_j(x)\right] \\
&= \frac{1}{L^2} E_X \left[\sum_{l=1}^{L} \epsilon_l^2(x)\right] \quad \text{since } E_X \left[\epsilon_i(x)\epsilon_j(x)\right] = 0 \text{ for } i \ne j \\
&= \frac{1}{L^2} \sum_{l=1}^{L} E_X \left[\epsilon_l^2(x)\right] \\
&= \frac{1}{L} E_{avg}
\end{aligned}
\tag{2}
$$

$\square$

### 2.2

*Proof.*

$$
\begin{aligned}
E_{bag} &= E_X\left[\epsilon_{bag}(x)^2\right] \\
&= E_X\left[\left(\sum_{l=1}^{L}\frac{\epsilon_l(x)}{L}\right)^2\right] \\
&\le E_X\left[\sum_{l=1}^{L}\frac{\epsilon_l^2(x)}{L}\right] \text{ using Jensen's inequality with } \lambda_i = \frac{1}{L}. \text{ For random variables } 0 \le U \le V;\ E[U] \le E[V] \\
&= \frac{1}{L}\sum_{l=1}^{L} E_X\left[\epsilon_l^2(x)\right] \\
&= E_{avg}
\end{aligned}
\tag{3}
$$

$\square$