Written Answers:

Question 1

WRITTEN EXERCISES.

1. Decision trees.

(a) Suppose $p_1 > n_1$, $p_2 < n_2$, then we have a tree like this:

$$\begin{tikzcd}[column sep=small] & (p_1+p_2,\ n_1+n_2) \arrow[dl] \arrow[dr] & \\ \underset{+}{(p_1, n_1)} & & \underset{-}{(p_2, n_2)} \end{tikzcd}$$

for training mistake we have: $n_1 + p_2$

for weighted impurity, we have:

$$(p_1+n_1)\cdot I\left(\frac{p_1}{p_1+n_1}\right) + (p_2+n_2)\cdot I\left(\frac{p_2}{p_2+n_2}\right)$$

$$= (p_1+n_1)\cdot\frac{n_1}{p_1+n_1} + (p_2+n_2)\frac{p_2}{p_2+n_2} = n_1+p_2$$, the same as training mistake.

thus the min-error impurity is equivalent to grow the tree greedily to minimize training error.

(b) We have trees:

$$\begin{tikzcd}[column sep=tiny] & a_1 \arrow[dl, "0"'] \arrow[dr, "1"] & & & a_2 \arrow[dl, "0"'] \arrow[dr, "1"] & & & a_3 \arrow[dl, "0"'] \arrow[dr, "1"] & \\ \underset{+\ -}{(2,2)} & & \underset{+\ -}{(1,5)} & \underset{+\ -}{(1,3)} & & \underset{+\ -}{(2,4)} & \underset{+\ -}{(3,4)} & & \underset{+\ -}{(0,3)} \end{tikzcd}$$

Gini index for a_1: $\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{6}\cdot\frac{5}{6} = \frac{7}{18} \approx 0.389$

a_2: $\frac{1}{4}\cdot\frac{3}{4}+\frac{1}{3}\frac{2}{3} = \frac{59}{144} \approx 0.410$

a_3: $\frac{3}{7}\cdot\frac{4}{7}+0\cdot 1 = \frac{12}{49} \approx 0.245$ (smallest)

min-error impurity for a_1: $2+1=3$

a_2: $1+2=3$

a_3: $3+0=3$

thus a_3 will be chosen at root for Gini index, while either of a_1, a_2, a_3 could be chosen by min-error impurity

(c). Same tree as that in (a).

for min-error, before making the split, is either p_1+p_2 or n_1+n_2.

for weighted impurity of the split:

if $p_1>n_1$, $p_2>n_2$ or $p_1<n_1$, $p_2<n_2$, it is the same as min-error (p_1+p_2 or n_1+n_2)

if $p_1>n_1$, $p_2<n_2$, it will be n_1+p_2 which is smaller than p_1+p_2 or n_1+n_2.

if $p_1<n_1$, $p_2>n_2$, it will be n_2+p_1, still smaller than p_1+p_2 or n_1+n_2.

thus the general condition will be $(p_1>n_1, p_2<n_2)$ or $(p_1<n_1, p_2>n_2)$.

(d) The answer of (b) and (c) suggest that min-error is suitable for growing a tree under some special case, but it should not be the best way to grow a tree (as a greedy method that always find local optimum).

Question 2

2. Bootstrap aggregation:

Since for N samples, we are drawing N samples with replacement $\Rightarrow$ each draw is independent

for one sample, the probability of not been selected in one draw is:

$$\frac{N-1}{N} = 1-\frac{1}{N}$$

$\Rightarrow$ for N times, the probability of one sample not been selected is:

$$(1-\frac{1}{N})^N$$

$\Rightarrow$ the fraction of samples does not appear at all is:

$$\frac{N\cdot(1-\frac{1}{N})^N}{N} = (1-\frac{1}{N})^N$$

The limit of this expectation as $N\to\infty$ is:

$$\lim_{n\to\infty}(1-\frac{1}{N})^N = \lim_{n\to\infty}\left[\left(1+\frac{1}{(-N)}\right)^{(-N)}\right]^{(-1)}$$

$$= e^{-1}$$

$$= \frac{1}{e}$$