hw4_sol Haoyuan Du 5500151
du000078@umn.edu

1. a. (-2, -1). (0, 1) (0.5, -0.5)

b. ① (0, 1) hyperplane: $x_1 - x_2 = 0$ ② (-2, -1) ③ (1, 3)

① $\because d^t = \frac{|w^T x^t + w_0|}{||w||} = \frac{|1\cdot 0 + 1\cdot(-1) + 0|}{\sqrt{1^2+1^2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$

② $d^t = \frac{|(-2)\cdot 1 + (-1)\cdot(-1) + 0|}{\sqrt{2}} = \frac{\sqrt{2}}{2}$

③ $d^t = \frac{|1\cdot 1 + 3\cdot(-1) + 0|}{\sqrt{2}} = \sqrt{2}$

c. Yes, it will change (broader). If remove (-1, -2.5), it won't change.

d. Yes. I'll use soft margin and slack variables.

e. When C is small, classification mistakes are give less importance and focus is more on maximizing the margin. When C is large, the focus is more on avoiding misclassification at the expense of keeping the margin small.

f. I'd prefer soft margin rather than hard margin. Because if choose hard margin, even one outlier can determine the boundary which makes classifier overly sensitive to noise. Then, use linear SVM for linear problems, and kernel for non-linear problems.

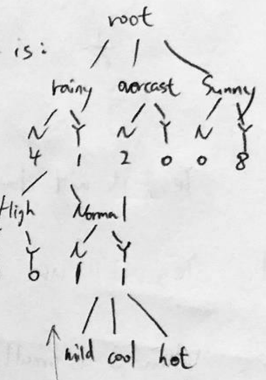
2. a.

$$\begin{cases} \text{High} & 3\ N,\ 2\ Y \\ \text{Normal} & 3\ N,\ 7\ Y \end{cases} \quad \text{Entropy}: -\left(\frac{5}{15}\times\frac{3}{5}\log\frac{3}{5}+\frac{5}{15}\times\frac{2}{5}\log\frac{2}{5}+\frac{10}{15}\times\frac{3}{10}\log\frac{3}{10}+\frac{10}{15}\times\frac{7}{10}\log\frac{7}{10}\right)=0.9112$$

$$\begin{cases} \text{Mild} & 2\ N,\ 4\ Y \\ \text{Cool} & 3\ N,\ 3\ Y \\ \text{hot} & 1\ N,\ 2\ Y \end{cases} \quad \text{Entropy}: -\left(\frac{6}{15}\cdot\frac{2}{6}\log\frac{2}{6}+\frac{6}{15}\frac{4}{6}\log\frac{4}{6}+\frac{6}{15}\frac{1}{2}\log\frac{1}{2}+\frac{6}{15}\frac{1}{2}\log\frac{1}{2}+\frac{3}{15}\frac{1}{3}\log\frac{1}{3}+\frac{3}{15}\frac{2}{3}\log\frac{2}{3}\right)=0.951$$

$$\begin{cases} \text{Rainy} & 4\ N,\ 1\ Y \\ \text{Overcast} & 2\ N,\ 0\ Y \\ \text{Sunny} & 0\ N,\ 8\ Y \end{cases} \quad \text{Entropy}: -\left(\frac{5}{15}\frac{4}{5}\log\frac{4}{5}+\frac{5}{15}\frac{1}{5}\log\frac{1}{5}+0\right)=0.2406$$

Conclusion 0.2406 is the best. So the first layer is:



$$\begin{cases} \text{High} & 3\ N,\ 0\ Y \\ \text{Normal} & 1\ N,\ 1\ Y \end{cases} \quad E: -\left(\frac{2}{5}\cdot\frac{1}{2}\log\frac{1}{2}+\frac{2}{5}\frac{1}{2}\log\frac{1}{2}\right)=0.4$$

$$\begin{cases} \text{mild} & 1\ N,\ 1\ Y \\ \text{cool} & 2\ N,\ 0\ Y \\ \text{hot} & 1\ N,\ 0\ Y \end{cases} \quad E: -\left(\frac{2}{5}\cdot\frac{1}{2}\log\frac{1}{2}+\frac{2}{5}\frac{1}{2}\log\frac{1}{2}+0\right)=0.4$$

Thus, choose less braches one, which is High/Normal.

The next layer is Cool, mild, hot. (This layer could be overfit ※).

$$\begin{cases} \text{mild} & 0\ N,\ 1\ Y \\ \text{cool} & 1\ N,\ 0\ Y \\ \text{hot} & 0\ N,\ 0\ Y \end{cases} \quad E=0$$

b. ①: No. $\frac{2}{2}>0$

②: No. $\frac{4}{5}>\frac{1}{5}$

3. a.

```
Training/validation accuracy for minimum node entropy 0.010000 is 1.000 / 0.863
Training/validation accuracy for minimum node entropy 0.050000 is 0.999 / 0.863
Training/validation accuracy for minimum node entropy 0.100000 is 0.997 / 0.865
Training/validation accuracy for minimum node entropy 0.200000 is 0.990 / 0.867
Training/validation accuracy for minimum node entropy 0.400000 is 0.979 / 0.861
Training/validation accuracy for minimum node entropy 0.800000 is 0.919 / 0.856
Training/validation accuracy for minimum node entropy 1.000000 is 0.871 / 0.840
Training/validation accuracy for minimum node entropy 2.000000 is 0.596 / 0.600
Test accuracy with minimum node entropy 0.200000 is 0.872
Press any key to continue . . .
```

Use $\theta = 0.2$, the accuracy is 0.872.

b.

the less the θ is, the more complex the tree is. Although when $\theta = 0.01$, the training accuracy and validation is high, but it’s overfitting. To avoid overfitting, we need to choose the tree that has the highest test accuracy.