CSE 6363 - *Machine Learning*

Homework 2- Spring 2019 - Sample Solution

Due Date: Mar. 16, 2019

Support Vector Machines

1. Consider the following linearly separable training data set:

$$
\begin{aligned}
D = \{ \quad & ((1,2), \quad -1), \\
& ((2,3), \quad 1), \\
& ((2,1), \quad -1), \\
& ((3,4), \quad 1), \\
& ((1,3), \quad -1), \\
& ((4,4), \quad 1) \qquad \}
\end{aligned}
$$

a) Formulate the optimization function as well as the constraints for the corresponding linear maximum margin optimization problem without a regularization term. Also show the corresponding Lagrangian as well as the Lagrangian Dual for this problem.

The general formulation of the optimization problem for the linear maximum margin optimization problem is:

$$
\begin{array}{ll}
Optimization\ Function: & min_{w,b} \frac{1}{2}||w||^2 \\
Constraints: & y^{(i)}(w^T x^{(i)} + b) \geq 1 \quad i = 1...n
\end{array}
$$

For this problem, this translates to:

$$
\begin{array}{ll}
Optimization\ Function: & min_{w,b} \frac{1}{2}(w_1^2 + w_2^2) \\
Constraints: & -(w_1 + 2w_2 + b) \geq 1 \\
& (2w_1 + 3w_2 + b) \geq 1 \\
& -(2w_1 + w_2 + b) \geq 1 \\
& (3w_1 + 4w_2 + b) \geq 1 \\
& -(w_1 + 3w_2 + b) \geq 1 \\
& (4w_1 + 4w_2 + b) \geq 1
\end{array}
$$

The corresponding Lagrangian in general is:

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

For this specific problem, this translates into:

$$\begin{aligned}\mathcal{L}(w, b, \alpha) = \;& \tfrac{1}{2}(w_1^2 + w_2^2) - \alpha_1 \left[-(w_1 + 2w_2 + b) - 1 \right] - \alpha_2 \left[(2w_1 + 3w_2 + b) - 1 \right] \\ & -\alpha_3 \left[-(2w_1 + w_2 + b) - 1 \right] - \alpha_4 \left[(3w_1 + 4w_2 + b) - 1 \right] \\ & -\alpha_5 \left[-(w_1 + 3w_2 + b) - 1 \right] - \alpha_6 \left[(4w_1 + 4w_2 + b) - 1 \right]\end{aligned}$$

The Lagrangian Dual problem is defined as:

$$\begin{array}{lll} Optimization\ Function: & max_\alpha \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j \left\langle x^{(I)}, x^{(j)} \right\rangle & \\ Constraints: & \alpha_i \geq 0 & i = 1...n \\ & \sum_{i=1}^{n} \alpha_i y^{(i)} = 0 & \end{array}$$

For this problem, this translates to:

$$\begin{array}{ll} Optimization\ Function: & max_\alpha \sum_{i=1}^{6} \alpha_i - \frac{1}{2}(\alpha_1(5\alpha_1 - 8\alpha_2 + 4\alpha_3 - 11\alpha_4 + 7\alpha_5 - 12\alpha_6) \\ & \quad + \alpha_2(-8\alpha_1 + 13\alpha_2 - 7\alpha_3 + 18\alpha_4 - 11\alpha_5 + 20\alpha_6) \\ & \quad + \alpha_3(4\alpha_1 - 7\alpha_2 + 5\alpha_3 - 10\alpha_4 + 5\alpha_5 - 12\alpha_6) \\ & \quad + \alpha_4(-11\alpha_1 + 18\alpha_2 - 10\alpha_3 + 25\alpha_4 - 15\alpha_5 + 28\alpha_6) \\ & \quad + \alpha_5(7\alpha_1 - 11\alpha_2 + 5\alpha_3 - 15\alpha_4 + 10\alpha_5 - 16\alpha_6) \\ & \quad + \alpha_6(-12\alpha_1 + 20\alpha_2 - 12\alpha_3 + 28\alpha_4 - 16\alpha_5 + 32\alpha_6)) \\ Constraints: & \alpha_i \geq 0 \quad i = 1...n \\ & -\alpha_1 + \alpha_2 - \alpha_3 + \alpha_4 - \alpha_5 + \alpha_6 = 0 \end{array}$$

b) Manually perform 4 iterations of the SMO algorithm on this data. You do not have to use any specific heuristic to pick the two α parameters in each iteration.
Initialize all parameters α_i to 0, which fulfills the dual constraints
For these values we can compute the optimal weights, W, using the dual function:

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0$$

and the offset b from the primal:

$$b = -\frac{max_{i:y^{(i)}=-1} w^T x^{(i)} + min_{i:y^{(i)}=1} w^T x^{(i)}}{2} = 0$$

Iteration 1

- Pick two constraints: e.g.: α_1 and α_2

- Find the optimum parameters for the two constraints:
  Rewriting α_1 in terms of α_2 and considering that we only optimize over these two constraints (thus everything else is constant and thus does not affect the optimization), and realizing that the constraints of the dual can be inherently fulfilled with a re-write of α_1 we get:

$$\xi = -\sum_{i \neq 1,2} \alpha_i y^{(i)} = 0$$

$$\alpha_1 = (\xi - \alpha_2 y^{(2)}) y^{(1)}$$

$$\begin{aligned}
\alpha_2^{new,uc} &= argmax_{\alpha_2} (\xi - \alpha_2 y^{(2)}) y^{(1)} + \alpha_2 \\
&\quad - \tfrac{1}{2}((\xi - \alpha_2 y^{(2)}) y^{(1)} (5(\xi - \alpha_2 y^{(2)}) y^{(1)} - 8\alpha_2 + 4\alpha_3 - 11\alpha_4 + 7\alpha_5 - 12\alpha_6 \\
&\qquad + \alpha_2(-8(\xi - \alpha_2 y^{(2)}) y^{(1)} + 13\alpha_2 - 7\alpha_3 + 18\alpha_4 - 11\alpha_5 + 20\alpha_6) \\
&\qquad + \alpha_3(4(\xi - \alpha_2 y^{(2)}) y^{(1)} - 7\alpha_2) \\
&\qquad + \alpha_4(-11(\xi - \alpha_2 y^{(2)}) y^{(1)} + 18\alpha_2) \\
&\qquad + \alpha_5(7(\xi - \alpha_2 y^{(2)}) y^{(1)} - 11\alpha_2) \\
&\qquad + \alpha_6(-12(\xi - \alpha_2 y^{(2)}) y^{(1)} + 20\alpha_2))
\end{aligned}$$

  Taking into account that all α values except for α_2 are 0 and that ξ is 0, this yields:

$$\begin{aligned}
\alpha_2^{new,uc} &= argmax_{\alpha_2} (-\alpha_2 y^{(2)}) y^{(1)} + \alpha_2 \\
&\quad - \tfrac{1}{2}((-\alpha_2 y^{(2)}) y^{(1)} (5(-\alpha_2 y^{(2)}) y^{(1)} - 8\alpha_2) \\
&\qquad + \alpha_2(-8(-\alpha_2 y^{(2)}) y^{(1)} + 13\alpha_2)) \\
&= argmax_{\alpha_2} 2\alpha_2 - \tfrac{1}{2}(\alpha_2(5\alpha_2 - 8\alpha_2) + \alpha_2(-8\alpha_2 + 13\alpha_2)) \\
&= argmax_{\alpha_2} 2\alpha_2 - \tfrac{1}{2} 2\alpha_2^2 \\
&= argmax_{\alpha_2} 2\alpha_2 - \alpha_2^2 = 1
\end{aligned}$$

  To clip these values we need to calculate H and L:
  The line is $\alpha_1 y^{(1)} + \alpha_2 y^{(2)} = \xi$ and thus we have $\alpha_2 = \alpha_1$ yielding $L = 0$, $H = \infty$
  Thus $\alpha_2 = \alpha_2^{new,clipped} = 1$, $\alpha_1 = \alpha_1^{new,clipped} = 1$
- Now we can resolve for the optimal weights and offset for the new α values:

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = -1(1,2)^T + 1(2,3)^T = (1,1)^T$$

  and the offset b from the primal:

$$b = -\frac{max_{i:y^{(i)}=-1} w^T x^{(i)} + min_{i:y^{(i)}=1} w^T x^{(i)}}{2} = -\frac{4+5}{2} = -4.5$$

Iteration 2

- Pick two constraints: e.g.: α_2 and α_5
- Find the optimum parameters for the two constraints:
  Rewriting α_2 in terms of α_5 and considering that we only optimize over these two

constraints (thus everything else is constant and thus does not affect the optimization), and realizing that the constraints of the dual can be inherently fulfilled with a re-write of α_2 we get:

$$\xi = -\sum_{i \neq 2,5} \alpha_i y^{(i)} = 1$$

$$\alpha_2 = (\xi - \alpha_5 y^{(5)}) y^{(2)}$$

$$\begin{aligned}
\alpha_5^{new,uc} \;=\;& argmax_{\alpha_5} (\xi - \alpha_5 y^{(5)}) y^{(2)} + \alpha_5 \\
& - \tfrac{1}{2}(\alpha_1(-8(\xi - \alpha_5 y^{(5)}) y^{(2)} + 7\alpha_5) \\
& \quad + (\xi - \alpha_5 y^{(5)}) y^{(2)} (-8\alpha_1 + 13(\xi - \alpha_5 y^{(5)}) y^{(2)} - 7\alpha_3 + 18\alpha_4 - 11\alpha \\
& \quad + \alpha_3(-7(\xi - \alpha_5 y^{(5)}) y^{(2)} + 5\alpha_5) \\
& \quad + \alpha_4(18(\xi - \alpha_5 y^{(5)}) y^{(2)} - 15\alpha_5) \\
& \quad + \alpha_5(7\alpha_1 - 11(\xi - \alpha_5 y^{(5)}) y^{(2)} + 5\alpha_3 - 15\alpha_4 + 10\alpha_5 - 16\alpha_6) \\
& \quad + \alpha_6(20(\xi - \alpha_5 y^{(5)}) y^{(2)} - 16\alpha_5))
\end{aligned}$$

Taking into account that all α values except for $\alpha_1 = 1$ and α_5 are 0 and that ξ is 1, as well as that constants do not affect the result of the optimization, this yields:

$$\begin{aligned}
\alpha_5^{new,uc} \;=\;& argmax_{\alpha_5} (1 - \alpha_5 y^{(5)}) y^{(2)} + \alpha_2 \\
& - \tfrac{1}{2}((-8(1 - \alpha_5 y^{(5)}) y^{(2)} + 7\alpha_5) \\
& \quad + (1 - \alpha_5 y^{(5)}) y^{(2)} (-8 + 13(1 - \alpha_5 y^{(5)}) y^{(2)} - 11\alpha_5) \\
& \quad + \alpha_5(7 - 11(1 - \alpha_5 y^{(5)}) y^{(2)} + 10\alpha_5) \\
=\;& argmax_{\alpha_5} 1 + \alpha_5 + \alpha_5 \\
& - \tfrac{1}{2}((-8(1 + \alpha_5) + 7\alpha_5) \\
& \quad + (1 + \alpha_5)(-8 + 13(1 + \alpha_5) - 11\alpha_5) \\
& \quad + \alpha_5(7 - 11(1 + \alpha_5) + 10\alpha_5) \\
=\;& argmax_{\alpha_5} 1 + 2\alpha_5 - \tfrac{1}{2}((-8 - \alpha_5) + (1 + \alpha_5)(5 + 2\alpha_5) + \alpha_5(-4 - \alpha_5)) \\
=\;& argmax_{\alpha_5} 1 + 2\alpha_5 - \tfrac{1}{2}(-3 + 2\alpha_5 + \alpha_5^2) \\
=\;& argmax_{\alpha_5} 2\alpha_5 - \alpha_5 - \tfrac{1}{2}\alpha_5^2 \\
=\;& argmax_{\alpha_5} \alpha_5 - \tfrac{1}{2}\alpha_5^2 = 1
\end{aligned}$$

To clip these values we need to calculate H and L:
The line is $\alpha_2 y^{(2)} + \alpha_5 y^{(5)} = \xi$ and thus we have $\alpha_2 = 1 + \alpha_5$ yielding $L = 0$, $H = \infty$
Thus $\alpha_5 = \alpha_5^{new,clipped} = 1$, $\alpha_2 = \alpha_2^{new,clipped} = 2$

- Now we can resolve for the optimal weights and offset for the new α values:

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = -1(1,2)^T + 2(2,3)^T - 1(1,3)^T = (2,1)^T$$

and the offset b from the primal:

$$b = -\frac{max_{i:y^{(i)}=-1} w^T x^{(i)} + min_{i:y^{(i)}=1} w^T x^{(i)}}{2} = -\frac{5+7}{2} = -6$$

Iteration 3

- Pick two constraints: e.g.: α_1 and α_3
- Find the optimum parameters for the two constraints:
  Rewriting α_1 in terms of α_3 and considering that we only optimize over these two constraints (thus everything else is constant and thus does not affect the optimization), and realizing that the constraints of the dual can be inherently fulfilled with a re-write of α_1 we get:

$$\xi = -\sum_{i\neq 1,3} \alpha_i y^{(i)} = -1$$

$$\alpha_1 = (\xi - \alpha_3 y^{(3)}) y^{(1)} = (1 - \alpha_3) =$$

$$\begin{aligned}
\alpha_3^{new,uc} &= argmax_{\alpha_3} (1-\alpha_3) + \alpha_3 \\
&\qquad - \tfrac{1}{2}((1-\alpha_3)(5(1-\alpha_3) - 8\alpha_2 + 4\alpha_3 - 11\alpha_4 + 7\alpha_5 - 12\alpha_6) \\
&\qquad\qquad + \alpha_2(-8(1-\alpha_3) - 7\alpha_3) \\
&\qquad\qquad + \alpha_3(4(1-\alpha_3) - 7\alpha_2 + 5\alpha_3 - 10\alpha_4 + 5\alpha_5 - 12\alpha_6) \\
&\qquad\qquad + \alpha_4(-11(1-\alpha_3) - 10\alpha_3) \\
&\qquad\qquad + \alpha_5(7(1-\alpha_3) + 5\alpha_3) \\
&\qquad\qquad + \alpha_6(-12(1-\alpha_3) - 12\alpha_3))
\end{aligned}$$

  Taking into account that all α values except for $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_5 = 1$, and α_3 are 0 and that ξ is 1, as well as that constants do not affect the result of the optimization, this yields:

$$\begin{aligned}
\alpha_3^{new,uc} &= argmax_{\alpha_3} 1 \\
&\qquad - \tfrac{1}{2}((1-\alpha_3)(5(1-\alpha_3) - 8*2 + 4\alpha_3 + 7) \\
&\qquad\qquad + 2(-8(1-\alpha_3) - 7\alpha_3) \\
&\qquad\qquad + \alpha_3(4(1-\alpha_3) - 7*2 + 5\alpha_3 + 5) \\
&\qquad\qquad + (7(1-\alpha_3) + 5\alpha_3) \\
&= argmax_{\alpha_3} - \tfrac{1}{2}((1-\alpha_3)(5 - 5\alpha_3 - 8*2 + 4\alpha_3 + 7) \\
&\qquad\qquad + 2(-8 + 8\alpha_3 - 7\alpha_3) \\
&\qquad\qquad + \alpha_3(4 - 4\alpha_3 - 7*2 + 5\alpha_3 + 5) \\
&\qquad\qquad + (7 - 7\alpha_3 + 5\alpha_3) \\
&= argmax_{\alpha_3} - \tfrac{1}{2}((1-\alpha_3)(-4-\alpha_3) + 2(-8+\alpha_3) + \alpha_3(-5+\alpha_3) + (7-2\alpha_3)) \\
&= argmax_{\alpha_3} - \tfrac{1}{2}(-13 - 2\alpha_3 + 2\alpha_3^2) \\
&= argmax_{\alpha_3} \alpha_3 - \alpha_3^2 = \tfrac{1}{2}
\end{aligned}$$

  To clip these values we need to calculate H and L:
  The line is $\alpha_1 y^{(1)} + \alpha_3 y^{(3)} = \xi$ and thus we have $\alpha_1 = 1 - \alpha_3$ yielding $L = 0$, $H = 1$
  Thus $\alpha_3 = \alpha_3^{new,clipped} = \frac{1}{2}$, $\alpha_1 = \alpha_1^{new,clipped} = \frac{1}{2}$
- Now we can resolve for the optimal weights and offset for the new α values:

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = -\frac{1}{2}(1,2)^T + 2(2,3)^T - \frac{1}{2}(2,1)^T - 1(1,3)^T = (1.5, 1.5)^T$$

and the offset b from the primal:

$$b = -\frac{max_{i:y^{(i)}=-1} w^T x^{(i)} + min_{i:y^{(i)}=1} w^T x^{(i)}}{2} = -\frac{6+7.5}{2} = -6.75$$

Iteration 4

- Pick two constraints: e.g.: α_1 and α_5
- Find the optimum parameters for the two constraints:
  Rewriting α_1 in terms of α_5 and considering that we only optimize over these two constraints (thus everything else is constant and thus does not affect the optimization), and realizing that the constraints of the dual can be inherently fulfilled with a re-write of α_1 we get:

$$\xi = -\sum_{i \neq 1,5} \alpha_i y^{(i)} = -1.5$$

$$\alpha_1 = (\xi - \alpha_5 y^{(5)}) y^{(1)} = -(-1.5 + \alpha_5) = (1.5 - \alpha_5)$$

$$\begin{aligned}
\alpha_5^{new,uc} &= argmax_{\alpha_5} (1.5 - \alpha_5) + \alpha_5 \\
&\quad - \tfrac{1}{2}((1.5 - \alpha_5)(5(1.5 - \alpha_5) - 8\alpha_2 + 4\alpha_3 + 7\alpha_5) \\
&\qquad + \alpha_2(-8(1.5 - \alpha_5) - 11\alpha_5) \\
&\qquad + \alpha_3(4(1.5 - \alpha_5) + 5\alpha_5) \\
&\qquad + \alpha_4(-11(1.5 - \alpha_5) - 15\alpha_5) \\
&\qquad + \alpha_5(7(1.5 - \alpha_5) - 11\alpha_2 + 5\alpha_3 + 10\alpha_5) \\
&\qquad + \alpha_6(-12(1.5 - \alpha_5) - 16\alpha_5))
\end{aligned}$$

Taking into account that all α values except for α_1, $\alpha_2 = 2$, $\alpha_3 = 0.5$, and α_5 are 0 and that ξ is 1.5, as well as that constants do not affect the result of the optimization, this yields:

$$\begin{aligned}
\alpha_5^{new,uc} &= argmax_{\alpha_5} 1.5 \\
&\quad - \tfrac{1}{2}((1.5 - \alpha_5)(5(1.5 - \alpha_5) - 8 * 2 + 4 * 0.5 + 7\alpha_5) \\
&\qquad + 2(-8(1.5 - \alpha_5) - 11\alpha_5) \\
&\qquad + 0.5(4(1.5 - \alpha_5) + 5\alpha_5) \\
&\qquad + \alpha_5(7(1.5 - \alpha_5) - 11 * 2 + 5 * 0.5 + 10\alpha_5)) \\
&= argmax_{\alpha_5} - \tfrac{1}{2}((1.5 - \alpha_5)(2\alpha_5 - 6.5) + 2(-12 - 3\alpha_5) \\
&\qquad + 0.5(6 + \alpha_5) + \alpha_5(-9 + 3\alpha_5)) \\
&= argmax_{\alpha_5} - \tfrac{1}{2}(-6.75 - 5\alpha_5 + \alpha_5^2) \\
&= argmax_{\alpha_5} (-5\alpha_5 + \alpha_5^2) = 2.5
\end{aligned}$$

To clip these values we need to calculate H and L:
The line is $\alpha_1 y^{(1)} + \alpha_5 y^{(5)} = \xi$ and thus we have $\alpha_1 = 1.5 - \alpha_5$ yielding $L = 0$, $H = 1.5$
Thus $\alpha_5 = \alpha_5^{new,clipped} = 1.5$, $\alpha_1 = \alpha_1^{new,clipped} = 0$

- Now we can resolve for the optimal weights and offset for the new α values:

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 2(2,3)^T - 0.5(2,1)^T - 1.5(1,3)^T = (1.5,1)^T$$

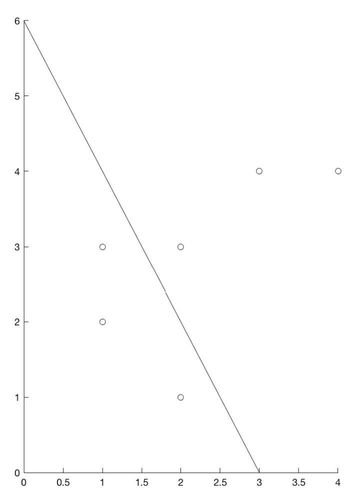
and the offset b from the primal:

$$b = -\frac{max_{i:y^{(i)}=-1} w^T x^{(i)} + min_{i:y^{(i)}=1} w^T x^{(i)}}{2} = -\frac{4.5+6}{2} = -5.25$$

Note that different orders will lead to different progress. The fastest order would be to use only α_2, α_3, and α_5 since these are the support vectors.

c) Use a SVM solver (e.g. MatLab's *fitcsvm* function) to learn the linear SVM parameters for this problem. Show the resulting decision boundary and identify the support vectors in this problem.

Depending on the solver used, the solution either normalizes the weight vector (resulting in $w = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)^T, b = -\frac{6}{\sqrt{5}}$ or the margin (this would be the case in the notation above, yielding $w = (2,1)^T, b = -6$. To plot the decision boundary it is easier to use the normalized weight vector since it is the normal to the line with b being the distance to the origin. This yields the line $\left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right) x = \frac{6}{\sqrt{5}}$ or $x_2 = -2x_1 + 6$



Decision Trees

2. Consider the problem where we want to predict whether a mushroom is edible or poisonous from a set of discrete attributes, namely cap-shape (6 possible values), cap-surface (4 possible values), cap-color (10 possible values), bruises (2 possible values), and odor (9 possible values). Data is given in the files as a comma separated list $\{e, x, s, y, t, a\}$ where the first entry is the class (e

or p), the second is the cap-shape (b, c, x, f, k, or s), the third is the cap-surface (f, g, y, or s), the fourth entry is the cap-color (n, b, c, g, r, p, u, e, w, y, t, or f), the fifth entry is whether it bruises (t, or f), and the last entry is the odor (a, l, c, y, f, m, n, p, or s). There is a training and a test data set for this problem (datasets are derived from the more expansive UCI machine learning mushroom data set).

a) Show the construction of a 2 level decision tree using minimum Entropy as the construction criterion on the training data set. You should include the entropy calculations and the construction decisions for each node you include in the 2-level tree.

   For construct the decision tree we have to incrementally split the dataset according to the attribute with the highest entropy loss.

   Level 1:

   - Compute the entropy loss for each attribute:
     - cap-shape: 0.26764617
     - cap-surface: 0.11127295
     - cap-color: 0.44963301
     - bruises: 0.09643129
     - odor: 1
   - Select node for level 1 split From this we can see that odor perfectly splits the set and we could use this as the first level which would make the second level arbitrary since there is no more entropy left (which you would have to show to fulfill the second level requirement in the assignment).

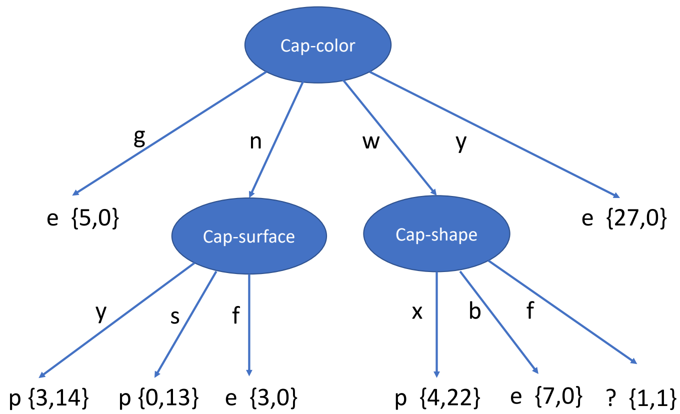
     To make it more interesting, we will ignore this attribute here and assume it were not present. Then we would select the attribute "cap-color" fo the first level.

   Level 2: cap-color splits according to 4 attributes into the following sets:

   - cap-color = g: 5 edible, 0 poisonous, Entropy 0 No further split to be performed.
   - cap-color = n: 6 edible, 27 poisonous, Entropy 0.684038436 For level 2 we consider all non-used attributes (i.e. all except cap-color and odor (since we ignore this one here) and compute their entropy losses:
     - cap-shape: 0.458398307
     - cap-surface: 0.569748317
     - bruises: 0.543339758

     From this we see that the next level for this branch should be cap-surface.
   - cap-color = w: 0.927526588 For level 2 we consider all non-used attributes (i.e. all except cap-color and odor (since we ignore this one here) and compute their entropy losses:
     - cap-shape: 0.746487218
     - cap-surface: 0.650564044
     - bruises: 0.61874595

     From this we see that the next level for this branch should be cap-shape.
   - cap-color = y: 0 No further split to be performed.



b) Implement a decision tree learner for this particular problem and derive the complete tree for the training data set.

c) Apply the tree from part *b*) to the test data set and compare the classification accuracy on this test set with the one on the training set. Does the result indicate overfitting ?

When including the last attribute, the final decision tree only has 1 level (based on odor) and perfectly classifies both the training and test set. There are therefore no indications of overfitting.

When not including the last attribute, we can first notice that there are attributes in the test data that did not occur in the training data. Thus the decision tree does not have an answer for those. Otherwise the tree does not show explicit signs of overfitting (despite relatively low accuracy due to the small training set).