Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2017-18-a

Assignment Number: 2
Student Name: Siddharth Agrawal
Roll Number: 150716
Date: October 10, 2017

QUESTION 1

Part 1

No, the first attribute (*name*) is **not** useful in learning a binary classifier from this data.
This is because it is impossible to find any correlation between the name of a professor and his/her ability to advise.
Mathematically, *name* attribute can have infinitely many possible values and hence is not suitable to train the classifier.

Part 2

No, it is **not** possible to perfectly classify this data without using name attribute. Otherwise it is possible.
There are two data points which have all the attributes (except *name*) as similar but their labels are different (Example: Prof. S. Snape and Prof. H. Slughorn have the same attribute values, viz., medum-no-heavy-(0-1) but have different label - yes and no respectively). In this case our classifier will not be able to properly distinguish between the two.

Part 3

As per the `ID3 Decision Algorithm`-

- Entropy of a set S over c outcomes is defined as -

$$\text{Entropy}(S) = -\sum_c P(I)\log_2 P(I)$$

  where $P(I)$ is the proportion of S belonging to class I.

- Gain(S, A) is the information gain of example S on attribute A, defined as -

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_v \left(\frac{|S_v|}{|S|} \times \text{Entropy}(S_v)\right)$$

  where:

  - S is each value v of all possible values of attribute A.
  - S_v = subset of S for which attribute A has value v
  - $|S_v|$ = number of elements in S_v
  - $|S|$ = number of elements in S

For the entire sample set S
$P(\text{Yes}) = \frac{5}{15}, P(\text{No}) = \frac{10}{15} \implies \text{Entropy}(S) = 0.9183$

The value of gain for the following attributes are -

- Attribute: # of meetings per week
  $\text{Entropy}(0-1) = 1.0$
  $\text{Entropy}(2-3) = 0.0$
  $\text{Entropy}(>3) = 0.0$
  $\text{Gain} = 0.2516$
- Attribute: Average workload?
  $\text{Entropy}(\text{average}) = 1.0$
  $\text{Entropy}(\text{heavy}) = 0.7219$
  $\text{Entropy}(\text{light}) = 0.8113$
  $\text{Gain} = 0.0613$
- Attribute: Like the research area?
  $\text{Entropy}(\text{Yes}) = 1.0$
  $\text{Entropy}(\text{No}) = 0.8454$
  $\text{Gain} = 0.0317$
- Attribute: Size of research group
  $\text{Entropy}(\text{small}) = 0.6500$
  $\text{Entropy}(\text{medium}) = 0.9710$
  $\text{Entropy}(\text{large}) = 1.0$
  $\text{Gain} = 0.0680$

Since the Gain is maximum for Attrbute: # of meetings per week, we choose this as the decision attribute in root node.
Also, since for this attribute - $\text{Entropy}(2-3) = \text{Entropy}(>3) = 0$, hence the set with these two values are *classified perfectly* and are assigned label 'no'.
For the remaining set with value $1-3$, we calculate gain as follows -

- Attribute: Average workload?
  $\text{Entropy}(\text{average}) = 0.0$
  $\text{Entropy}(\text{heavy}) = 0.7219$
  $\text{Entropy}(\text{light}) = 1.0$
  $\text{Gain} = 0.4391$
- Attribute: Like the research area?
  $\text{Entropy}(\text{Yes}) = 0.9183$
  $\text{Entropy}(\text{No}) = 0.9852$
  $\text{Gain} = 0.0349$
- Attribute: Size of research group
  $\text{Entropy}(\text{small}) = 0.0$

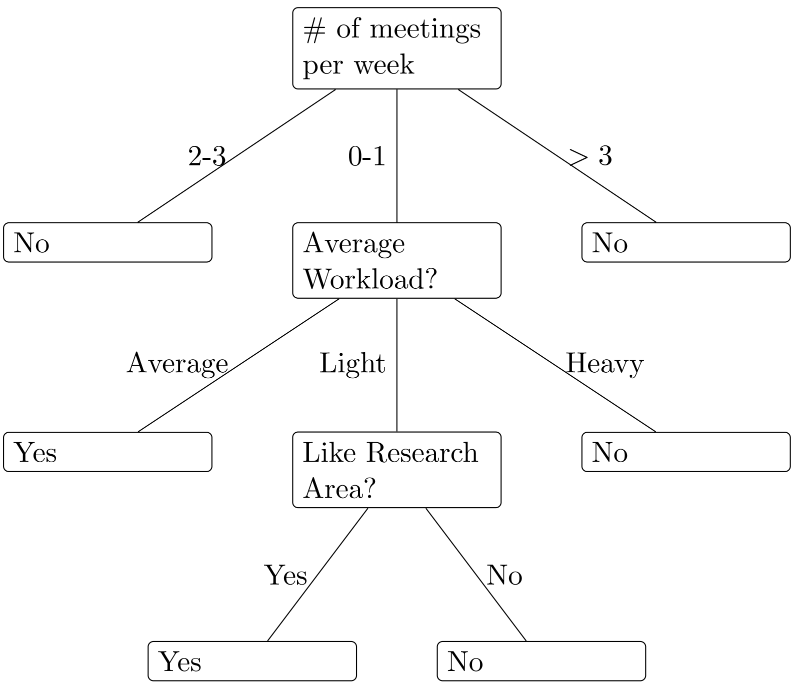
Entropy(medium) = 0.9710
Entropy(large) = 1.0
Gain = 0.1145

Since the Gain is maximum for Attrbute: Average workload?, we choose this as the decision attribute in second level node.
Also, since for this attribute - Entropy(average) = 0, hence the set with this value is *classified perfectly* and is assigned the label 'yes'.

For the set with value heavy, the other two attributes fail to classify the data ($\because$ Prof. S. Snape and Prof. H. Slughorn have the same attribute values, viz., medum-no-heavy-(0-1) but have different label - yes and no respectively).
Hence we assign the majorly dominant label 'no' to this set.

For the set with value light, 'Size of research group' attribute- large and small valued sets have indeterminate entropies, while 'Like the research area' attribute has gain = 1.
Hence we choose the Attribute: Like the research area? as the decision attribute in third level node. For the set with value 'no' we assign the label 'no' and for the set with value 'yes' we assign the label 'yes'.

The learnt decision tree is:-



Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2017-18-a

Assignment Number: 2
Student Name: Siddharth Agrawal
Roll Number: 150716
Date: October 10, 2017

QUESTION 2

Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2017-18-a

Assignment Number: 2
Student Name: Siddharth Agrawal
Roll Number: 150716
Date: October 10, 2017

QUESTION 3

Part 1

To prove that the given constraints $\xi_i \geq 0$ are vacuous, we just need to show that for any optimal solution $(\mathbf{w}, \{\xi_i\})$, $\nexists i \in [n]$ such that $\xi_i < 0$.

Claim: $\nexists k \in [n]$ *such that* $\xi_k < 0$ *where* $(\mathbf{w}, \{\xi_i\})$ *is the optimal solution.*
Proof by Contradiction:
Let the optimized solution be $(\mathbf{w}, \{\xi_i\})$ such that for some $k \in [n]$, $\xi_k < 0$ and $\xi_i \geq 0 \ \forall i \in [n]\backslash\{k\}$.
Let $\xi_k = -\epsilon$ where $\epsilon > 0$.
Hence, $y^k \left\langle \mathbf{w}, \mathbf{x}^k \right\rangle \geq 1 - \xi_k = 1 + \epsilon$.
$\because \ \epsilon > 0 \implies y^k \left\langle \mathbf{w}, \mathbf{x}^k \right\rangle > 1$
Now, consider the pair $(\mathbf{w}, \{\xi'_i\})$ such that $\xi'_k = 0$ and $\xi'_i \geq 0 \ \forall i \in [n]\backslash\{k\}$.
Clearly $\forall i \in [n]\backslash\{k\}$, $y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle \geq 1 - \xi'_i$. Also, $y^k \left\langle \mathbf{w}, \mathbf{x}^k \right\rangle > 1 \implies 1 - \xi'_k$. Thus $\{\xi'_i\}$ satisfies the given constraint.

$$
\begin{aligned}
f(\mathbf{w}, \{\xi'_i\}) &= \|\mathbf{w}\|_2^2 + \sum_{i=1}^{n} \xi_i'^2 \\
&= \|\mathbf{w}\|_2^2 + \sum_{\substack{i=0 \\ i \neq k}}^{n} \xi_i'^2 + \xi_k'^2 \\
&= \|\mathbf{w}\|_2^2 + \sum_{\substack{i=0 \\ i \neq k}}^{n} \xi_i'^2 && \because \ \xi'_k = 0 \\
&= \|\mathbf{w}\|_2^2 + \sum_{\substack{i=0 \\ i \neq k}}^{n} \xi_i^2 \\
&< \|\mathbf{w}\|_2^2 + \sum_{\substack{i=0 \\ i \neq k}}^{n} \xi_i^2 + \xi_k^2 && \because \ \xi_k < 0 \implies \xi_k^2 > 0 \\
&= \|\mathbf{w}\|_2^2 + \sum_{i=1}^{n} \xi_i^2 \\
&= f(\mathbf{w}, \{\xi_i\})
\end{aligned}
$$

Thus $f(\mathbf{w}, \{\xi'_i\}) < f(\mathbf{w}, \{\xi_i\})$.
But we assumed $(\mathbf{w}, \{\xi'_i\})$ to be the optimal solution – which is a clear Contradiction.
Hence our original assumption was wrong. Thus our claim is proved by contradiction.

Part 2

The Lagrangian for $(P1)$ is:

$$\mathcal{L}(\mathbf{w}, \{\xi_i\}, \boldsymbol{\alpha}) = \|\mathbf{w}\|_2^2 + \sum_{i=1}^{n} \xi_i^2 + \sum_{i=1}^{n} \alpha_i \left[1 - \xi_i - y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle\right] \tag{1}$$

Part 3

Primal Problem:

$$(\widehat{\mathbf{w}}_P, \{\widehat{\xi_i}\}_P) = \underset{\mathbf{w},\{\xi_i\}}{\arg\min} \underset{\substack{\alpha_i \geq 0 \\ i \in [n]}}{\arg\max} \ \mathcal{L}(\mathbf{w}, \{\xi_i\}, \boldsymbol{\alpha})$$

Dual Problem:

$$(\widehat{\mathbf{w}}_D, \{\widehat{\xi_i}\}_D) = \underset{\substack{\alpha_i \geq 0 \\ i \in [n]}}{\arg\max} \underset{\mathbf{w},\{\xi_i\}}{\arg\min} \ \mathcal{L}(\mathbf{w}, \{\xi_i\}, \boldsymbol{\alpha})$$

Derivation:
Differentiating (1) w.r.t. $\mathbf{w}$:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{w}} &= 2\mathbf{w} - \sum_{i=1}^{n} \alpha_i y^i \mathbf{x}^i \\
&= 0 \\
\implies \mathbf{w} &= \frac{1}{2} \sum_{i=1}^{n} \alpha_i y^i \mathbf{x}^i \qquad (2)
\end{aligned}$$

Substituting (2) in (1):

$$\begin{aligned}
\mathcal{L}(\{\xi_i\}, \boldsymbol{\alpha}) &= \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} y^i y^j \alpha_i \alpha_j \langle \mathbf{x}^i, \mathbf{x}^j \rangle + \sum_{i=1}^{n} \big(\xi_i^2 + \alpha_i(1 - \xi_i)\big) - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y^i y^j \alpha_i \alpha_j \langle \mathbf{x}^i, \mathbf{x}^j \rangle \\
&= -\frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} y^i y^j \alpha_i \alpha_j \langle \mathbf{x}^i, \mathbf{x}^j \rangle + \sum_{i=1}^{n} \big(\xi_i^2 + \alpha_i(1 - \xi_i)\big) \qquad (3)
\end{aligned}$$

$\forall i \in [n]$, Differentiating (1) w.r.t ξ_i:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \xi_i} &= 2\xi_i - \alpha_i \\
&= 0 \\
\implies \xi_i &= \frac{\alpha_i}{2} \qquad (4)
\end{aligned}$$

Substituting (4) in (3):

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}) &= -\frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} y^i y^j \alpha_i \alpha_j \langle \mathbf{x}^i, \mathbf{x}^j \rangle + \sum_{i=1}^{n} \left(\left(\frac{\alpha_i}{2}\right)^2 + \alpha_i \left(1 - \frac{\alpha_i}{2}\right) \right) \\
&= -\frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} y^i y^j \alpha_i \alpha_j \langle \mathbf{x}^i, \mathbf{x}^j \rangle + \sum_{i=1}^{n} \left(\alpha_i - \frac{\alpha_i^2}{4} \right) \qquad (5)
\end{aligned}$$

Thus the Lagrangian Dual Problem for $(P1)$ is:

$$\underset{\substack{\alpha_i \geq 0 \\ i \in [n]}}{\arg\min}\ \mathcal{L}(\boldsymbol{\alpha}) = \underset{\substack{\alpha_i \geq 0 \\ i \in [n]}}{\arg\min} \left(-\frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} y^i y^j \alpha_i \alpha_j \langle \mathbf{x}^i, \mathbf{x}^j \rangle + \sum_{i=1}^{n} \left(\alpha_i - \frac{\alpha_i^2}{4} \right) \right)$$

Part 4

Part 5

No, the positivity constraints $\xi_i \geq 0$ are **not** vacuous for the original SVM problem.
This is because in the original SVM problem, the expression to be minimized has the term $\sum \xi_i$, thus negative and positive value combination for ξ_i can help minimize the expression. Thus the positivity constraint needs to be *explicitly* mentioned.
For the given SVM problem, since the expression to be minimized has the term $\sum \xi_i^2$, thus positive and negative value combination of xi_i will give the same expression (since sign gets lost while squaring), hence the positivity constraint need *not* be *explicitly* mentioned.

Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2017-18-a

QUESTION 4

Assignment Number: 2
Student Name: Siddharth Agrawal
Roll Number: 150716
Date: October 10, 2017

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis feugiat vehicula dolor, sed ultricies leo. Phasellus euismod dictum felis in euismod. Proin pretium vel neque in placerat. Proin imperdiet egestas vulputate. Etiam faucibus accumsan ante non viverra. Duis ultrices ac odio vel sodales. In maximus gravida dolor, ut commodo lacus. Pellentesque ante massa, venenatis id aliquam et, posuere sed dui. Duis dignissim justo sit amet augue posuere fringilla. Suspendisse at nisi gravida, mattis justo sit amet, elementum elit. Praesent et massa ornare, consequat dui eget, ornare risus. Duis est nibh, sollicitudin nec mattis non, mattis in leo. Donec finibus justo sed massa sagittis, non fermentum nibh dictum. Pellentesque et congue purus. Donec porta pretium porttitor.

Morbi euismod risus eu tortor ornare malesuada. Nunc sed sollicitudin neque, efficitur rhoncus tellus. Cras malesuada augue arcu. Sed sem odio, tincidunt quis laoreet ac, facilisis ut nibh. Quisque gravida dolor at egestas aliquam. Aenean mollis massa sit amet enim mattis, vel fermentum tortor facilisis. Donec pellentesque est velit, vitae posuere lorem tristique ut.