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

QUESTION 1

Assignment Number: 2
Student Name: Deepanshu Bansal
Roll Number: 150219
Date: October 5, 2017

---

1. No. It is not useful in learning a binary classifier from this data because it just defines different persons same as different S.No. and has nothing to do with good or bad advisor. Moreover each name is different and those cannot be classified into different classes. If we try to do so that would mean as many classes as many names. So not at all useful.

2. No. Consider S.No. entries "6" and "14" they have same values for all decision making attributes but still different final result. One is good advisor while other is not. So no matter how complicated the classification algorithm is we cannot perfectly classify this data.

3. The decision tree is as follows:

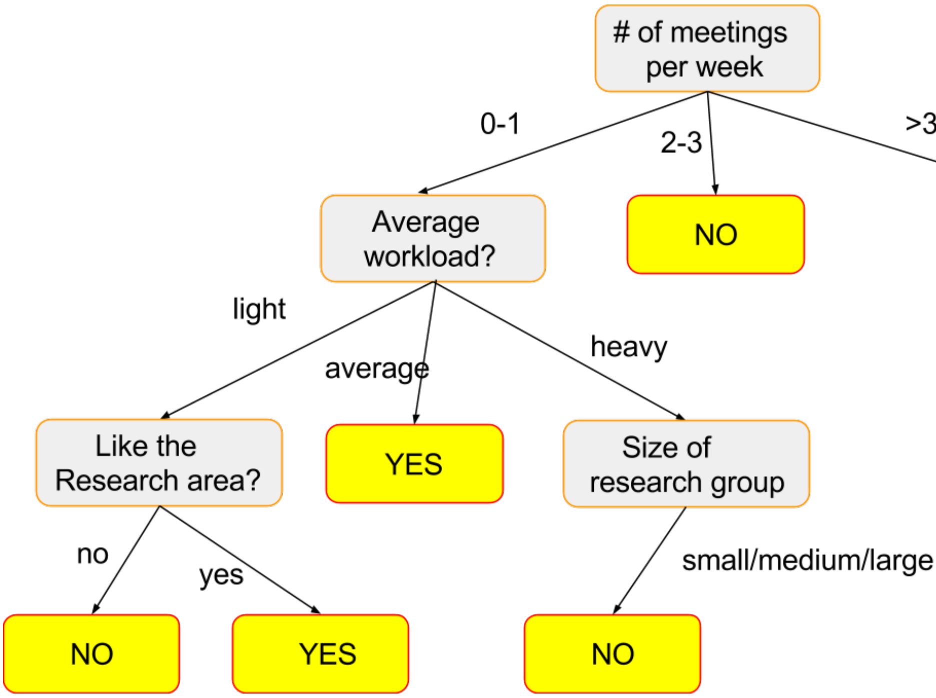

Figure 1: Decision Tree for given data.

Justification

1. I chose root node to be # of meetings per week because it has the highest information gain of 0.251.

   - $Gain(S,\text{Size of Research Group}) = 0.068$
   - $Gain(S,\text{Like Research Area}) = 0.032$
   - $Gain(S,\text{Workload}) = 0.061$

- $Gain(S$,no. of meetings) = 0.251

Since it has three possible values, the root node has three branches (0-1, 2-3, >3). Now the node corresponding to (0-1) branch is for Average Workload? because it has highest information gain of 0.439

- $Gain(S_{no.of meetings=0-1}$,Size of Research Group) = 0.115
- $Gain(S_{no.of meetings=0-1}$,Like Research Area) = 0.035
- $Gain(S_{no.of meetings=0-1}$,Workload) = 0.439

Now the node corresponding to (2-3) and (>3) branches are leaves beacause gain for them is zero.

- $Gain(S_{no.of meetings=2-3}$,Size of Research Group) = 0.000
- $Gain(S_{no.of meetings=2-3}$,Like Research Area) = 0.000
- $Gain(S_{no.of meetings=2-3}$,Workload) = 0.000
- $Gain(S_{no.of meetings>3}$,Size of Research Group) = 0.000
- $Gain(S_{no.of meetings>3}$,Like Research Area) = 0.000
- $Gain(S_{no.of meetings>3}$,Workload) = 0.000

For the splitting of Average Workload? node according to "light" information gain is highest for Like the research area? as 1.000 hence splitted to it.

- $Gain(S_{LightWorkload\_no.of meetings=0-1}$,Size of Research Group) = 0.000
- $Gain(S_{LightWorkload\_no.of meetings=0-1}$,Like Research Area) = 1.000

For "average" answer is sure to be "yes" as all three yes and for "heavy" it splitted to Size of research group as it has more information gain of 0.170

- $Gain(S_{HeavyWorkload\_no.of meetings=0-1}$,Size of Research Group) = 0.170
- $Gain(S_{HeavyWorkload\_no.of meetings=0-1}$,Like Research Area) = 0.072

2. I chose not to split nodes emerging from root with branch (2-3) and (>3) as leaves because **information gain** from them is **zero** and decision is sure to be NO. For the leaf emerging from "Average Workload?" it is so because at this point it have only 3 Yes hence making sure as YES leaf. Now for "Like research area" attribute only two points are given and only for "medium" value for "Size of research group" attribute we cant really decide but now it is decided on the basis of whether now the value of "Like research area?" is yes or no, as given in training if it's yes then result is yes else no hence they are two leaves. Similarly for "Size of research group" attribute answer is always no as it have 4 No's and 1 Yes hence resulting output as No and into a leaf.

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

Assignment Number: 2
Student Name: Deepanshu Bansal
Roll Number: 150219
Date: October 5, 2017

QUESTION 2

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.

Fusce pulvinar convallis lobortis. Mauris iaculis lacus vitae dui suscipit, ut ornare neque placerat. In mattis malesuada rutrum. Vivamus consectetur tempus ex sit amet aliquam. In blandit libero at mi rutrum, nec iaculis orci cursus. Maecenas a dolor lorem. Donec pretium turpis sapien, dapibus sollicitudin odio scelerisque eget. Maecenas egestas tellus a quam scelerisque, et pretium magna condimentum. In dapibus feugiat ornare. Nunc eget nulla convallis, laoreet tortor nec, convallis dui. Etiam in leo vitae nulla facilisis congue. Curabitur blandit sodales augue. Vivamus et aliquam orci, non suscipit elit. Quisque vestibulum lacus at velit congue semper.

Nulla efficitur risus nunc, in posuere turpis tempor eget. Sed efficitur id tellus non vestibulum. Praesent elementum condimentum sollicitudin. Integer eget quam dictum, varius est sit amet, aliquam mauris. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Morbi in pretium dui. Sed luctus magna rutrum ex mollis, ut blandit lacus tincidunt. In tincidunt urna neque, placerat consequat velit sagittis id. Morbi pretium maximus fermentum. Interdum et malesuada fames ac ante ipsum primis in faucibus. Duis vehicula efficitur rhoncus. Nullam lacinia semper scelerisque. Maecenas eleifend nisi et ante auctor, a tincidunt arcu molestie. Aenean faucibus feugiat arcu ac mattis. Morbi quis suscipit sapien. Pellentesque pulvinar fermentum tellus at malesuada. Nam id metus vitae risus dignissim laoreet. Pellentesque massa velit, vehicula in convallis et, vestibulum sed turpis. In venenatis massa vel mattis tincidunt. Donec varius faucibus elit, in blandit metus interdum sit amet. Nullam vel nibh non nisl mollis volutpat. Donec cursus iaculis lorem, id elementum metus iaculis nec. Cras a diam porttitor, suscipit dolor id, vestibulum nisi. Morbi maximus mauris a iaculis hendrerit. Duis rutrum quam in ex lobortis gravida. Aenean iaculis lacinia metus. Fusce sit amet dignissim elit, sed sodales lacus. Mauris ac neque finibus, bibendum tortor id, scelerisque neque. In nec quam ullamcorper, egestas sem ut, iaculis ante. Nam porta diam ut lacus sagittis euismod.

proto_blank.png

proto_euclid_sample.png

Figure 2: Learning with Prototypes: the figure on the left shows the two prototypes. The figure on the right shows what the decision boundary if the distance measure used is $d(\mathbf{z}^1, \mathbf{z}^2) = \left\|\mathbf{z}^1 - \mathbf{z}^2\right\|_2$, for any two points $\mathbf{z}^1, \mathbf{z}^2 \in \mathbb{R}^2$. The decision boundary in this case is the line $y = x$.

QUESTION 3

Assignment Number: 2
Student Name: Deepanshu Bansal
Roll Number: 150219
Date: October 5, 2017

We have the following problem:

$$\begin{aligned}\underset{\mathbf{w},\{\xi_i\}}{\arg\min}\ & \|\mathbf{w}\|_2^2 + \sum_{i=1}^{n} \xi_i^2 \\ \text{s.t. } & y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle \geq 1 - \xi_i, \text{ for all } i \in [n] \qquad (P1) \\ & \xi_i \geq 0, \text{ for all } i \in [n]\end{aligned}$$

1. We know that $\xi^2 \geq 0$ for all real ξ. Here we are trying to minimize the optimization problem hence we want all ξ_i to be equal to zero if constraints weren't there. Now we have constraint :

$$y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle \geq 1 - \xi_i, \text{ for all } i \in [n]$$
$$\xi_i \geq 1 - y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle, \text{ for all } i \in [n]$$

Case-I: $y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle \leq 1$, we get

$$1 - y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle \geq 0, \text{ for all } i \in [n]$$
$$\xi_i \geq 0, \text{ for all } i \in [n]$$

Here we get constraints $\xi_i \geq 0$ for all $i \in [n]$ from the first constraint itself hence specifying constraints $\xi_i \geq 0$ for all $i \in [n]$ is vacuous in this case.

Case-II: $y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle > 1$, we get

$$1 - y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle < 0, \text{ for all } i \in [n]$$
$$\text{Let } 1 - y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle = \alpha_i, \text{ for all } i \in [n]$$
$$\text{Now } \xi_i \geq \alpha_i, \text{ for all } i \in [n]$$

Since $\alpha_i < 0$ hence this means that ξ_i is greater than a negative value for all $i \in [n]$. Since we are minimizing the optimization problem we will always choose $\xi_i = 0$ for all $i \in [n]$ thus minimizing target as square of negative number is also greater then zero and we don't want that and thus the constraints $\xi_i \geq 0$ for all $i \in [n]$ are vacuous in this case also.

So the constraints $\xi_i \geq 0$ are vacuous i.e. the optimization problem does not change even if we remove the all constraints $\xi_i \geq 0$ for all $i \in [n]$.

2. Without vacuous constraints optimization problem becomes

$$\begin{aligned} \arg\min_{\mathbf{w},\{\xi_i\}} \; & \|\mathbf{w}\|_2^2 + \sum_{i=1}^{n} \xi_i^2 \\ \text{s.t. } & 1 - \xi_i - y^i \left\langle \mathbf{w}, \mathbf{x}^i \right\rangle \leq 0, \text{ for all } i \in [n] \end{aligned}$$

Lagrangian using multipliers α_i's s.t $\alpha_i \geq 0$ for all $i \in [n]$ is

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

Now the optimization problem becomes

$$\arg\min_{\mathbf{w},\{\xi_i\}} \left\{ \arg\max_{\boldsymbol{\alpha} \geq 0} \{L(\mathbf{w}, \boldsymbol{\xi}, \boldsymbol{\alpha})\} \right\}$$

3. Derivation for the Lagrangian dual problem for $(P1)$ is as follows

$$\begin{aligned} \nabla_{\mathbf{w}} L &= 2\mathbf{w} - \sum_{i=1}^{n} \alpha_i y^i \mathbf{x}^i \\ \nabla_{\mathbf{w}} L &= 0 \\ \mathbf{w} &= \frac{1}{2} \sum_{i=1}^{n} \alpha_i y^i \mathbf{x}^i \end{aligned}$$

$$\begin{aligned} L(\boldsymbol{\xi}, \boldsymbol{\alpha}) &= \left\| \frac{1}{2} \sum_{i=1}^{n} \alpha_i y^i \mathbf{x}^i \right\|_2^2 + \sum_{i=1}^{n} \xi_i^2 + \sum_{i=1}^{n} \alpha_i (1 - \xi_i - y^i \left\langle \frac{1}{2} \sum_{i=1}^{n} \alpha_i y^i \mathbf{x}^i, \mathbf{x}^i \right\rangle) \\ L(\boldsymbol{\xi}, \boldsymbol{\alpha}) &= \frac{1}{4} \sum_{n} \sum_{m} \alpha_n \alpha_m y^n y^m \left\langle \mathbf{x}^n, \mathbf{x}^m \right\rangle + \sum_{i=1}^{n} \xi_i^2 + \sum_{i=1}^{n} \alpha_i (1 - \xi_i) - \frac{1}{2} \sum_{n} \sum_{m} \alpha_n \alpha_m y^n y^m \left\langle \mathbf{x}^n, \mathbf{x}^m \right\rangle \\ L(\boldsymbol{\xi}, \boldsymbol{\alpha}) &= -\frac{1}{4} \sum_{n} \sum_{m} \alpha_n \alpha_m y^n y^m \left\langle \mathbf{x}^n, \mathbf{x}^m \right\rangle + \sum_{i=1}^{n} (\xi_i^2 + \alpha_i - \alpha_i \xi_i) \end{aligned}$$

$$\begin{aligned} \nabla_{\xi_i} L &= 2\xi_i - \alpha_i, \text{ for all } i \in [n] \\ \nabla_{\xi_i} L &= 0, \text{ for all } i \in [n] \\ \xi_i &= \frac{\alpha_i}{2}, \text{ for all } i \in [n] \end{aligned}$$

$$L(\boldsymbol{\alpha}) = -\frac{1}{4} \sum_{n} \sum_{m} \alpha_n \alpha_m y^n y^m \left\langle \mathbf{x}^n, \mathbf{x}^m \right\rangle + \sum_{i=1}^{n} (\alpha_i - \frac{\alpha_i^2}{4})$$

Thus Lagrangian dual problem for $(P1)$ is $\arg\max_{\boldsymbol{\alpha} \geq 0} \{L(\boldsymbol{\alpha})\}$ or $\arg\min_{\boldsymbol{\alpha} \geq 0} \{-L(\boldsymbol{\alpha})\}$

$$\arg\min_{\boldsymbol{\alpha} \geq 0} \left\{ \frac{1}{4} \sum_{n} \sum_{m} \alpha_n \alpha_m y^n y^m \left\langle \mathbf{x}^n, \mathbf{x}^m \right\rangle - \sum_{i=1}^{n} (\alpha_i - \frac{\alpha_i^2}{4}) \right\}$$

4. The original SVM problem is

$$\underset{0 \le \boldsymbol{\alpha} \le C}{\arg\min} \left\{ \frac{1}{2} \sum_n \sum_m \alpha_n \alpha_m y^n y^m \langle \mathbf{x}^n, \mathbf{x}^m \rangle - \sum_{i=1}^{n} \alpha_i \right\}$$

The differences are

(a) The constant in the term $\sum_n \sum_m \alpha_n \alpha_m y^n y^m \langle \mathbf{x}^n, \mathbf{x}^m \rangle$.
We have $\frac{1}{4}$ while the original SVM problem has $\frac{1}{2}$.

(b) There is extra $\sum_{i=1}^{n} \frac{\alpha_i^2}{4}$ term in dual problem for $(P1)$.

(c) We dont get any other constraint on α. We have only $\boldsymbol{\alpha} \ge 0$ while original have $0 \le \boldsymbol{\alpha} \le C$.

5. No, the positivity constraints $\xi_i \ge 0$ are not vacuous for the original SVM problem.

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

Assignment Number: 2
Student Name: Deepanshu Bansal
Roll Number: 150219
Date: October 5, 2017

QUESTION 4

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.