CS 344 (Spring 2017): Class Test 1

Instructor: Shivaram Kalyanakrishnan

8.30 a.m. – 9.30 a.m., January 24, 2017, 101/103/105 New CSE Building

Total marks: 15

Note. Provide brief justifications and/or calculations along with each answer to illustrate how you arrived at the answer.

Question 1. What is the "AI Effect", which is sometimes called the "AI Paradox" or the "Odd Paradox"? Explain using an example. [2 marks]

Question 2. What is a simple reflex agent? What is a goal-based agent? What advantage does the former have over the latter? What advantage does the latter have over the former? [3 marks]

Question 3. A data set D contains 4 classes: l_1, l_2, l_3, and l_4. The fraction of data points with label l_1 is $f_1 = 20\%$, with label l_2 is $f_2 = 40\%$, with label l_3 is $f_3 = 15\%$, and with label l_4 is $f_4 = 25\%$.

3a. Consider a predictor P_1 that predicts the label of a given input point by picking **uniformly** at random from $\{l_1, l_2, l_3, l_4\}$. In other words, P_1 makes its prediction by picking one of the four labels, each with a probability of 1/4. What is the expected accuracy of P_1 on D?; equivalently, what is the expected fraction of points in D classified correctly by P_1? [1 mark]

3b. Consider another predictor P_2 that predicts the label of a given input point by picking label l_1 with probability p_1, label l_2 with probability p_2, label l_3 with probability p_3, and label l_4 with probability p_4, where $p_1 + p_2 + p_3 + p_4 = 1$. For what value(s) of (p_1, p_2, p_3, p_4), is the expected accuracy of P_2 on D maximised? [2 marks]

Question 4. This question corresponds to an exercise given in Class Note 1. Let

$$D = ((\mathbf{x}^1, y^1), (\mathbf{x}^2, y^2), \ldots, (\mathbf{x}^n, y^n))$$

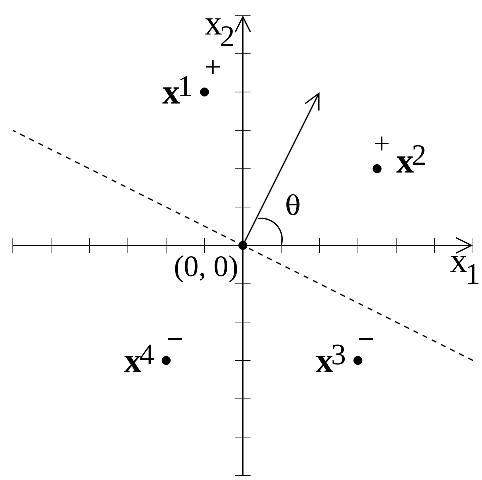
be a set of labeled data points such that for $i \in \{1, 2, \ldots, n\}$: $\mathbf{x}^i \in \mathbb{R}^d$ and $y^i \in \{-1, 1\}$. This data set is linearly separable by a non-origin-centred hyperplane (there is no guarantee that it is also separable by an origin-centred hyperplane). In other words, all we can assume is that there exist $\mathbf{w}^\star \in \mathbb{R}^d$ and $b \in \mathbb{R} \setminus \{0\}$ such that for all $i \in \{1, 2, \ldots, n\}$,

$$y^i(\mathbf{w}^\star \cdot \mathbf{x}^i + b) > 0.$$

Describe a procedure that uses the Perceptron Learning Algorithm discussed in class (which can only learn origin-centred hyperplanes) as a blackbox in order to find a non-origin-centred hyperplane in $\mathbb{R}^d$ that classifies all the points in D correctly. Provide a proof that your procedure is correct. [3 marks]

Question 5. Consider the data set shown in the table and plot below. Clearly, the + and the − points in this data set can be separated by an origin-centred hyperplane. Since our space is 2-dimensional, we may represent a hyperplane by the angle $\theta \in (-180°, 180°]$ that its **normal** makes with the x_1 axis. In addition to showing the four data points, the plot below also shows an example of a separating hyperplane (dotted line, with perpendicular arrow showing the normal direction).

i	x_1^i	x_2^i	y^i
1	-1	4	$+1$
2	$2\sqrt{3}$	2	$+1$
3	3	-3	-1
4	-2	-3	-1



5a. Give the set of all θ corresponding to hyperplanes that correctly classify all four input points. [1 mark]

5b. Assume, as we did in class, that we initialise the Perceptron Learning Algorithm with the zero vector, and iteratively update the weight vector based on misclassified points. Recall that the algorithm may pick an *arbitrary* misclassified point at each step to update its weight vector. For the data set shown above, what is the set of all hyperplanes that the algorithm can possibly return upon termination? [3 marks]

Solutions

1. The "AI Effect" refers to the phenomenon that every time AI reaches a goal, society ceases to view that goal as a part (and a success!) of AI. For example, the game of Chess was pursued for several decades (even as far back as Turing) as a holy grail for AI. In the 1990's, Chess-playing programs could finally outplay humans. Since we now know how these programs are engineered, they are no longer considered exemplars of "intelligence", nor in the domain of AI. The goals of AI have moved forward.

2. A simple reflex agent implements a mapping from sensation to action, whereas a goal-directed agent, armed with some knowledge about how the world, takes actions by explicitly reasoning about their consequences, such that a goal can be reached. As the name suggests, simple reflex agents have very little "thinking" to do—they are easy to implement and computationally lightweight. However, a goal-directed agent is a lot more robust to changes that could occur in the world. For example, if there is a road-closure on your route home from work, you could replan if you knew the map. A simple reflex agent that follows the rule "if at Saki Naka signal, turn right" might never reach home in such an eventuality.

3a. Assume there are n points. The expected accuracy is

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}\{\text{Point } i \text{ is classified correctly by } P_1\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{4}\mathbb{P}\{\text{Point } i \text{ has label } l \text{ and is classified as } l \text{ by } P_1\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{4}\mathbb{P}\{\text{Point } i \text{ has label } l\}\cdot\mathbb{P}\{\text{Point } i \text{ is classified as } l \text{ by } P_1\}\\
=&\sum_{l=1}^{4}\left(\sum_{i=1}^{n}\frac{\mathbb{P}\{\text{Point } i \text{ has label } l\}}{n}\right)\cdot\mathbb{P}\{\text{Point } i \text{ is classified as } l \text{ by } P_1\}\\
=&\sum_{l=1}^{4} f_l\cdot\frac{1}{4}=\frac{1}{4}.
\end{aligned}
$$

3b. By a working similar to that in 3a, we get that the expected accuracy of P_2 is

$$\sum_{l=1}^{4} f_l \cdot p_l.$$

Observe that regardless of the actual values of p_1, p_2, p_3, and p_4,

$$f_1p_1 + f_2p_2 + f_3p_3 + f_4p_4 \leq \max(f_1, f_2, f_3, f_4) = f_2.$$

Indeed an expected accuracy of f_2 can (only) be achieved by setting $(p_1, p_2, p_3, p_4) = (0, 1, 0, 0)$. This strategy amounts to predicting the "most frequent class".

4. We transform our original data set D into a data set D' in $d+1$ dimensions, where

$$D' = ((\mathbf{x}'^1, y^1), (\mathbf{x}'^2, y^2), \ldots, (\mathbf{x}'^n, y^n)),$$

where for $i \in \{1, 2, \ldots, n\}$: $\mathbf{x}'^i \in \mathbb{R}^{d+1}$ is identical to $\mathbf{x}^i$ in the first d dimensions, and has its $(d+1)^{\text{st}}$ component as 1. We also define a vector $\mathbf{w}'^\star \in \mathbb{R}^{d+1}$ which is identical to $\mathbf{w}^\star$ in the first d dimensions, and has its $(d+1)^{\text{st}}$ component as b. Observe that

$$\forall i \in \{1, 2, \ldots, n\} : y^i(\mathbf{w}^\star \cdot \mathbf{x}^i + b) > 0 \iff \forall i \in \{1, 2, \ldots, n\} : y^i(\mathbf{w}'^\star \cdot \mathbf{x}'^i) > 0.$$

Thus, D' is linearly separable using an origin-centred hyperplane $\mathbf{w}^\star$. Therefore, we can find an an origin-centred separating hyperplane in $d+1$ dimensions by running the Perceptron Learning Algorithm on D'. The answer returned by the algorithm can be interpreted as a non-origin-centred hyperplane in d dimensions (with the first d dimensions giving the normal, and the $(d+1)^{\text{st}}$ dimension giving the intercept).

5a. From the figure, it is clear that every separating hyperplane must lie "between" $\mathbf{x}^3$ and $\mathbf{x}^2$ (and pointing towards the + points), and in fact that every hyperplane in this region is a legal separating hyperplane. Thus, θ must be in the set $(45°, 120°)$.

5b. Initially, when $\mathbf{w}^1 = \mathbf{0}$, any of $\mathbf{x}^1$, $\mathbf{x}^2$, $\mathbf{x}^3$, and $\mathbf{x}^4$ can be picked (since they all have a zero dot product with $\mathbf{w}^1$). Thus, $\mathbf{w}^2$ must be in the set $\{\mathbf{x}^1, \mathbf{x}^2, -\mathbf{x}^3, -\mathbf{x}^4\}$. It is easy to verify that $\mathbf{x}^1$ and $-\mathbf{x}^4$ are already valid separating hyperplanes, and will be returned if they are found at the end of this update. On the other hand, (1) if point $\mathbf{x}^2$ was picked, then point $\mathbf{x}^3$ is misclassified (since $\mathbf{x}^2 \cdot \mathbf{x}^3 > 0$), and (2) if point $\mathbf{x}^3$ was picked, then point $\mathbf{x}^2$ is misclassified (since $-\mathbf{x}^3 \cdot \mathbf{x}^2 < 0$). Consequently, the only possible value $\mathbf{w}^3$ can take is $\mathbf{x}^2 - \mathbf{x}^3$, which can be seen to be a separating hyperplane. The required answer is

$$\{\mathbf{x}^1, -\mathbf{x}^4, \mathbf{x}^2 - \mathbf{x}^3\} = \{(-1, 4), (2, 3), (2\sqrt{3} - 3, 5)\}.$$