CSE 251A Winter 2021 - Take Home Exam 1 Solution

1 hour **Name:** **PID:**

(1) (10 points) Draw the decision boundary for the 1-nearest neighbor classifier on the following labeled data points in 2 dimensions.

$$((1, 2), 1), ((-3, 4), 2), ((-2, -1), 3)$$

For full credit, write down the equations for each segment of the decision boundary. Clearly label all the decision regions with the label assigned by the 1-nearest neighbor classifier.

Solution: Let's consider all points whose nearest neighbor will be $(1, 2)$; these points are closer to $(1, 2)$ than $(-3, 4)$ as well as closer to $(1, 2)$ than $(-2, -1)$. If $\mathrm{x} = (x_1, x_2)$ is such a point, then we can write these down as the following equations:

$$(x_1 - 1)^2 + (x_2 - 2)^2 \leq (x_1 + 3)^2 + (x_2 - 4)^2$$
$$(x_1 - 1)^2 + (x_2 - 2)^2 \leq (x_1 + 2)^2 + (x_2 + 1)^2$$

If we expand and solve these, we get the decision boundary segments:

$$x_2 - 2x_1 - 5 \leq 0$$
$$x_2 + x_1 \geq 0$$

If we do the same and write the inequalities for the rest of the points as well and solve them we get the decision boundaries below for point $(-3, 4)$:

$$x_2 - 2x_1 - 5 \geq 0, \quad x_2 - \frac{1}{5}x_1 - 2 \geq 0$$

and for the last point we get:

$$x_2 + x_1 \leq 0, \quad x_2 - \frac{1}{5}x_1 - 2 \leq 0$$

Figure 1 shows these boundaries and the labels of the regions. If we put all of these together, we get the classifier's piece-wise labeling function:

$$y = \begin{cases} 1 & \text{if} \quad x_2 - 2x_1 - 5 < 0, \quad x_2 + x_1 \geq 0 \\ 2 & \text{if} \quad x_2 - 2x_1 - 5 \geq 0, \quad x_2 - \frac{1}{5}x_1 - 2 \geq 0 \\ 3 & \text{if} \quad x_2 - \frac{1}{5}x_1 - 2 < 0, \quad x_2 + x_1 < 0 \end{cases}$$

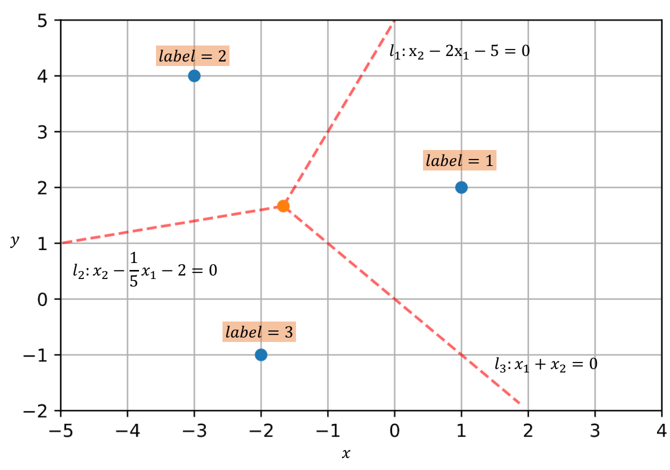

FIGURE 1. Decision boundaries of the classifier

(2) Suppose we have instances X drawn from an instance space $\mathcal{X}$ and labels Y drawn from $\{0, 1\}$. In class and the homework, we talked about how the data distribution D is a joint distribution over $X \times Y$ and can be factorized into the marginal over X times the conditional distribution of Y given X as: $D(X = x, Y = y) = \mu(X = x) \cdot \eta(Y = y|X = x)$. It turns out that D can also be factorized in a different manner:

$$D(X = x, Y = y) = \pi(Y = y) \cdot f(X = x|Y = y)$$

where $\pi(Y = y)$ is the marginal over Y and $f(X = x|Y = y)$ is the conditional likelihood of X given a specific value of Y.

(a) (5 points) Write down expressions for $\pi(Y = y)$ and $f(X = x|Y = y)$ as functions of $\mu(X = x)$ and $\eta(Y = y|X = x)$.

Solution:

$$\pi(Y = y) = \sum_x D(X = x, Y = y) = \sum_x \eta(Y = y|X = x).\mu(X = x)$$
$$f(X = x|Y = y) = \frac{\eta(Y = y|X = x).\mu(X = x)}{\pi(Y = y)} = \frac{\eta(Y = y|X = x).\mu(X = x)}{\sum_x \eta(Y = y|X = x).\mu(X = x)}$$

(b) (5 points) Suppose now that we change $f(X = x|Y = y)$ while keeping $\pi(Y = y)$ fixed. Does the Bayes Optimal classifier for D change? If yes, give an example where it happens. If no, give a brief justification.

Solution: Yes, the classifier would be different. Recall that the Bayes Optimal Classifier depends only on $\eta(Y|X)$. We can write this as:

$$\eta(Y|X) = \frac{f(X|Y).\pi(Y)}{\mu(X)}$$

When we change f while keeping π the same, we can end up changing η which in turn will lead to a change in the Bayes optimal. Let's look at a concrete example. Let's assume we have two cases, a and b, where $f_a(X|Y)$ is shown in likelihood Table 1a (where X is sunny or not and Y is playing golf or not) and $f_b(X|Y)$ is shown in likelihood Table 1b. For both cases we assume $\pi(Y = yes) = 0.9$ and $\pi(Y = no) = 0.1$, since π is supposed to remain unchanged. Now, given the input $X = sunny$, the Bayes optimal classifier would return a y that would maximize $\eta(Y = y|X = sunny)$. For case a, using Bayes rule, we write:

$$\eta_a(Y = y|X = sunny) = \frac{f_a(X = sunny|Y = y).\pi(Y = y)}{\mu_a(X = sunny)}$$

To find the $Y = y$ that maximizes this probability, we can only focus on maximizing the numerator.Based on the table:

$$\eta_a(Y = yes|X = sunny) = \frac{0.9 * 0.9}{\mu_a(X = sunny)} = \frac{0.81}{\mu_a(X = sunny)}$$
$$\eta_a(Y = no|X = sunny) = \frac{0.3 * 0.1}{\mu_a(X = sunny)} = \frac{0.03}{\mu_a(X = sunny)}$$

for case b, we write:

$$\eta_b(Y = yes|X = sunny) = \frac{0.05 * 0.9}{\mu_b(X = sunny)} = \frac{0.045}{\mu_b(X = sunny)}$$
$$\eta_b(Y = no|X = sunny) = \frac{0.95 * 0.1}{\mu_b(X = sunny)} = \frac{0.095}{\mu_b(X = sunny)}$$

TABLE 1. Likelihood tables for question 2.b.

(A) Before

	Play Golf	
	Yes	No
Sunny	0.90	0.30
Not sunny	0.10	0.70

(B) After

	Play Golf	
	Yes	No
Sunny	0.05	0.90
Not sunny	0.95	0.10

So, for case a, $0.81 > 0.03$ so $Y = yes$ yields the highest probability, and the classifier would return $Y = yes$. For case b, $0.095 > 0.045$ so the classifier returns ($Y = no$), which is different from the previous one.