CMPE 257 Machine Learning Spring 2019

HW#2 Due March 11th, 11:59 PM, on Canvas

1. (10 points) Problem 2.1 from the textbook.

Problem 2.1 In Equation (2.1), set $\delta = 0.03$ and let

$$\epsilon(M, N, \delta) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}.$$

(a) For $M = 1$, how many examples do we need to make $\epsilon \leq 0.05$?

(b) For $M = 100$, how many examples do we need to make $\epsilon \leq 0.05$?

(c) For $M = 10,000$, how many examples do we need to make $\epsilon \leq 0.05$?

2. (10 points) Problem 2.3 from the textbook.

Problem 2.3 Compute the maximum number of dichotomies, $m_{\mathcal{H}}(N)$, for these learning models, and consequently compute d_{VC}, the VC dimension.

(a) Positive or negative ray: $\mathcal{H}$ contains the functions which are $+1$ on $[a, \infty)$ (for some a) together with those that are $+1$ on $(-\infty, a]$ (for some a).

(b) Positive or negative interval: $\mathcal{H}$ contains the functions which are $+1$ on an interval $[a, b]$ and -1 elsewhere or -1 on an interval $[a, b]$ and $+1$ elsewhere.

(c) Two concentric spheres in $\mathbb{R}^d$: $\mathcal{H}$ contains the functions which are $+1$ for $a \leq \sqrt{x_1^2 + \ldots + x_d^2} \leq b$.

3. (10 points) Problem 2.12 from textbook.

Problem 2.12 For an $\mathcal{H}$ with $d_{\text{VC}} = 10$, what sample size do you need (as prescribed by the generalization bound) to have a 95% confidence that your generalization error is at most 0.05?

4. (10 points) Prove that selecting the hypothesis h that maximizes the likelihood $\prod_{n=1}^{N} P(y_n|x_n)$ is equivalent to minimizing the cross-entropy error

$$E_{in}(\boldsymbol{w}) = \frac{1}{N} \sum_{n=1}^{N} \ln(1 + e^{-y_n \boldsymbol{w}^T x_n})$$

5. (10 points) Derive the gradient of the in-sample error $\nabla E_{in}(\boldsymbol{w}(t))$ used in the gradient descent algorithm.

6. (10 points) Exercise 3.13 (a) (b) (c) from textbook

Exercise 3.13

Consider the feature transform $\mathbf{z} = \Phi_2(\mathbf{x})$ in (3.13). How can we use a hyperplane $\tilde{\mathbf{w}}$ in $\mathcal{Z}$ to represent the following boundaries in $\mathcal{X}$?

(a) The parabola $(x_1 - 3)^2 + x_2 = 1$.

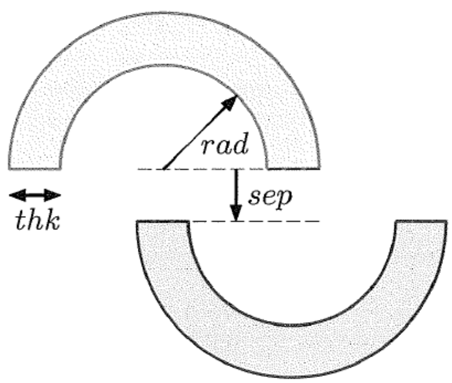
(b) The circle $(x_1 - 3)^2 + (x_2 - 4)^2 = 1$.

(c) The ellipse $2(x_1 - 3)^2 + (x_2 - 4)^2 = 1$.

$$\Phi_2(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_1 x_2, x_2^2), \qquad (3.13)$$

7. (20 points) Problem 3.1 from textbook. You can use the attached "HW2_3.1.ipynb" as a starting point to generate the data. Feel free to write your own code to generate the data.

Problem 3.1 Consider the double semi-circle "toy" learning task below.



There are two semi circles of width thk with inner radius rad, separated by sep as shown (red is -1 and blue is $+1$). The center of the top semi circle is aligned with the middle of the edge of the bottom semi circle. This task is linearly separable when $sep \geq 0$, and not so for $sep < 0$. Set $rad = 10$, $thk = 5$ and $sep = 5$. Then, generate $2{,}000$ examples uniformly, which means you will have approximately $1{,}000$ examples for each class.

(a) Run the PLA starting from $\mathbf{w} = \mathbf{0}$ until it converges. Plot the data and the final hypothesis.

(b) Repeat part (a) using the linear regression (for classification) to obtain $\mathbf{w}$. Explain your observations.

8. (10 points) Problem 3.2 from textbook

Problem 3.2 For the double semi circle task in Problem 3.1, vary sep in the range $\{0.2, 0.4, \ldots, 5\}$. Generate $2{,}000$ examples and run the PLA starting with $\mathbf{w} = \mathbf{0}$. Record the number of iterations PLA takes to converge.

Plot sep versus the number of iterations taken for PLA to converge. Explain your observations. *[Hint: Problem 1.3.]*

9. (20 points) Problem 3.3 from textbook

Problem 3.3 For the double semi circle task in Problem 3.1, set $sep = -5$ and generate $2,000$ examples.

(a) What will happen if you run PLA on those examples?

(b) Run the pocket algorithm for $100,000$ iterations and plot E_{in} versus the iteration number t.

(c) Plot the data and the final hypothesis in part (b).

(d) Use the linear regression algorithm to obtain the weights $\mathbf{w}$, and compare this result with the pocket algorithm in terms of computation time and quality of the solution.

(e) Repeat (b) – (d) with a 3rd order polynomial feature transform.

10. (10 points) Problem 3.16 from textbook

Problem 3.16 In Example 3.4, it is mentioned that the output of the final hypothesis $g(\mathbf{x})$ learned using logistic regression can be thresholded to get a 'hard' (± 1) classification. This problem shows how to use the risk matrix introduced in Example 1.1 to obtain such a threshold.

Consider fingerprint verification, as in Example 1.1. After learning from the data using logistic regression, you produce the final hypothesis

$$g(\mathbf{x}) = \mathbb{P}[y = +1 \mid \mathbf{x}],$$

which is your estimate of the probability that $y = +1$. Suppose that the cost matrix is given by

		True classification	
		$+1$ (correct person)	-1 (intruder)
you say	$+1$	0	c_a
	-1	c_r	0

For a new person with fingerprint $\mathbf{x}$, you compute $g(\mathbf{x})$ and you now need to de cide whether to accept or reject the person (i.e., you need a hard classification). So, you will accept if $g(\mathbf{x}) \geq \kappa$, where κ is the threshold.

(a) Define the cost(accept) as your expected cost if you accept the person. Similarly define cost(reject). Show that

$$\begin{aligned} \text{cost(accept)} &= (1 - g(\mathbf{x}))c_a, \\ \text{cost(reject)} &= g(\mathbf{x})c_r. \end{aligned}$$

(b) Use part (a) to derive a condition on $g(\mathbf{x})$ for accepting the person and hence show that

$$\kappa = \frac{c_a}{c_a + c_r}.$$

(c) Use the cost matrices for the Supermarket and CIA applications in Example 1.1 to compute the threshold κ for each of these two cases. Give some intuition for the thresholds you get.

Submission instructions:

- Please read the submission instructions on Canvas for naming conventions. Please use meaningful names for variables and file names.
- Discussions are encouraged, but do not copy your answers from external sources or each other.
- Please cite all the sources you used in your submission.
- For questions on the textbook problems, you can check out the book forum: http://book.caltech.edu/bookforum/