Nicholas Nhat Tran 1002027150

CSE 4309 Assignment 3

Task 2 (15 pts)

We are given these training examples for a linear regression problem:

$x_1 = 5.3$, $t_1 = 9.6$
$x_2 = 7.1$, $t_2 = 4.2$
$x_3 = 6.4$, $t_3 = 2.2$

We want to fit a line to this data, so we want to find the 2-dimensional vector **w** that minimizes $\tilde{E}_D(\mathbf{w})$ as defined in slide 60 of the linear regression slides. What is the value of **w** in the limit where λ approaches positive infinity? Justify your answer. Correct answers with insufficient justification will not receive credit.

In the case where λ approaches infinity, the 2-dimensional vector $w = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ will minimize $\tilde{E}_D(w)$ because the zero vector effectively minimizes the seemingly infinite penalty created by the regularization term. The regularization term needs to be reduced because it would dominate the entire error function, resulting in an insignificant sum of squared errors.

Task 3 (15 pts)

We are given these training examples for a linear regression problem:

$x_1 = 5.3$, $t_1 = 9.6$

$x_2 = 7.1$, $t_2 = 4.2$

$x_3 = 6.4$, $t_3 = 2.2$

We are also given these two lines as possible solutions:

- `f(x) = 3.1x + 4.2`
- `f(x) = 2.4x - 1.5`

Which of these lines is a better solution according to the sum-of-squares criterion? This criterion is defined as function $E_D(w)$ in slide 25 of the linear regression slides. Justify your answer. Correct answers with insufficient justification will not receive credit.

The line f(x) = 2.4x - 1.5 is the better solution according to the sum-of-squares criterion.

Line 1

f(x) = 3.1x + 4.2	Error (y - t)	Squared Error $(y - t)^2$
f(5.3) = 3.1(5.3) + 4.2 = 20.63	20.63 - 9.6 = 11.03	11.03^2 = 121.6609
f(7.1) = 3.1(7.1) + 4.2 = 26.21	26.21 - 4.2 = 22.01	22.01^2 = 484.4401
f(6.4) = 3.1(6.4) + 4.2 = 24.04	24.04 - 2.2 = 21.84	21.84^2 = 476.9856
	Total Squared Error:	1083.0866

Line 2

f(x) = 2.4x - 1.5	Error (y - t)	Squared Error $(y - t)^2$
f(5.3) = 2.4(5.3) - 1.5 = 11.22	11.22 - 9.6 = 1.62	1.62^2 = 2.6244
f(7.1) = 2.4(7.1) - 1.5 = 15.54	15.54 - 4.2 = 11.34	11.34^2 = 128.5956
f(6.4) = 2.4(6.4) - 1.5 = 13.86	13.86 - 2.2 = 11.66	11.66^2 = 135.9556
	Total Squared Error:	267.1756

The lower total SSE is considered the better fit, therefore Line 2 mathematically is proven to be the better fit.

Task 4 (10 pts)

> Your colleague, Bob Hacker, has come up with a new (and better, he claims) training algorithm for the regularized version of linear regression. Bob's algorithm, instead of requiring lambda as an input, computes both the optimal w and the optimal lambda that minimize $\tilde{E}_D(\mathbf{w})$ as defined in slide 60 of the linear regression slides. Bob says that his algorithm is better, since it does not require lambda as a hyperparameter, and instead computes the best lambda automatically (together with the best w).
>
> Your boss is asking for your opinion: should your company use Bob's algorithm instead of the standard training algorithm that you were asked to implement in Task 1? First you check Bob's algorithm, and you verify that indeed it works correctly, and computes both the optimal w and the optimal lambda. Based on that, what is your recommendation? Explain as specifically as possible why Bob's algorithm should or should not replace the standard algorithm.

Bob's algorithm should not replace the standard algorithm because the optimal lambda will always be 0. This occurs because the goal of regularization is to essentially zero out the error (or to minimize) and to reduce or prevent overfitting altogether (which generalizes the function and allows the introduction of new data). Since the minimum lambda of the function is 0, the algorithm mathetically will always zero out the regularization term, making the regularization term inconsequential and return the non-regularized least-squares solution.