

CS 146 - Homework 3

Zoey Nguyen

zoeyn@ucla.edu

April 26, 2021

Question 1.

Partial derivative with respect to a single parameter ω_j .

$$\begin{aligned}\frac{\partial J}{\partial \omega_j} &= \frac{\partial}{\partial \omega_j} J(\omega_j) \\ &= - \sum_n \frac{\partial}{\partial \omega_j} \left(y^{(n)} \ln \sigma(\omega_j x_j^{(n)}) + (1 - y^{(n)}) \ln (1 - \sigma(\omega_j x_j^{(n)})) \right) + \omega_j \\ &= - \sum_n \frac{\partial}{\partial \omega_j} \left(-y^{(n)} \ln (1 + e^{-\omega_j x_j^{(n)}}) + (1 - y^{(n)}) \ln (e^{-\omega_j x_j^{(n)}}) - (1 - y^{(n)}) \ln (1 + e^{-\omega_j x_j^{(n)}}) \right) + \omega_j \\ &= \sum_n \frac{\partial}{\partial \omega_j} \left(\ln (1 + e^{-\omega_j x_j^{(n)}}) + (1 - y^{(n)}) \ln (e^{-\omega_j x_j^{(n)}}) \right) + \omega_j \\ &= \sum_n \left(\frac{1}{1 + e^{-\omega_j x_j^{(n)}}} - y^{(n)} x_j^{(n)} \right) + \omega_j \\ &= \boxed{\sum_n \left(h_{\omega_j}(x^{(n)}) - y^{(n)} x^{(n)} \right) + \omega_j}\end{aligned}$$

Question 2.

Find the MAP estimate of ω^* , where derivative wrt ω is 0. Note taking the log of the MAP function will also give maximum so we will be adding logs rather than multiplying.

$$\begin{aligned}0 &= \frac{\partial}{\partial \omega_j} \ln \left(\frac{1}{(2\pi)^{\frac{m}{2}}} \exp \left\{ \sum_j \omega_j^2 / 2 \right\} \sum_i P(y_i | x_i, \omega) \right) \\ &= \frac{\partial}{\partial \omega_j} \left(\sum_i \ln P(y_i | x_i, \omega) - \sum_j \omega_j^2 / 2 \right) \\ &= \end{aligned}$$

no time

Question 3.

Question 4.

Question 5.

- (a) Plots 1 and 3 can be fully separated using a depth 2 decision tree.
- (b) Plot 2 would have the most complex decision tree because splitting horizontally and vertically does not help narrow down the range very much in the worse case since there are many points with similar horizontal and vertical coordinates along the dividing line.
- (c) Plot 4 is more separable with a depth 4 decision tree. You could, for example, use the four splitting lines: $f_1 = 2, f_2 = -2, f_1 = 2, f_2 = 2$. These partitions will let you identify whether a point is in the blue ring, which is fully contained in the square from $(-2, -2)$ to $(2, 2)$. If it's not in within that range, it has to be part of the outer red ring.

Question 6.

Question 7.

- (a) Let's get k for the first dataset.
- (b) If you use too small a k you will overfit as only the smallest neighborhood is explored, which may not capture the overall cluster that a point is a part of. If you use too large a k you will underfit as points that may not be in a reasonable neighborhood are weighted to the classification of a point.