CS188, Winter 2017
Problem Set 2: part 1
Due feb 9,2017

Rohan Varma

2/9/17

1 Problem 1

(a) Problem 1a

Solution: Our training data are

$$X_1 = [-1, -1], y_1 = -1$$

$$X_2 = [-1, 1], y_2 = -1$$

,

$$X_3 = [1, -1], y_1 = -1$$

,

$$X_4 = [1, 1], y_4 = 1$$

If we let our vector of parameters

$$\Theta = [w_0, w_1, b]$$

and append a 1 as the last element of each training sample X, we obtain X as a 3 x 1 column vector and Θ as a 3 x 1 column vector. Then, we can obtain our predictions with

$$\Theta^T X_i = y_i, i = [1, 2, 3, 4]$$

. In particular, where s represents the sign function:

$$s(-w_0 - w_1 + b) = -1$$

,

$$s(-w_0 + w_1 + b) = -1$$

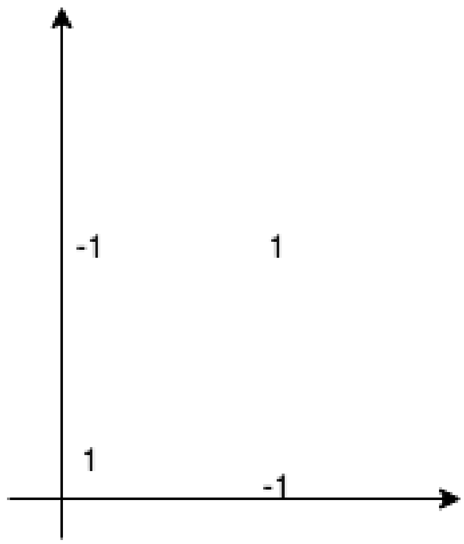
$$s(w_0 - w_1 + b) = -1$$

$$s(w_0 + w_1 + b) = 1$$

We can let $w_0 = 3, w_1 = 4, b = -2$ which satisfies all of the equations. Another set of parameters that satisfies the equations are $w_0 = 7, w_1 = 9, b = -11$.

(b) Problem 1b

Solution: No perceptron exists to compute the XOR function. This is because the data given by XOR are not linearly separable. Perceptrons can only separated data linearly, and the decision boundary is given by the plane $\Theta^T X = 0$. However, the data for XOR looks as follows:



which we cannot draw a single line through to separate.

2 Problem 2

Problem 2a

Solution: We have the negative log-likelihood $J(\theta) == \sum_{n=1}^{N} y_n log(\sigma(\theta^T x_n)) + (1 - y_n) log(1 - \sigma(\theta^T x_n))$. Differentiating with respect to a parameter θ_j,

$$\frac{\delta J}{\delta \theta_j} = -\sum_{n=1}^{N} y_n x_{n,j}(1 - \sigma(\theta^T x_n)) - x_{n,j}(1 - y_n)(\sigma(\theta^T x_n))$$

Simplifying,

$$-\sum_{n=1}^{N} y_n x_{n,j} - y_n x_{n,j} h_\theta(x_n) - x_{n,j} h_\theta(x_n) + x_{n,j} y_n h_\theta(x_n)$$

$$= -\sum_{n=1}^{N} x_{n,j}(y_n - h_\theta(x_n)$$

giving us

$$\frac{\delta J}{\delta \theta_j} = \sum_{n=1}^{N} x_{n,j}(h_\theta(x_n) - y_n)$$

where $x_{n,j}$ denotes the jth feature in the nth training example.

Problem 2b

Solution: Substituting the sigmoid function and differentiating the quantity from 2A with respect to θ_j, we have

$$\frac{\delta^2 J}{\delta \theta_j \theta_k} = \sum_{n=1}^{N} x_{n,j} x_{n,k}(\sigma(\theta^T x_n))(1 - \sigma(\theta^T x_n))$$

which is one of the components of the vector of partial second derivatives. This is some arbitrary element in the Hessian $H_{j,k}$. In general, the Hessian is composed of a vector of these θ_j, θ_k partial derivatives, so we have $x_{n,j}, x_{n,k}$ as vectors, giving us their product as $x_n x_n^T$ for the n training example. This gives us the Hessian:

$$H = \sum_{n=1}^{N} h_\theta(x_n)(1 - h_\theta(x_n)) x_n x_n^T$$

Problem 2c

Solution: J is convex if $z^T H z = \sum_{j,k} z_j z_k H_{j,k} \geq 0$ By substitution, we have:

$$\sum_{j,k} z_j z_k \sum_{n=1}^{N} x_{n,j} x_{n,k} \sigma(\theta^T x_n)(1 - \sigma(\theta^T x_n))$$

From 2b, we can rewrite the inner product, obtaining:

$$\sum_{j,k} z_j z_k \sum_{n=1}^{N} h_\theta(x_n)(1 - h_\theta(x_n)) x_n x_n^T$$

. The inner product is greater than or equal to 0 since it is the product of the square of a vector's magnitude (always positive or 0) and two probabilities (which are both positive or 0). So we are left with

$$\sum_{j,k} z_j z_k H, H \geq 0$$

. Since we know $\sum_{j,k} z_j z_k = ||z||^2 \geq 0$ by the definition of a vector's dot product with itself, and that sum in this case is just being multiplied each time by a number that's greater than or equal to 0, this quantity has to be greater than or equal to 0.

3 Problem 3

Problem 3a

Solution:

$$\frac{\delta J}{\delta \theta_0} = 2\sum_{n=1}^{N} w_n(\theta_0 + \theta_1 x_{n,1} - y_n)$$

$$\frac{\delta J}{\delta \theta_1} = 2\sum_{n=1}^{N} w_n x_{n,1}(\theta_0 + \theta_1 x_{n,1} - y_n)$$

Problem 3b

Solution:

$$\frac{\delta J}{\delta \theta_0} = 0$$

$$\theta_0 \sum_{n=1}^{N} w_n + \theta_1 \sum_{n=1}^{N} x_{n,1} w_n = \sum_{n=1}^{N} w_n y_n$$

We obtain an expression for θ_0 in terms of θ_1:

$$\theta_0 = \frac{\sum_{n=1}^{N} w_n y_n - \theta_1 \sum_{n=1}^{N} x_{n,1} w_n}{\sum_{n=1}^{N} w_n}$$

Now solving explicitly for θ_1

$$\frac{\delta J}{\delta \theta_1} = 0$$

$$\theta_0 \sum_{n=1}^{N} w_n x_{n,1} + \theta_1 \sum_{n=1}^{N} x_{n,1}^2 w_n = \sum_{n=1}^{N} w_n x_{n,1} y_n$$

Substituting,

$$\frac{\sum_{n=1}^{N} w_n y_n - \theta_1 \sum_{n=1}^{N} x_{n,1} w_n}{\sum_{n=1}^{N} w_n} \sum_{n=1}^{N} w_n x_{n,1} + \theta_1 \sum_{n=1}^{N} w_n x_{n,1}^2 = \sum_{n=1}^{N} w_n x_{n,1} y_n$$

$$(\sum_{n=1}^{N} w_n y_n - \theta_1 \sum_{n=1}^{N} x_{n,1} w_n)(\sum_{n=1}^{N} w_n x_{n,1}) + \theta_1 (\sum_{n=1}^{N} w_n)(\sum_{n=1}^{N} w_n x_{n,1}^2) = (\sum_{n=1}^{N} w_n)(\sum_{n=1}^{N} w_n x_{n,1} y_n)$$

Expanding and factoring to isolate θ_1, we get:

$$\theta_1((\sum_{n=1}^{N} w_n)(\sum_{n=1}^{N} w_n x_{n,1}^2) - (\sum_{n=1}^{N} x_{n,1} w_n)^2) = (\sum_{n=1}^{N} w_n)(\sum_{n=1}^{N} w_n x_{n,1} y_n) - (\sum_{n=1}^{N} w_n y_n)(\sum_{n=1}^{N} w_n x_{n,1})$$

Dividing to solve for theta 1, we get

$$\theta_1 = \frac{(\sum_{n=1}^{N} w_n)(\sum_{n=1}^{N} w_n x_{n,1} y_n) - (\sum_{n=1}^{N} w_n y_n)(\sum_{n=1}^{N} w_n x_{n,1})}{((\sum_{n=1}^{N} w_n)(\sum_{n=1}^{N} w_n x_{n,1}^2) - (\sum_{n=1}^{N} x_{n,1} w_n)^2)}$$

Now we can go back to our expression for θ_0 that was in terms of θ_1, and plug the explicit form of θ_1 back in to get θ_0 explicitly as well.