Machine Learning WS 2018

Solution to Assignment 4

Submitted By : Vindhya Singh **Enrollment Number :** 03693296
Collaborated By : Wasiq Rumaney **Enrollment Number :** 03694978

Solution 1

The posterior class probability in the given case is given by logistic sigmoid acting on a linear function a(x) given by :

$$a(\mathbf{x}) = (\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2)^{\mathrm{T}}\mathbf{x} + \ln g(\boldsymbol{\lambda}_1) - \ln g(\boldsymbol{\lambda}_2) + \ln p(\mathcal{C}_1) - \ln p(\mathcal{C}_2).$$

[Bishop, Chapter 4, 4.2.4 Exponential Family]

Thus, it is a Bernoulli distribution

Solution 2

The maximum likelihood solution for the decision boundary w of a logistic regression model should be able to classify the data into their respective classes. MLE solution occurs in linearly separable data set only when $\sigma = 0.5, \textit{equivalently } w^T \phi = 0$. Also, the MLE for linearly separable data takes the magnitude to infinity.

MLE for Logistic Regression is not a winning choice because of primarily 2 reasons :

1. It causes **over fitting.**
2. When training is done on a linearly separable data, it can break down. In a linearly separable data, all those data points from which the function passes through will have **equal likelihood**. Therefore, there will exist many solutions when we try to optimize the likelihood which will render the main objective of the MLE meaningless unless **we use a prior and find a MAP solution for w, or by using regularization in the error function to correct these shortcomings.**

Solution 3

We know that, the Sigmoid function is defined as :

$$\sigma(a)=\frac{1}{(1+\exp(-a))} \qquad \text{[Slide 25 Lecture 4]} \qquad \textbf{(1)}$$

where a is defined as $a=\ln(\frac{p(x|y=1).\,p(y=1)}{p(x|y=0).\,p(y=0)})$ **(2)**

And for Multi-class Logistic Regression, Softmax is defined as :

$$\sigma(x)_i=\frac{\exp(x_i)}{\sum_{k=1}^{k}\exp(x_k)} \qquad \text{[Slide 29 Lecture 4]} \qquad \textbf{(3)}$$

Thus from the equations 1 and 2, the predicted probabilities in 2-class case is:

$$P(y_i=0)=\frac{e^{-(\beta X_i)}}{1+e^{-(\beta_o X_i)}} \quad \text{and} \quad P(y_i=1)=1-P(y_i=0)=\frac{1}{1+e^{-(\beta X_i)}} \qquad \textbf{(4)}$$

Also, for multi-class logistic regression, using the eq 3, for softmax, we see that,

$$P(y_i=0)=\frac{e^{-(\beta X_i)}}{1+e^{-(\beta X_i)}} \quad \text{and} \quad P(y_i=1)=1-P(y_i=0)=\frac{1}{1+e^{-(\beta X_i)}}$$

If we consider only 2 classes, then using 3, Softmax can be written as :

$$P(y=1|X)=\frac{1}{\sum_{i=0}^{1}\exp^{-(\theta_1-\theta_0)^{T.x}}}$$

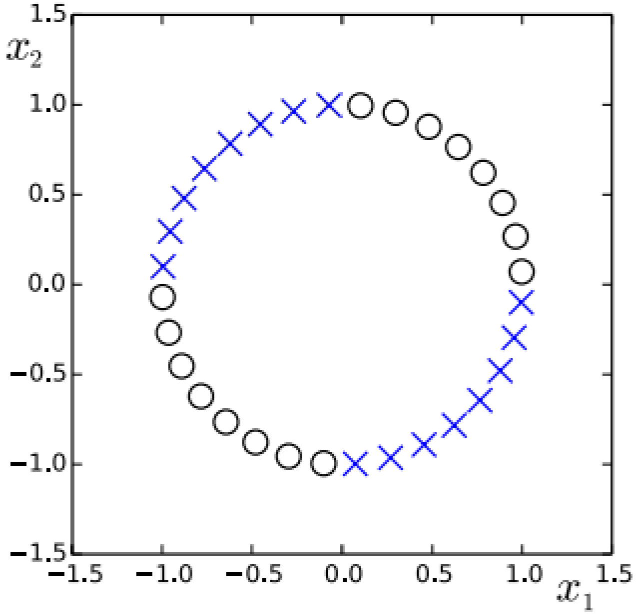
$$=\frac{1}{1+\exp^{-(\theta_1-\theta_0)^{T.x}}}$$

$$=\frac{1}{1+\exp^{-(\theta)^{T.x}}} \quad \text{where} \quad \theta=\theta_1-\theta_0=\sigma(\theta^T*X)$$

which is equivalent to the sigmoid function as defined above in (4).

Thus, Softmax is a special case of Sigmoid function for 2 class case.

Solution 4



From the graph above, we can see that the given points are not linearly separable into two different classes. Therefore, taking a basis function into account $\phi(x_1, x_2)$.

Clearly from the graph above, the (X) and (O) lie in the opposite quadrants with the values of (x_1, x_2) reversed and the signs changed.

For ex. , a coordinate of X(cross) can be **P(0.5,-1)** or **Q(-0.5,1)** so if you multiply, the product will always remain the same i.e. (0.5*-1) = (-0.5*1). So, the basis function to make them linearly separable will be

$$\phi(x_1, x_2) = x_1 x_2$$