Homework – DSC 540 Week Three

Colton Proctor

Grand Canyon University

Homework Week 3 DSC540

Part 1:

Begin with the definition

$$P(x|y_q) = \frac{1}{(2\pi)^{\frac{n}{2}}\,|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_q)^T\Sigma^{-1}(x-\mu_q)\right)$$

Then from Bayes Theorem

$$P(x|y_q) = \frac{P(y_q|x)P(x)}{P(y_q)}$$

This leads to

$$P(y_q|x) = \frac{P(x|y_q)P(y_q)}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}\sum_{q=1,2}P(x|y_q)P(y_q)}$$

$$= \frac{\exp\left(-\frac{1}{2}(x-\mu_q)^T\Sigma^{-1}(x-\mu_q)\right)P(y_q)}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}\sum_{q=1,2}P(x|y_q)P(y_q)}$$

Since the denominator is the same for both classes the log can be taken

$$\ln\left(P(y_q|x)\right) = -\frac{(x-\mu_q)^T\Sigma^{-1}(x-\mu_q)}{2} + \ln\left(P(y_q)\right)$$

$$= -\frac{x^T\Sigma^{-1}x - 2x^T\Sigma^{-1}\mu_q + \mu_q^T\Sigma^{-1}\mu_q}{2} + \ln\left(P(y_q)\right)$$

Note that $x^T\Sigma^{-1}x$ is a constant term for both classes so then

$$g_q(x) = \ln\left(P(y_q|x)\right) = \mu_q^T\Sigma^{-1}x - \frac{1}{2}\mu_q^T\Sigma^{-1}\mu_q + \ln\left(P(y_q)\right)$$

Therefore, that is the discriminant function

Then it follows that

$$g_1(x) = \ln\left(P(y_1|x)\right) = \mu_1^T\Sigma^{-1}x - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \ln\left(P(y_1)\right)$$

$$g_2(x) = \ln\left(P(y_2|x)\right) = \mu_2^T\Sigma^{-1}x - \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \ln\left(P(y_2)\right)$$

$$g(x) = g_1(x) - g_2(x) = 0$$

$$= \mu_1^T \Sigma^{-1} x - \frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1 + \ln\big(P(y_1)\big) - \Big(\mu_2^T \Sigma^{-1} x - \frac{1}{2}\mu_2^T \Sigma^{-1}\mu_2 + \ln\big(P(y_2)\big)\Big)$$

$$= \mu_1^T \Sigma^{-1} x - \frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1 + \ln\big(P(y_1)\big) - \mu_2^T \Sigma^{-1} x + \frac{1}{2}\mu_2^T \Sigma^{-1}\mu_2 - \ln\big(P(y_2)\big)$$

$$= (\mu_1^T - \mu_2^T)\Sigma^{-1} x - \frac{1}{2}(\mu_1^T \Sigma^{-1}\mu_1 - \mu_2^T \Sigma^{-1}\mu_2) + \ln\left(\frac{P(y_1)}{P(y_2)}\right)$$

$$q.e.d.$$

Part Two: Perform two iterations of the gradient algorithm to find the minima of

$$E(w) = 2w_1^2 + 2w_1 w_2 + 5w_2^2$$

With starting point

$$w = [2 \quad -2]^T$$

Initial E(w) is

$$E(w) = (2)(2)^2 + (2)(2)(-2) + 5(-2)^2$$

$$= 8 - 8 + 20 = 20$$

The partial derivations with respect to w_1 and w_2 are

$$\frac{\partial E(w)}{\partial w_1} = 4w_1 + 2w_2$$

$$\frac{\partial E(w)}{\partial w_2} = 2w_1 + 10w_2$$

Then to iterate the algorithm assume that the base learning rate is .1

Iteration One

$$W_1 = w_1 - \alpha \frac{\partial E(w)}{\partial w_1}$$

$$= 2 - (0.1)[(4)(2) + (2)(-2)$$

$$= 2 - .4 = 1.6$$

$$W_2 = w_2 - \alpha \frac{\partial E(w)}{\partial w_2}$$

$$= -2 - .1[(2)(2) + (10)(-2)]$$

$$= -2 + 1.6 = -.4$$

$$E(w_1) = 2(1.6)^2 + 2(1.6)(-.4) + 5(-.4)^2$$

$$= 5.12 + (1.28) + .8$$

$$= 4.64$$

Iteration Two

$$W_1 = w_1 - \alpha \frac{\partial E(w)}{\partial w_1}$$

$$= 1.6 - .1[(4)(1.6) + (2)(-.4)]$$

$$= 1.6 - .56 = 1.04$$

$$W_2 = w_2 - \alpha \frac{\partial E(w)}{\partial w_2}$$

$$= -.4 - .1[2(1.6) + 10(-.4)]$$

$$= -.4 + .08 = -.32$$

$$E(w_2) = 2(1.04)^2 + 2(1.04)(-.32) + 5(-.32)^2$$

$$= 2.16 + (-.66) + .512$$

$$= 2.012$$

Then after two iterations

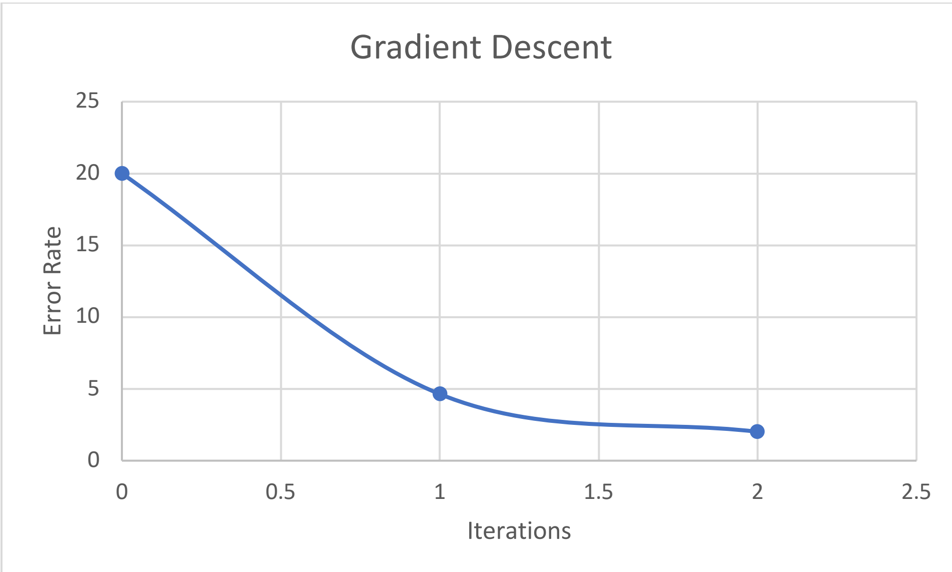
W_1 is 1.04

W_2 is -.32

Homework Week 3 DSC540

E(w) = 2.012

Plotted this looks like



Part Three:

To explain why logistic regression a non-linear regression problem, look at the mathematical definition. First, logistic regression is a non-linear transformation of $\beta^T x$.

Then the probability is:

$$P(Y=1): p = \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2}}{1+e^{\alpha+\beta_1 x_1+\beta_2 x_2}}$$

Then the odds are:

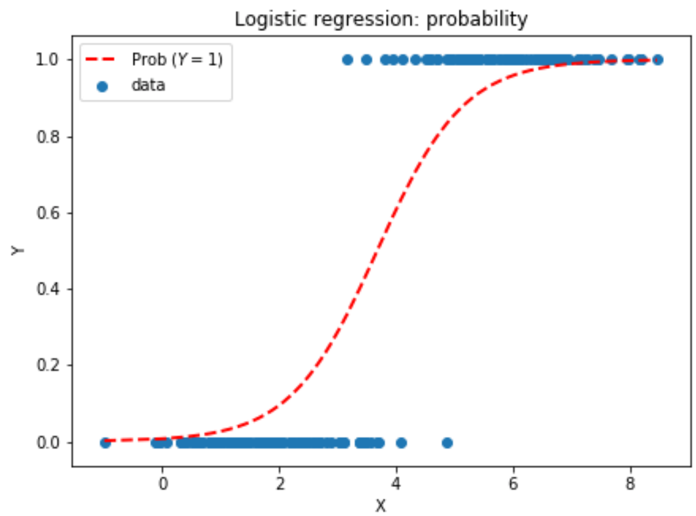
$$Odds\ (Y=1): \left(\frac{p}{1-p}\right) = e^{\alpha+\beta_1 x_1+\beta_2 x_2}$$

The log odds are then:

$$Log\ Odds\ (Y=1): \log\left(\frac{p}{1-p}\right) = \alpha+\beta_1 x_1+\beta_2 x_2$$

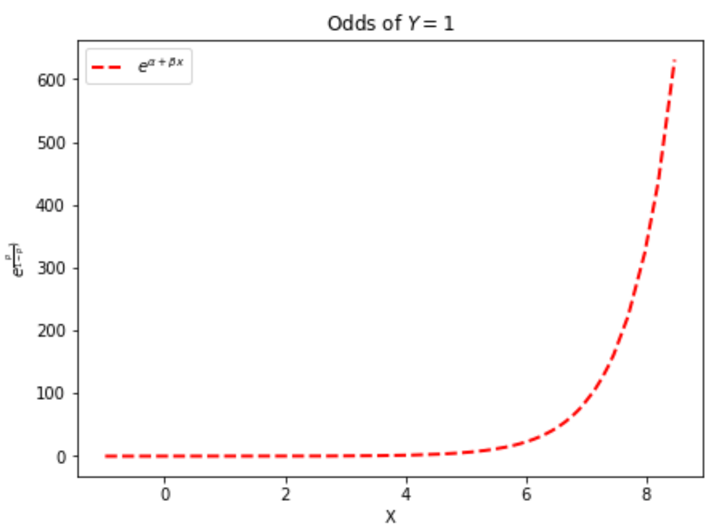
As can be seen by these definitions of logistic regression, it is non-linear it terms of the probability and the odds. However, it is linear in terms of the log odds. This can be further shown in a graphed example.

Plotting the probability is as follows:



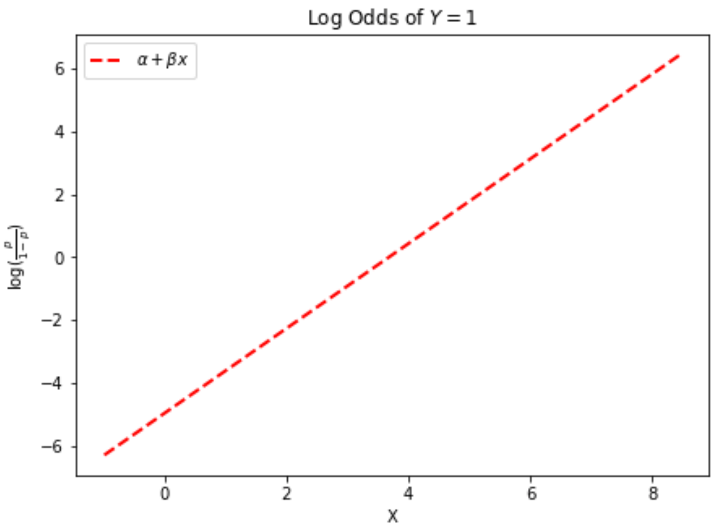
This shows that the non-linear relationship of probability.

Plotting the odds then looks like:



This is also a non-linear relationship, as it is clearly exponential.

Finally, the linear relationship of logistic and its log odds is shown in the following plot.



Then as can be seen by both the plots and the equations, logistic regression is a non-linear regression problem. However, it can be described as a linear problem if the log odds are used.

References

Donges, N. (n.d.). *Gradient descent: An introduction to 1 of Machine Learning's most popular algorithms*. Built In. Retrieved November 16, 2021, from https://builtin.com/data-science/gradient-descent.

Gopal, M. (2019). *Applied machine learning.* McGraw-Hill Education. ISBN-13-9781260456844

ironmanironman 59266 silver badges2222 bronze badges, Xavier Bourret SicotteXavier Bourret Sicotte 7, & KolmogorovwannabeKolmogorovwannabe 1781414 bronze badges. (1966, November 1). *Logistic regression is a nonlinear regression problem?* Cross Validated. Retrieved November 16, 2021, from https://stats.stackexchange.com/questions/365391/logistic-regression-is-a-nonlinear-regression-problem.

Pandey, P. (2021, June 10). *Understanding the mathematics behind gradient descent.* Medium. Retrieved November 16, 2021, from https://towardsdatascience.com/understanding-the-mathematics-behind-gradient-descent-dde5dc9be06e.