DSC-540: Machine Learning for Data Science

Topic 3: Gradient Descent and Logistic Regression

Rajkumar Cheney Krishnan

17 November 2021

Part 1

According to Bayes' Theorem, the posteriors can be calculated as follows:

$$P(y_k|x) = \frac{P(y_k)(P x|y_k)}{\sum_{q=1}^{M} P(x|y_q)\,P(y_q)}$$

Plugging in given expression for $P(x|y_q)$,

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

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

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

Linear discriminant function is denoted as $g_q(x)$

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

#

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

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

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

$$g_1(x) = g_2(x)$$

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

Part 2

$$E(w) = 2w_1^2 + 2w_1w_2 + 5w_2^2$$

$$\frac{dE(w)}{dw_1} = 4w_1 + 2w_2$$

$$\frac{dE(w)}{dw_2} = 2w_1 + 10w_2$$

First iteration, using learning rate $\alpha = 0.1$

$$w_1 = w_1 - \alpha \frac{dE(w)}{dw_1}$$

$$= 2 - (0.1)(4 \times 2 + 2 \times (-2))$$

$$= 1.6$$

$$w_2 = w_2 - \alpha \frac{dE(w)}{dw_2}$$

$$= -2 - (0.1)[(2 \times 2) + (10 \times (-2))]$$

$$= -0.4$$

$$E(w)_1 = 2 \times (1.6)^2 + 2(1.6)(-0.4) + 5(-0.4)^2$$

$$= 4.64$$

Second iteration, again using learning rate $\alpha = 0.1$

$$w_1 = w_1 - \alpha \frac{dE(w)}{dw_1}$$

$$= 1.6 - (0.1)[4(1.6) + 2(-0.4)]$$

$$= 1.04$$

$$w_2 = w_2 - \alpha \frac{dE(w)}{dw_2}$$

$$= -0.4 - (0.1)[2(1.6) + 10(-0.4)]$$

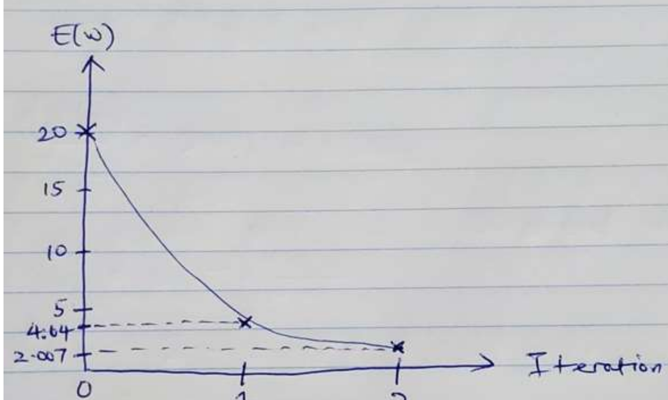
$$= -0.32$$

$$E(w)_2 = 2 \times (1.04)^2 + 2(1.04)(-0.32) + 5(-0.32)^2$$

$$= 2.007$$

$$E(w)_0 = 2 \times (2)^2 + (2)(2)(-2) + 5(-2)^2$$

$$= 20$$



Part 3

Linear regression models, by themselves, are meant for estimating continuous values over a wide range of values, and yield predictions that are not necessarily restricted between 0 and 1. This is inherently unsuited for prediction in a binary classification scenario, even if the 2 outcomes are quantified as 0 and 1. Based on these characteristics, logistic regression is a nonlinear regression problem.

However, it is possible to transform the logistic regression into a model resembling a linear regression model by using the concept of odds and taking the log of odds. While odds is a number between 0 and infinity, the transformation to log(odds), with logit as a dependent variable, converts the problem to a linear regression task, with output strictly between 0 and 1, facilitating the binary classification requirement of the original logistics regression problem.