Topic 3 Assignment: Gradient Descent and Logistic Regression

This assignment mixes statistical theory and application, in the form of three fairly short problems. Perform the tasks described in each.

Part 1

In the case of normally distributed classes, discriminant functions are linear (straight lines, planes, and hyperplanes for two-, three-, and n-dimensional feature vectors, respectively) when the covariances matrices of corresponding classes are equal. Confirm this by deriving discriminant functions for a binary classification problem.

Given:

$$P(\mathbf{x}|y_q) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_q)^T\ \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_q)\right);\ q = 1, 2$$

Prove that linear discriminant functions

$$g_q(\mathbf{x}) = \boldsymbol{\mu}_q^T\mathbf{\Sigma}^{-1}\,\mathbf{x} - \tfrac{1}{2}\,\boldsymbol{\mu}_q^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_q + \ln P(y_q);\ q = 1, 2$$

And decision boundary $g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x}) = 0$ is given by

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0$$

$$\mathbf{w}^T\mathbf{x} + w_0 = (\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_2^T)\mathbf{\Sigma}^{-1}\mathbf{x} - \tfrac{1}{2}(\boldsymbol{\mu}_1^T\,\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T\,\mathbf{\Sigma}^{-1}\,\boldsymbol{\mu}_2) + \ln\frac{P(y_1)}{P(y_2)}$$

*(**Hint**: Use equations 3.61–3.62 in the textbook)*

Using Bayes' Theorem: $P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_i^n P(A|B_i)P(B_i)}$

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

$$= \frac{e^{-\frac{1}{2}(x-\mu_q)\Sigma^{-1}(x-\mu_q)}\, p(y_q)}{2\pi^{n/2}|\Sigma|^{1/2}\sum P(x|y_q)P(y_q)}$$, take $ln(P(y_q|x))$

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

- Eliminate $x^T\Sigma^{-1}$ from both classes

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

Which gives us the discriminant function.

Part 2

Perform two iterations of the gradient algorithm to find the minima of $E(\boldsymbol{w}) = 2w_1^2 + 2w_1w_2 + 5w_2^2$

The starting point is $\boldsymbol{w} = [\mathbf{2} \quad \mathbf{-2}]^T$

Draw the contours and show your learning path graphically.

Finding the minima of $E(w) = 2w_1^2 + 2w_1w_2 + 5w_2^2$ with only 2 iterations of the gradient algorithm.

- $E(w) = 20$, starting point: $w = [2 \; -2]^T$
- $\frac{\partial E(w)}{\partial w_1} = 4w_1 + 2w_2 + 0 = 4(2) + 2(-2) = 4$
- $\frac{\partial E(w)}{\partial w_2} = 0 + 2w_1 + 10w_2 = 2(2) + 10(-2) = -16$
- Gradient algorithm: $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} j(\theta)$

Iteration 1: learning rate $= \alpha = 0.1$

$w_1 := w_1 - \alpha \frac{\partial}{\partial w_1} E(w) = 2 - .1(4) = 1.6$

$w_2 := w_2 - \alpha \frac{\partial}{\partial w_2} E(w) = -2 - .1(-16) = -0.4$

$E(w)_1 = 2w_1^2 + 2w_1w_2 + 5w_2^2 = 2(1.6)^2 + 2(1.6)(-.4) + 5(-.4)^2$

$= 5.12 - 1.28 + 0.8 = 4.64$

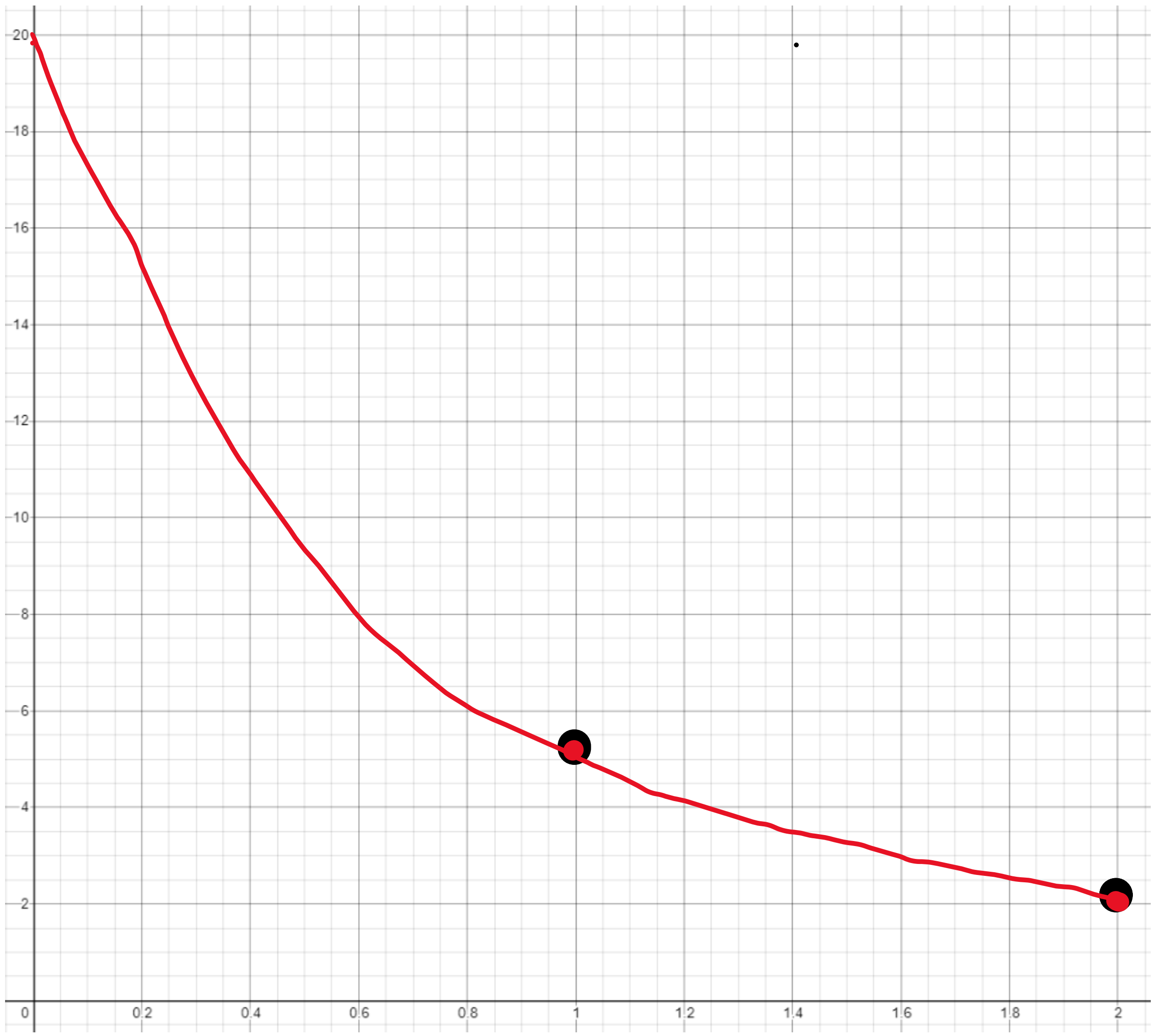
Iteration 2:

$W_{12} = 1.6 - 0.1(4(1.6) + 2(-0.4)) = 1.04$

$W_{22} = -0.4 - 0.1(2(1.6) + 10(-0.4)) = -0.32$

$E(w)_2 = 2(1.04)^2 + 2(1.04)(-0.32) + 5(-0.32)^2 = 2.007$

$Min: E(w)_2 = 2.007$



x-axis: Iterations 1 and 2
y-axis: E(w)
Points: (0,20) (1, 4.64) (2, 2.007)

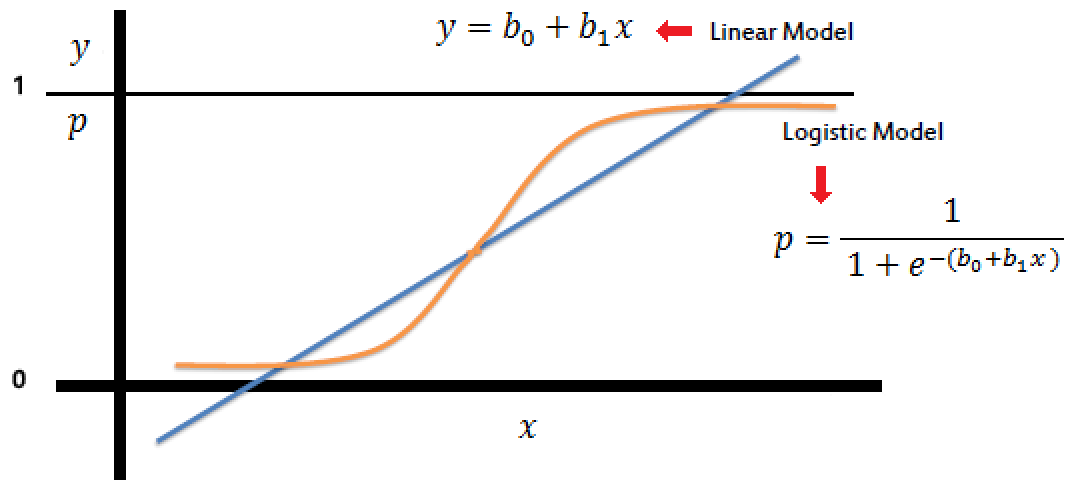
Part 3

Show that logistic regression is a nonlinear regression problem. Is it possible to treat logistic discrimination in terms of an equivalent linear regression problem? Justify your answer.

The easiest way to denote the difference between Logistic and linear regression is the formation of the probability curve. Linear regresison models focus on the conditional probability of distribution, while logistic regression models find the probability of the output in terms of input.

Linear Regression Model: $p = b_0 + b_1 x$

Logistic Regression Model: $p = \frac{1}{1+e^{-(b_0+b_1 x)}}$



Logistic regression can be utilized as a linear classifer to separate observations by linear boundaries to form a distinct class. It undergoes linear transformation through logarithmic means by taking the log of the odds. This transformation allows logistic discrimination to be treated as linear classification.

$$\log\left(\frac{1}{1+e^{-(b_0+b_1 x)}}\right) = b_0 + b_1 x$$

References:

Humanunsupervied, (2019), [L1] Regression (Univariate). Cost Function. Hypothesis. Gradient., https://humanunsupervised.github.io/humanunsupervised.com/topics/L1-regression-hypothesis-cost-gradient.html

Sayad Saed, (2021), Logistic Regression, Rutgers University, https://www.saedsayad.com/logistic_regression.htm