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:

Prove that linear discriminant functions

And decision boundary

$g(\mathbf{x}) = g_1(\mathbf{x}) – g_2(\mathbf{x}) = 0$ is given by*(**Hint**:Use equations 3.61–3.62 in the textbook)*

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

$$\frac{\partial E(w)}{\partial w_1} = 4w_1 + 2w_2 \ , \quad \frac{\partial E(w)}{\partial w_2} = 2w_1 + 10w_2$$

First iteration $\quad w_1 = w_1 - \alpha \frac{\partial E(w)}{\partial w_1}$

$$= 2 - (0.1)\left[4 \times 2 + 2(-2)\right]$$

$$= 1.6$$

$$w_2 = w_2 - \alpha \frac{\partial E(w)}{\partial w_2} = -2 - 0.1\left[2(2) + 10(-2)\right]$$

$$= -2 + 1.6 = -0.4$$

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

$$= 4.64$$

Second Iteration

$$w_1 = w_1 - \alpha \frac{\partial E(w)}{\partial w_1} = 1.6 - 0.1\left[4(1.6) + 2(-0.4)\right]$$

$$= 1.6 - 0.56$$

$$w_1 = 1.04$$

$$w_2 = w_2 - \alpha \frac{\partial E(w)}{\partial w_2} = -0.4 - 0.1\left[2(1.6) + 10(-0.4)\right]$$

$$= -0.4 + 0.08 = -0.32$$

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

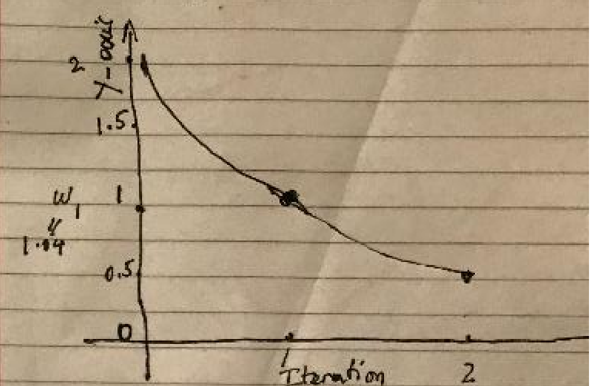
$$= 2.007$$

$w_1 = 1.04$

$w_2 = -0.32$

$E(w)_1 = 4.64$

$E(w)_2 = 2.007$



Part 2

Perform two iterations of the gradient algorithm to find the minima of

E(w)=2w1^2+2w1w2+5w2^2

The starting point is w =[2-2]^T

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

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

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

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

Draw the contours and show your learning path graphically.

$$From\ Bayes'\ theorem,\ we\ have$$

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

$$= \frac{exp(-0.5(x-\mu_q)^T\Sigma^{-1}(x-\mu_q))P(y_q)}{(2\pi)^{n/2}|\Sigma|^{1/2}\sum_{q=1,2}P(x|y_q)P(y_q))}$$

$Note\ that\ the\ denominator\ is\ same\ for\ both\ classes,\ so\ by\ taking\ log$
$of\ above\ expression\ we\ get$

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

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

$now\ note,\ that\ x^T\Sigma^{-1}x,\ is\ constant\ term\ for\ both\ classes,\ so\ we\ can\ drop$
$it\ from\ above\ expression\ then\ simplifying\ we\ get,$

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

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.

A linear classifier, or logistic regression, is traditionally used when there is a linear boundary between classes in feature space. However, if we had a better sense of what the decision boundary looked like, that could be rectified...
Logistic regression is known and used as a linear classifier. It is used to come up with a hyperplane in feature space that separates observations belonging to a class from all other observations that do not belong to the class. The decision boundary is therefore linear. Several robust and efficient implementations (e.g. scikit-learn) are readily available to use logistic regression as a linear classifier.
Logistic regression makes core assumptions about the observations such as IID (each observation is independent of the others and they have the same probability distribution), but the use of a linear decision boundary is not one of them. Linear decision boundaries are used due to their simplicity, following the Zen maxim - when in doubt, simplify. When the decision boundary appears to be nonlinear, it might make sense to formulate logistic regression with a nonlinear model and see how much better we can do.

Or in another way:
Recall that the Logistic regression model is a non-linear transformation of

$\beta_{T}xe\hat{}\beta Tx$

- **Probability** of
- $(Y=1)$
- $(Y=1)$: $p=e_{\alpha+\beta1x1+\beta2x2}1+e_{\alpha+\beta1x1+\beta2x2}$ $p=e\alpha+\beta1x1+\beta2x21+e\alpha+\beta1x1+\beta2x2$
- **Odds** of
- $(Y=1)$
- $(Y=1)$:
- $(p/1-p)=e_{\alpha+\beta_1x_1+\beta_2x_2}$
- $(p1-p)=e\alpha+\beta1x1+\beta2x2$
- **Log Odds** of
- $(Y=1)$
- $(Y=1)$:
- $\log(p/1-p)=\alpha+\beta_1x_1+\beta_2x_2$
- $\log(p1-p)=\alpha+\beta1x1+\beta2x2$

So to answer your question, Logistic regression is indeed non-linear in terms of Odds and Probability, however, it is linear in terms of Log Odds.

A simple example

Fitting a logistic regression model on the following toy example gives the coefficients

$\alpha=-5.05$

$\alpha=-5.05$ and

$\beta=1.3$

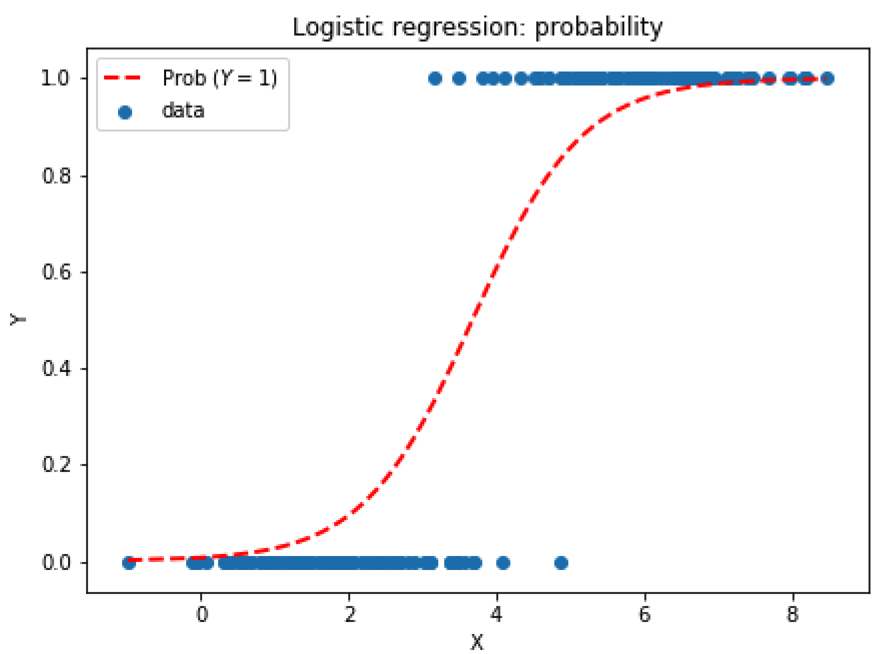
$\beta=1.3$

Plotting the probability

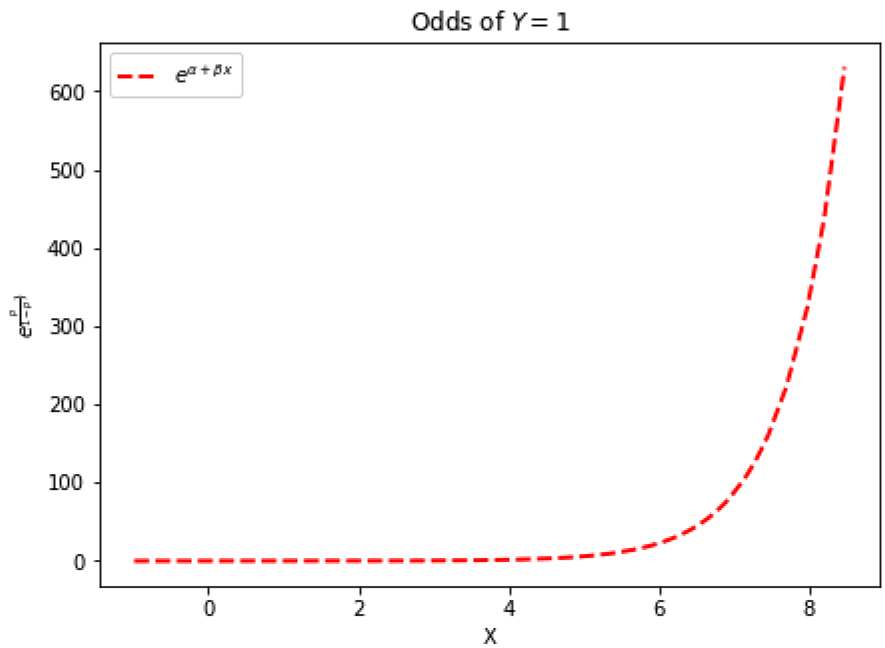
$P(Y=1)$

$P(Y=1)$ as a function of X

X clearly shows the non-linear relationship



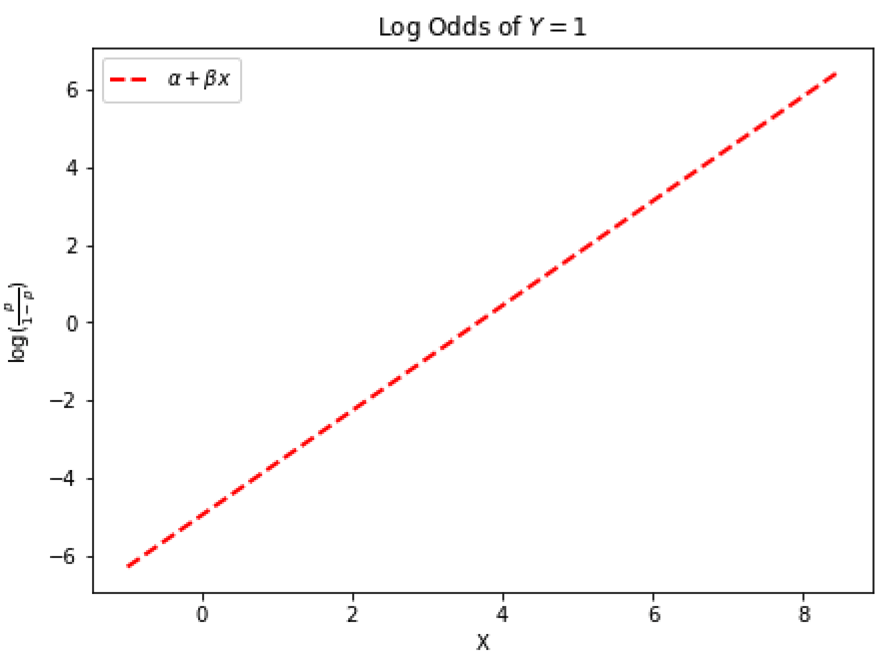
The *Odds* of Y, Y being 1 given X, X is also non-linear.



Finally the log odds of

Y

Y being 1 is a linear relationship



References:

1. https://stats.stackexchange.com/questions/365391/logistic-regression-is-a-nonlinear-regression-problem.
2. https://xplordat.com/2019/03/13/logistic-regression-as-a-nonlinear-classifier
3. Gopal, M. (2019). *Applied machine learning.* McGraw-Hill Education. ISBN-13-9781260456844