Max Shi

CS559

Homework 2

1) Through the results of the three regressions, we see that the score of the pair between Iris-versicolor and Iris-virginica have a score of 97.0%, while the other two pairs have a score of 100%, which mean that the first pair are not linearly separable, as there is no line directly separating these pairs of points. The other two pairs, however, are linearly separable, as seen in the models. A score of 100% means that the regression for these pairs have developed a line that completely separates the two data classes.

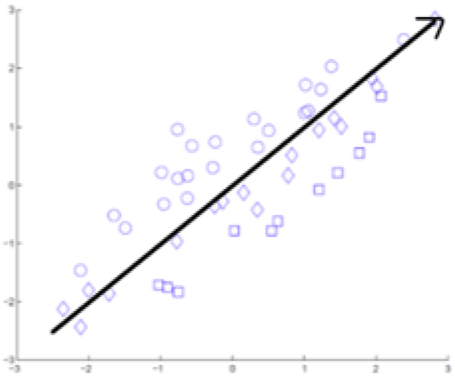
2)

$$f(x) = \sigma(w^T x + w_0)$$
$$\sigma(a) = \frac{1}{1+e^{-a}}$$
$$f(x) = \frac{1}{1+e^{-(w^T x + w_0)}} = \left(1 + e^{-w^T x} * e^{-w_0}\right)^{-1}$$
$$\frac{d}{dw} f(x) = \sigma(w^T x + w_0)\big(1 - \sigma(w^T x + w_0)\big) * \frac{d}{dw}(w^T x + w_0) = f(x)\big(1 - f(x)\big) * x$$
$$\epsilon(w) = -\sum_{n=1}^{N} \left(y_n \ln f(x_n) + (1 - y_n)\ln\big(1 - f(x_n)\big)\right)$$
$$\frac{d}{dw}\epsilon(w) = \frac{d}{dw} - \sum_{n=1}^{N} \left(y_n \ln f(x_n) + (1 - y_n)\ln\big(1 - f(x_n)\big)\right)$$
$$= -\sum_{n=1}^{N} \left(\frac{d}{dw}(y_n \ln f(x_n)) + \frac{d}{dw}\Big((1 - y_n)\ln\big(1 - f(x_n)\big)\Big)\right)$$
$$\frac{d}{dw}(y_n \ln f(x_n)) = \frac{y_n}{f(x_n)} * f(x_n)\big(1 - f(x_n)\big) * x_n = x_n y_n\big(1 - f(x_n)\big) = x_n y_n - f(x_n) x_n y_n$$
$$\frac{d}{dw}\Big((1 - y_n)\ln\big(1 - f(x_n)\big)\Big) = \frac{1 - y_n}{1 - f(x_n)} * -f(x_n) * \big(1 - f(x_n)\big) * x_n = -f(x_n) x_n + y_n f(x_n) x_n$$
$$\frac{d}{dw}\epsilon(w) = -\sum_{n=1}^{N} \left(\frac{d}{dw}(y_n \ln f(x_n)) + \frac{d}{dw}\Big((1 - y_n)\ln\big(1 - f(x_n)\big)\Big)\right)$$
$$= -\sum_{n=1}^{N} (x_n y_n - f(x_n) x_n y_n - f(x_n) x_n + y_n f(x_n) x_n) = -\sum_{n=1}^{N} (x_n y_n - f(x_n) x_n)$$
$$= -\sum_{n=1}^{N} ((y_n - f(x_n) x_n) = \sum_{n=1}^{N} (f(x_n) - y_n) x_n$$

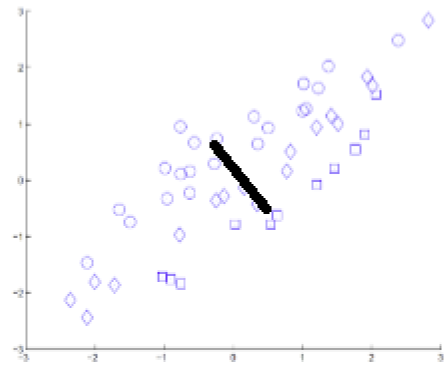
Part 2 is in the code.

(3.1)

(a)



(b)



(3.2)

$$\begin{bmatrix} 2 & 2 \\ 0 & 0 \\ -2 & -2 \end{bmatrix}, \mu_1 = 0, \mu_2 = 0, \sigma_1 = 2, \sigma_2 = 2$$

$$Standardized\ with \frac{x-\mu}{\sigma}: \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ -1 & -1 \end{bmatrix}$$

$$Covariance\ matrix: \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$Var(F_1) = \frac{(1-0)^2 + (0-0)^2 + (-1-0)^2}{2} = 1$$

$$Var(F_2) = \frac{(1-0)^2 + (0-0)^2 + (-1-0)^2}{2} = 1$$

$$Cov(F_1, F_2) = \frac{(1-0)*(1-0) + (0-0)*(0-0) + (-1-0)*(-1-0)}{2} = 1$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \lambda I = \begin{bmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix}$$

$$\det\left(\begin{bmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix}\right) = (1-\lambda)^2 - 1 = \lambda^2 - 2\lambda = \lambda(\lambda - 2)$$

$$Eigenvalues = 2{,}0$$

$$Using\ only\ non-zero\ eigenvalues:$$

$$\begin{bmatrix} 1-2 & 1 \\ 1 & 1-2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} => Eigenvectors = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$$

$$Multiply$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ -1 & -1 \end{bmatrix} * \frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}{||w||} = \frac{\begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}}{||w||} = \begin{bmatrix} \sqrt{2} \\ 0 \\ -\sqrt{2} \end{bmatrix}$$

On 1-D space, the coordinates are now (sqrt 2), (0), (-sqrt 2). The variance of this data is:

$$\frac{\left(\sqrt{2}-0\right)^2+(0-0)^2+\left(-\sqrt{2}-0\right)^2}{2}=\frac{2+2}{2}=2$$

The cumulative explained variance of this component is 100%, as the variance of the data is the same as the variance before the projection.

(4)

$$w=\sum_{i=1}^{N}\alpha_i y_i x_i = 0.414*1*\begin{bmatrix}4\\2.9\end{bmatrix}+0.018*-1*\begin{bmatrix}2.5\\1\end{bmatrix}+0.018*1*\begin{bmatrix}3.5\\4\end{bmatrix}+0.414*-1*\begin{bmatrix}2\\2.1\end{bmatrix}$$

$$=\begin{bmatrix}4*0.414-2.5*0.018+3.5*0.018-2*0.414\\2.9*0.414-1*0.018+4*0.018-2.1*0.414\end{bmatrix}=\begin{bmatrix}0.828\\0.3852\end{bmatrix}$$

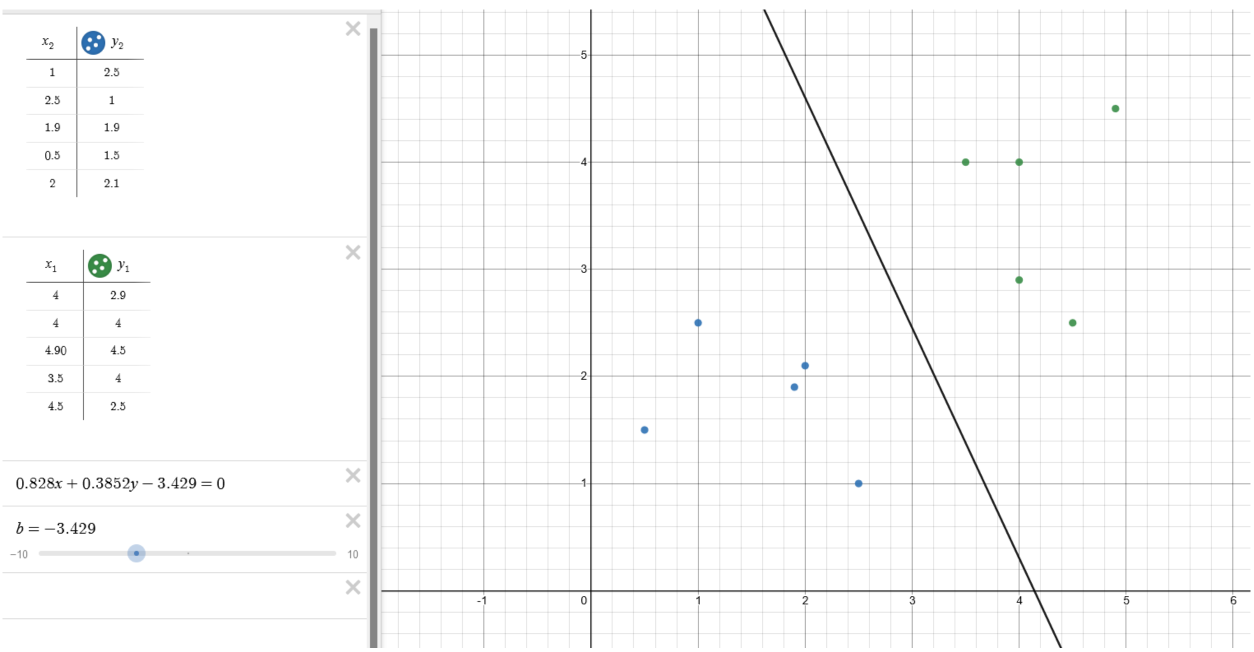
$$\alpha_i[y_i(w^T x_i+b)-1]=0$$

$$0.414\left[1\left(\left(\begin{bmatrix}0.828\\0.3852\end{bmatrix}\right)^T*\begin{bmatrix}4\\2.9\end{bmatrix}+b\right)-1\right]=0$$

$$(4*0.828+2.9*0.3852)+b-1=0$$

$$4.429+b-1=0$$

$$b=-3.429$$



$$Distance\ of\ point\ (x_0,x_1)\ to\ plane\ Ax+By+C=0$$

$$\frac{|Ax_0+Bx_1+C|}{\sqrt{A^2+B^2}}$$

$$For\ x_6: d=\frac{|0.828*1.9+0.3852*1.9-3.429|}{\sqrt{0.828^2+0.3852^2}}=\frac{1.124}{0.913}=1.23$$

$$Margin\ of\ classifier=\frac{1}{||w||}=\frac{1}{\sqrt{0.828^2+0.3852^2}}=\frac{1}{0.913}=1.10$$

X6 is not within the margin of the classifier.

Using the classifier above, point (3,3) lies above the line on the side of y=1, therefore, it is classified as y=1.