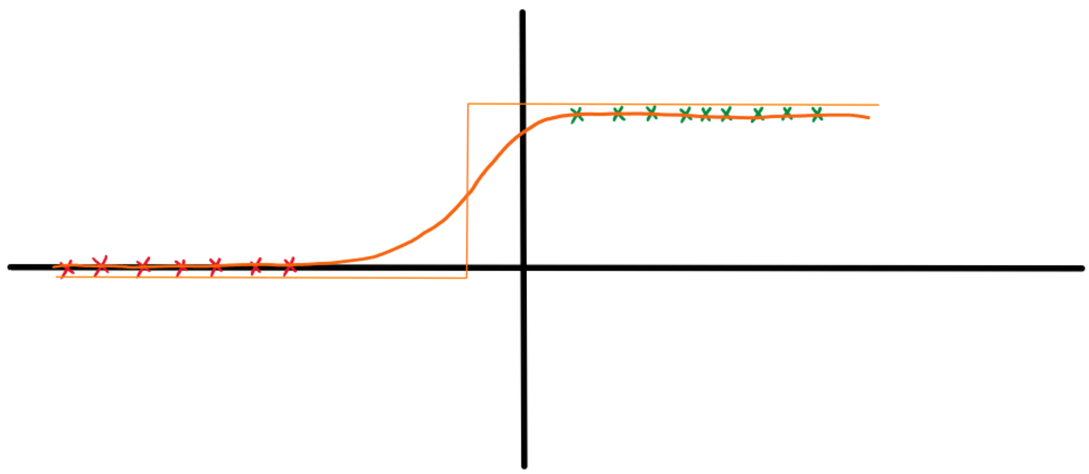
ECE 50024

Homework 3

Parth Sagar Hasabnis phasabni@purdue.edu

Exercise 1:



i. If the two classes of data are linearly separable, then with increasing iterations, the sigmoid function would tend to a step curve.
As $\theta \to \infty$, the angle of w would tend to $\pi/2$, and hence of slope of $||w||_2$ would tend closer to $\tan\left(\frac{\pi}{2}\right) = \infty$. Consequently, the transition of the sigmoid hyperplane would tend closer and closer to be parallel to the y-axis, resulting in the y-intercept w_0 to tend to ∞

ii. If we restrict $||w||_2 \leq c_1$ and $w_0 < c2$, then the iterations would stop after some finite number, causing the algorithm to converge. We can also counter non-convergence by:
   a. Stopping the algorithm by a finite number of iterations
   b. Applying a threshold on the maximum permissible error
   c. Adding a regularization term to penalize θ

iii. No, conversely, linear separability promotes convergence for linear-classifiers and convex problems, as in such cases, an analytical solution exists that can derive the optimal θ in a singular step.

Exercise 2:

Logistic regression loss is given by:

$$J(\theta) = \sum_{n=1}^{N} -\{y_n \log h_{\bar\theta}(\bar x_n) + (1-y_n)\log(1-h_{\bar\theta}(\bar x_n))$$

$$h_{\bar\theta}(\bar x) = \frac{1}{1+e^{-(\bar\omega^T\bar x+\omega_0)}} \qquad \bar\theta = \begin{bmatrix}\bar\omega \\ \omega_0\end{bmatrix}$$

$$= \sum_{n=1}^{N} -\left\{ y_n \underbrace{\log\frac{h_{\bar\theta}(\bar x_n)}{1-h_{\bar\theta}(\bar x_n)}} + \log(1-h_{\bar\theta}(\bar x_n))\right\}$$

$$= \log\frac{1/(1+\exp(-\theta^T\bar x_n))}{1-1/(1+\exp(-\theta^T\bar x_n))} = \log\frac{1}{1+\exp(-\theta^T\bar x_n)-1} = \bar\theta^T\bar x_n$$

$$= \sum_{n=1}^{N} -\left\{ y_n\bar\theta^T\bar x_n + \log(1-h_{\bar\theta}(x_n))\right\}$$

$$= -\sum_{n=1}^{N}(y_n\bar x_n)^T\theta - \sum_{n=1}^{N}\log\underbrace{\left(1-\frac{1}{1+\exp(-\theta^T\bar x_n)}\right)}$$

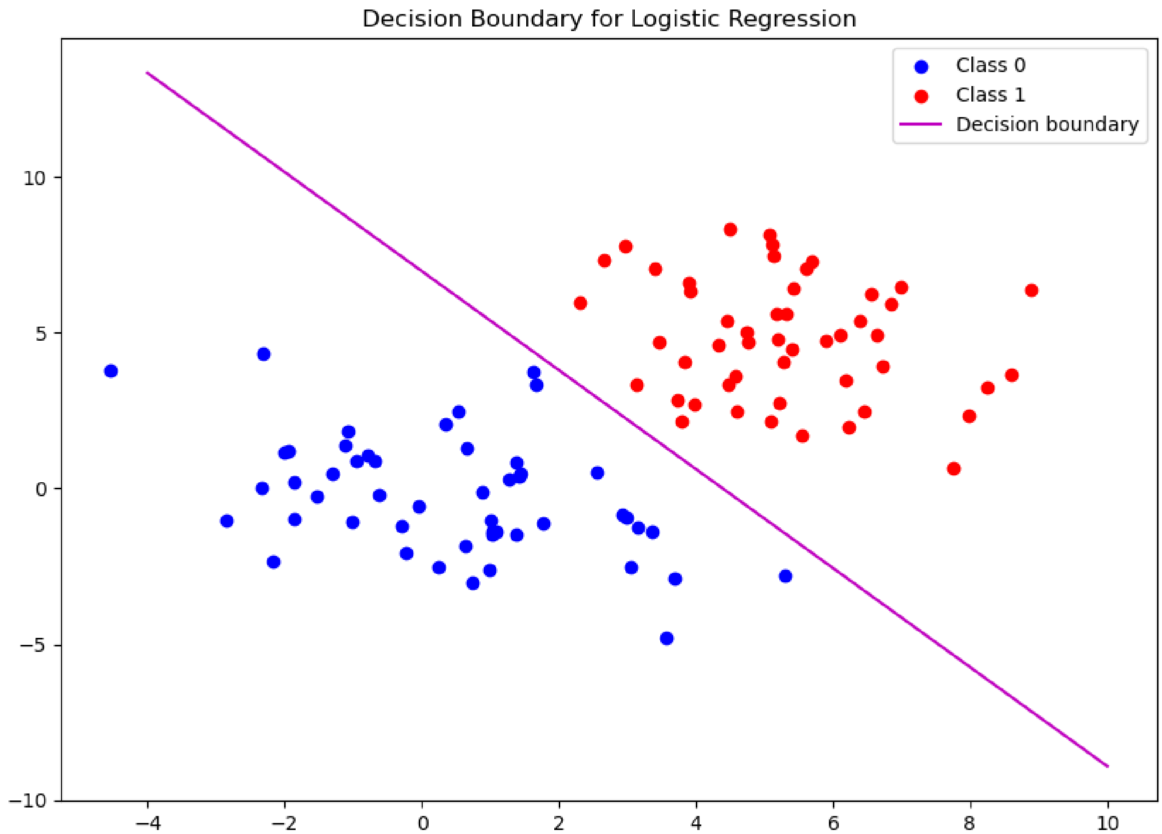
$$= \log\frac{1}{1+\exp(\bar\theta^T\bar x_n)}$$

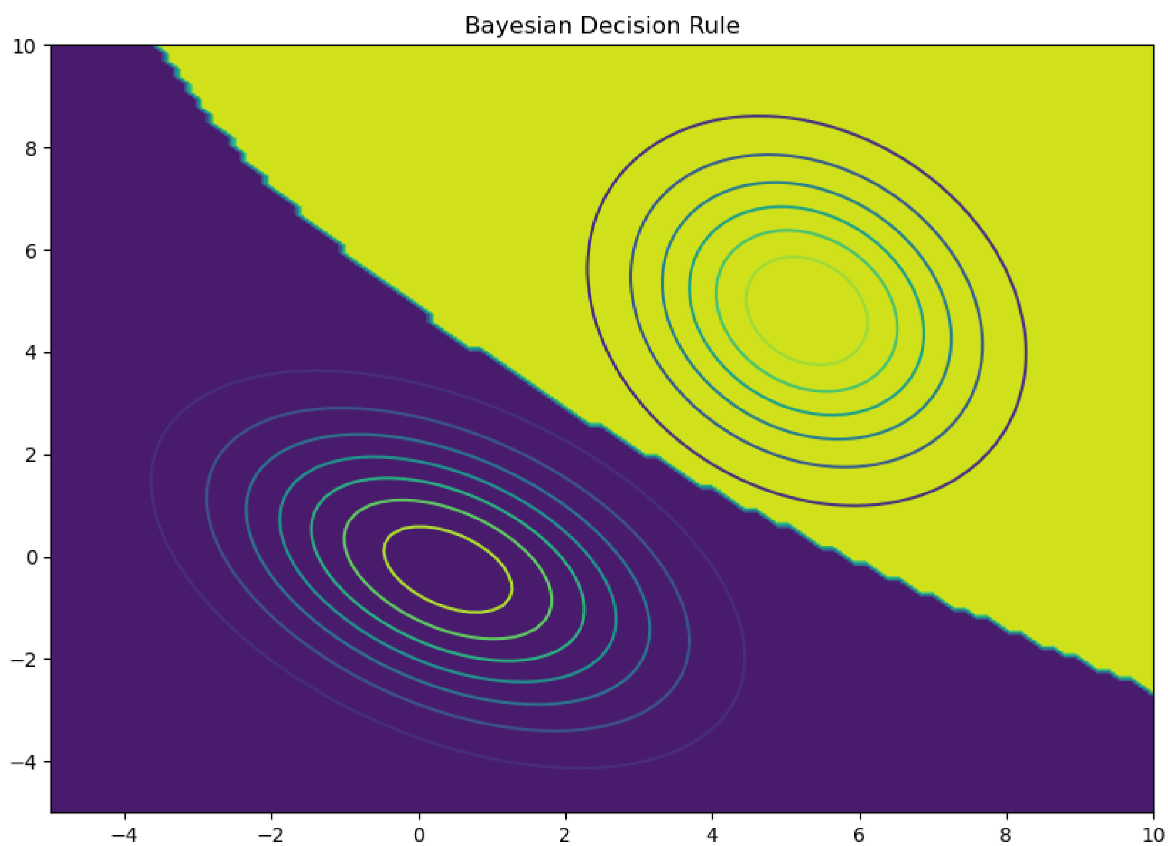
$$= -\log(1+\exp(\bar\theta^T\bar x_n))$$

$$= -\left\{\sum_{n=1}^{N}(y_n\bar x_n)^T\theta - \sum_{n=1}^{N}\log(1+\exp(\bar\theta^T\bar x_n))\right\}$$

Theta is given by:

$$\theta = \begin{matrix}[-10.44]\\ [2.38] \\ [1.50]\end{matrix}$$


Decision Boundary for Logistic Regression
Class 0
Class 1
Decision boundary
10
5
0
−5
−10
−4
−2
0
2
4
6
8
10


Bayesian Decision Rule
10
8
6
4
2
0
−2
−4
−4
−2
0
2
4
6
8
10

Exercise 3:

```python
K[47:52, 47:52]
```

```
array([[1.00000000e+00, 5.05310080e-25, 6.06536602e-20, 4.65474122e-29,
        4.06890793e-17],
       [5.05310080e-25, 1.00000000e+00, 3.95931666e-13, 2.69357110e-33,
        5.38775392e-12],
       [6.06536602e-20, 3.95931666e-13, 1.00000000e+00, 2.30352619e-65,
        3.78419625e-34],
       [4.65474122e-29, 2.69357110e-33, 2.30352619e-65, 1.00000000e+00,
        2.16278503e-06],
       [4.06890793e-17, 5.38775392e-12, 3.78419625e-34, 2.16278503e-06,
        1.00000000e+00]])
```

Using kernel trick:

$$\theta^T x = \sum_{n=1}^{N} \alpha_n \langle x_n, x \rangle$$

$$= \sum_{n=1}^{N} \alpha_n \cdot K(x_n, x)$$

$$J(\theta) = \frac{-1}{N} \left\{ \left(\sum_{n=1}^{N} y_n x_n \right)^T \theta - \sum_{n=1}^{N} \log\left(1 + \exp(\theta^T x_n)\right) \right\} + \lambda \|\theta\|^2$$

$$= \sum_{n=1}^{N} y_n \theta^T x_n$$

$$= \sum_{n=1}^{N} y_n \sum_{n=1}^{N} \alpha_n \cdot K(x_n, x_n)$$

$$= \sum_{n=1}^{N} y_n K\alpha$$

$$= y^T K \alpha$$

$$= \sum_{n=1}^{N} \log\left(e^0 + e^{\theta^T x_n}\right)$$

$$= \sum_{n=1}^{N} \log\left(e^0 + e^{\theta^T x_n}\right)$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \log \begin{bmatrix} e^0 + e^{\theta^T x_1} \\ e^0 + e^{\theta^T x_2} \\ \vdots \\ e^0 + e^{\theta^T x_n} \end{bmatrix}$$

$$= 1^T \log\left(e^0 + e^{K\alpha}\right)$$

$$= \lambda \theta^T \theta$$

$$= \lambda \sum_{n=1}^{N} \alpha_n \cdot K(x_n, \theta)$$

$$= \lambda \sum_{n=1}^{N} \alpha_n \cdot K\left(x_n, \sum_{n=1}^{N} \alpha_n x_n\right)$$

$$= \lambda \sum_{n=1}^{N} \alpha_n \cdot K(x_n, x_n)\, \alpha$$

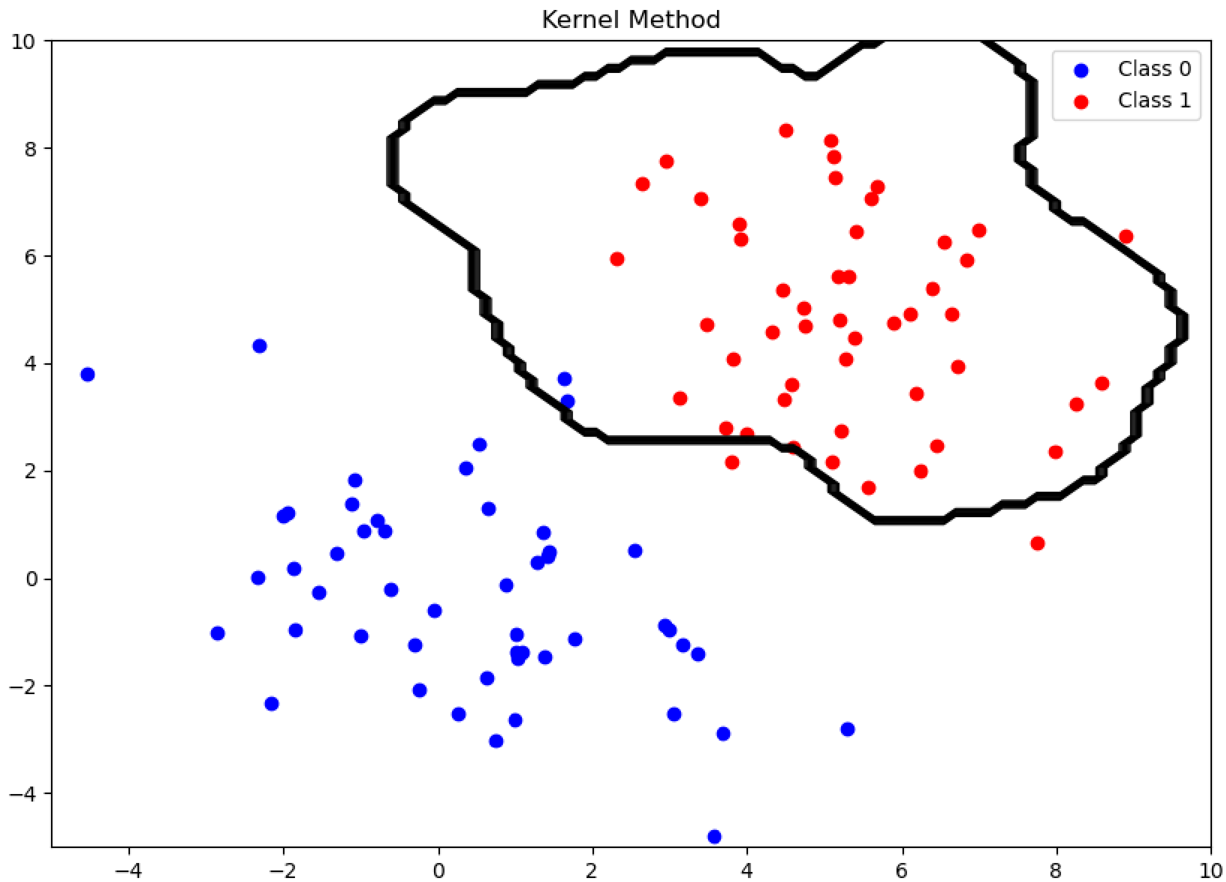
$$= \lambda \alpha^T K \alpha$$

J is expressed as a func^n of α

$$\therefore J(\alpha) = \frac{-1}{N} \left\{ y^T K \alpha - 1^T \log\left(e^0 + e^{K\alpha}\right) \right\} + \lambda \alpha^T K \alpha$$

Alpha is given by:

$$\alpha[0:2] = \begin{matrix} -0.95 \\ -1.21 \end{matrix}$$


Kernel Method
Class 0
Class 1
10
8
6
4
2
0
−2
−4
−4
−2
0
2
4
6
8
10

APPENDIX