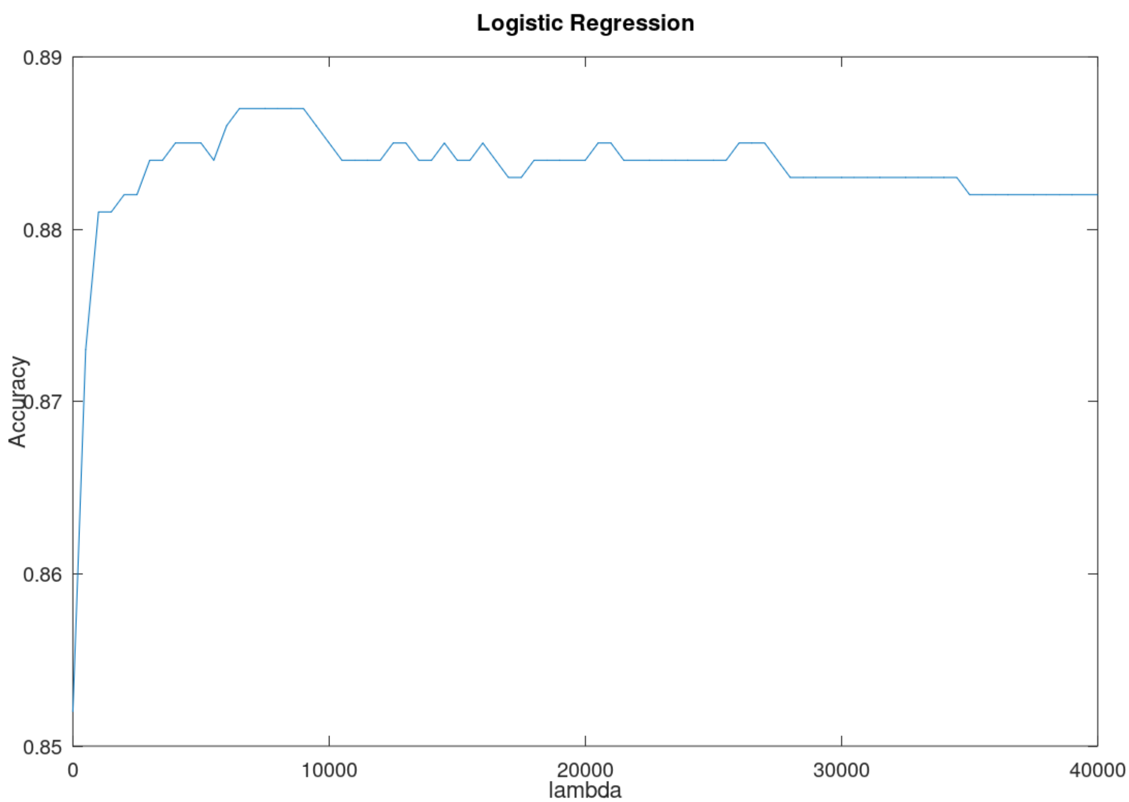
Assignment 2 Solutions

CS480/680 – Spring 2019

1. Question 1

   - Graph of 10-fold cross validation accuracy for logistic regression. The best λ is between 6500 and 9000 with a validation accuracy of 88.7%.



   - The test accuracy is 89.1% for mixture of Gaussians and also 89.1% for logistic regression (when $\lambda = 6500$).

   Discussion:

   - The Mixture of Gaussians model has more parameters (square in the dimensionality of the input space) than logistic regression (linear in the dimensionality of the input space). This is due to the fact that the mixture of Gaussians encodes a joint distribution over the inputs and outputs, where as logistic regression only encodes a conditional distribution over the outputs given the inputs. The maximum likelihood objective leads to a closed form solution for the mixture of Gaussians and therefore computation is quite fast. In contrast, logistic regression does not have a closed form solution and therefore an iterative technique must be used, which requires more computation. Logistic regression does not assume that the class conditional distributions are Gaussians and therefore might fit better the data, but regularization is needed to prevent overfitting. Mixture of Gaussians assumes that the class conditionals are Gaussians, but when this assumption is reasonable, less data is needed to obtain good results and there is a smaller risk of overfitting.

- Since the separator for KNN is non-linear, it has greater flexibility to fit the data in comparison to the linear separators for mixture of Gaussians and logistic regression. As a result, the non-linear separator tends to achieve greater training accuracy (for a small number of neighbours k) than the linear separators, but the linear separators overfit less and achieve higher test accuracy. While the data is not linearly separable, the linear separators perform better overall for this dataset.

2. Linear separability

   (a) Logic gates

   - AND: linearly separable, $\mathbf{w} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $w_0 = -1.5$
   - OR: linearly separable, $\mathbf{w} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $w_0 = -0.5$
   - XOR: not linearly separable. Use $\phi(\mathbf{x}) = \begin{pmatrix} x_1 \\ x_2 \\ x_1 x_2 \end{pmatrix}$ with weights $\mathbf{w} = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$, $w_0 = -0.5$.
   - IFF: not linearly separable. Use $\phi(\mathbf{x}) = \begin{pmatrix} x_1 \\ x_2 \\ x_1 x_2 \end{pmatrix}$ with weights $\mathbf{w} = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$, $w_0 = 0.5$.

   (b) No. If the data was linearly separable, the logistic regression technique should have found a separator with 100% training accuracy, but this was not the case. Since the objective is convex, it is possible to guarantee that the global optimum will be found. Hence, if the global optimum does not separate the data linearly, then the data is not linearly separable. Note that this argument assumes that we find the global optimum exactly, but in practice, iterative methods approximate the global optimum arbitrarily closely.