Theodore Kim
CS-UY 4563 – Introduction to Machine Learning
HW #5

LAB REPORT:

data1.csv

- **Logistic Regression (L1 Regularization):** resulted in a maximum testing accuracy of 0.517 using the c-value 0.001
- **Logistic Regression (L2 Regularization):** resulted in a maximum testing accuracy of 0.509 using the c-value 0.0001
- **Logistic Regression (L1 Regularization, Polynomial Transformation):** resulted in a maximum testing accuracy of 0.508 using the c-value 0.0001
- **Logistic Regression (L2 Regularization, Polynomial Transformation):** resulted in a maximum testing accuracy of 0.501 using the c-value 1000. Interestingly, using this method resulted in two local maxima and a divergence between the c-value which attained the global maximum in the training (c=10000) and testing dataset.
- **Support Vector Machine (Linear Kernal)**: resulted in a maximum testing accuracy of 0.543 using a c-value less than 10^{-6}.
- **Support Vector Machine (Radial Basis Function)**: required 2 parameters: a c-value and g-value. Increasing the value of gamma shifts the minimum accuracy attained by the model to the "left."
  - The pattern exhibited by the function has the following trend:
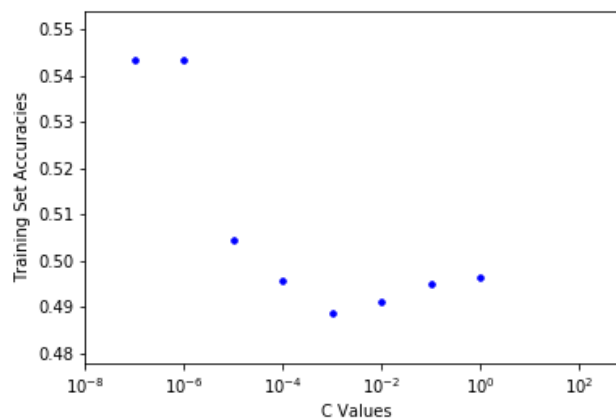
  Where at lower values of C, the accuracy of the testing set applied to the SVM model is constant then suddenly decreases and attains a global minimum before increasing slightly afterwards. Therefore, at lower values of gamma, the c-value of the global minimum also decreases. Overall, the maximum test accuracy attained by the model is 0.5172.

- **Support Vector Machine (Polynomial Kernal):** The training data exhibits an "S-curve" exhibiting a minimum accuracy of 0.5 at less than 10e-1 and an accuracy of 1 at c-values greater than 1. The testing data demonstrated an almost "Reflected" trend in which values less than 1 had a probability <u>0.533</u>
- **Overall the model and c-value that resulted in the best accuracy for the testing data was the linear kernel applied to the support vector machine, which achieves an maximum accuracy of 0.533.**

cryotherapy

- **Logistic Regression (L1 Regularization):** resulted in a maximum testing accuracy of <u>0.507</u> using the c-value <u>0.1</u>
  - The training and testing values' c-values for their accuracy maxima are divergent, the maximum accuracy / c-value of the test set was 0.501 and 100 respectively
- **Logistic Regression (L2 Regularization):** resulted in a maximum testing accuracy of <u>0.504</u> using the c-value <u>0.1</u>
  - The training and testing values' c-values for their accuracy maxima are divergent, the maximum accuracy / c-value of the test set was 0.501 and 1000 respectively
- **Logistic Regression (L1 Regularization, Polynomial Transformation):** resulted in a maximum testing accuracy of <u>0.509</u> using the c-value <u>1</u>
- **Logistic Regression (L2 Regularization, Polynomial Transformation):** resulted in a maximum testing accuracy of <u>0.507</u> using the c-value <u>1</u>.
- **Support Vector Machine (Linear Kernal):** resulted in a maximum testing accuracy of <u>0.533</u> using a c-value less than <u>10^{-7}</u>.
- **Support Vector Machine (Polynomial Kernal):** The training data exhibits an "S-curve" exhibiting a minimum accuracy of 0.5 at less than 10e-1 and an accuracy of 1 at c-values greater than 1. The testing data demonstrated an almost "Reflected" trend in which values less than 1 had a probability <u>0.533</u>
- **Both the SVM utilizing a linear kernel and the SVM utilizing the polynomial kernel achieved the maximum accuracy of 0.533 out of all the models. Both models exhibited an "S-like" trend in which a constant maximum was achieved for lower values of C, which then quickly decreased. For the Linear Kernal SVM, the accuracy value changed at around 10e-4 whereas the polynomial kernel transitioned at around 10e0**

Immunotherapy

- **Logistic Regression (L1 Regularization):** resulted in a maximum testing accuracy of <u>0.777</u> using the c-value <u>0.001</u>
- **Logistic Regression (L2 Regularization):** resulted in a maximum testing accuracy of <u>0.777</u> using the c-value <u>0.0001</u>
  - Unlike the L1 regularization, lower values of "C" resulted in higher accuracies in the L2 regularized model, whereas the inverse was true for L1 regularization

- The model also had a plateaued maximum and minimum in the range (-inf, 10e-1] and [10e2, inf)
- **Logistic Regression (L1 Regularization, Polynomial Transformation):** resulted in a maximum testing accuracy of <u>0.777</u> using the c-value <u>1e-5.</u>
  - Like the untransformed L1 regularized model, there is a significant drop off of accuracy at values of c less than 10e-5
- **Logistic Regression (L2 Regularization, Polynomial Transformation):** resulted in a maximum testing accuracy of <u>0.777</u> using the c-value <u>1e-6</u>.
  - This model has multiple local maxima, possibly related to the higher degree of the transformed training data
- **Support Vector Machine (Linear Kernal)**: resulted in a maximum testing accuracy of <u>0.777</u> using a c-value less than <u>10^{-2}</u>.
- **Support Vector Machine (Polynomial Kernal):** The training data exhibits a constant accuracy score of 0.77 at all tested values of c.
- **All models resulted in an accuracy of 0.777 amongst the testing data set at relatively low values of "C"**