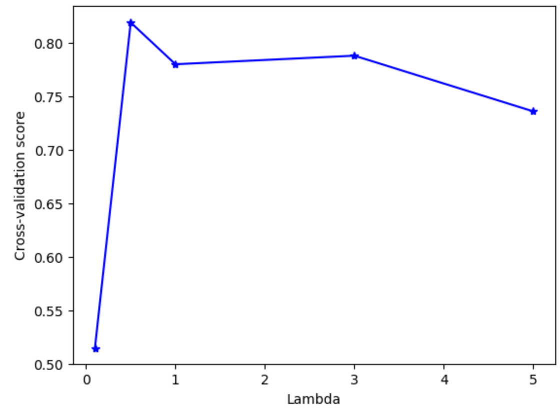
1

a) Cross-validation accuracy of logistic regression as λ varies



b) Report the accuracy of mixtures of Gaussians and logistic regression (with the best λ for regularization) on the test set.

Dataset	Gaussian Mixture	Logistic Regression
Test Data	0.89	0.87

c) Print the parameters π, μ1, μ2, Σ found for mixtures of Gaussian. Since the covariance Σ is quite big, print only the diagonal of Σ. Print also the parameters w, w0 found for logistic regression.

```
Pi_1 = 0.49818181818181817
Mean_1 = [ 4.78467153  4.88686131  8.35948905  9.93248175 10.71532847  9.39963504
  6.25364964  4.65693431  5.08576642  6.10583942 10.64051095  9.72992701
  8.56021898  7.82116788  5.3649635   4.40875912  4.52372263  6.43248175
 10.09124088  6.87408759  5.73905109  4.76824818  4.45255474  4.34124088
  4.80109489  6.46532847 10.20985401  9.59489051  7.97445255  6.34671533
  4.49087591  4.78649635  4.94525547  5.33029197  7.65145985  8.09306569
  8.53467153  7.79927007  5.71167883  4.26642336  4.60766423  4.83394161
  4.96350365  5.77737226  7.36131387  8.70620438  5.86131387  4.47080292
  4.33576642  5.0419708   7.09306569  7.86678832  8.78467153  7.93430657
  5.49635036  4.68248175  4.54927007  4.78649635  8.59854015 10.22080292
  8.72262774  6.28832117  5.08759124  4.97627737]
Mean_2 = [ 5.15036232  5.04891304  5.7826087   9.56521739  8.71195652  5.5615942
  5.49094203  4.89311594  4.81521739  4.83333333  8.24094203 10.38043478
  7.14311594  5.2807971   4.91123188  4.75362319  4.92753623  5.13043478
  9.78804348  8.51811594  4.99094203  5.17572464  4.86050725  4.93478261
  4.75724638  5.77355072 10.61594203  8.01086957  6.39855072  5.57971014
  5.17753623  4.69021739  5.06521739  6.27355072 10.69384058  9.91485507
  9.63768116  8.86413043  5.9384058   4.66485507  4.81521739  5.73550725
 10.5326087   9.19927536  7.78442029  9.42210145  8.42210145  4.74818841
  4.72101449  5.38768116  9.27898551 10.05253623  7.4365942   9.67572464
  9.0307971   5.00543478  4.93297101  5.07789855  5.875       9.2192029
 11.00724638  9.76268116  6.43478261  5.24275362]
```

```
Diagonals of Sigma = [39.60937846 40.20824861 43.80154824 42.61216334 40.50706774 42.44753911
 43.93051502 35.91092205 38.24783595 40.41761819 41.84800766 42.10213538
 43.74563842 40.40434607 39.8393938  37.00995089 36.63186003 41.31796473
 42.48053284 46.8853722  40.42681958 38.84473158 35.9435544  36.67287944
 36.89143623 40.84976046 40.35651693 46.39525856 46.00908187 43.9204288
 37.26551415 37.5592467  38.35337852 42.17204192 42.05802435 44.35816831
 45.13833476 45.11360559 42.33000478 33.18405631 37.49890553 40.07399842
 40.52155658 44.33597581 46.49162106 43.43930986 44.12577022 35.39756235
 35.77162814 40.73868096 44.04125306 44.4924987  46.18795151 44.89490291
 43.62974828 39.22378732 37.88905    37.8849315  45.32246523 42.60542762
 42.78673506 43.96937056 43.02679791 40.20232685]
```

W +wo

```
Weigts: [ 1.16486158e-01  1.51604771e+01 -3.11052318e+01 -1.96585981e+01
 -1.67518211e+00  1.42723396e+01 -1.27490050e+02 -3.81541477e+01
 -3.52147098e+01 -5.78306623e+01 -1.41269399e+01 -3.88909299e+01
  2.16404658e+00 -3.71288869e+01 -1.23714357e+02  1.44410700e+01
  2.36063222e+01 -2.64751750e+01 -8.01563006e+01  2.50630411e+01
  4.46911496e+01 -2.17617393e+01  1.65104288e+01  3.90146002e+01
 -2.68745473e+00 -2.36380836e+01 -1.88026887e+01 -6.06163555e+00
  4.65486484e+00 -7.19610384e+01  5.13410219e+01  7.34741559e+01
 -7.18200308e+01 -6.67744772e+01  3.57121681e+01  2.24497911e+01
 -6.47564001e-01  8.21133053e+00 -9.76179705e+00  2.72001242e+01
  1.83913821e+01  1.11396231e+00  7.39436057e+00  1.85868860e+02
  9.00492906e+01  1.13748044e+01 -3.69974267e-01  2.56325113e+00
  5.00739980e+01 -2.07991750e+00  6.75070428e+01  5.05512928e+01
  2.45610609e+01 -3.77578406e+01 -5.33189799e+00  3.54635597e+01
  2.24821585e+01  1.99249252e+01  9.10698440e+01 -7.18597291e+01
 -6.11443764e+01 -1.33387598e+01  4.28813558e+01 -1.28127622e+01
  1.02488128e+00]
```

d) Briefly discuss the results:

- In Gaussian Mixture model, we have π, µ1, µ2, Σ, the number of parameters of which is basically 1+2N + N^2 (N is number of features). In Logistic Regression, we have only **W** and wo, the number of parameters of which is basically N+1. The accuracies of both models on test dataset are comparable.
- In case of Mixture of Gaussians and logistic regression, the separators are line (2D), plane (3D) and hyperplane (n-dim) while KNN seeks to find curve, or surfaces to separate the data. Mixture of Gaussians and logistic regression are applicable when data are linearly separable while KNN can work well non-linear data.
- Then KNN algorithm in Assignment 1 only accomplished an 81% accuracy. Thus, we can conclude that a linear separator was more suited for this data because both methods that found a linear separator achieved a higher accuracy. We could say that the data was more linearly separable that “K Nearest Neighbour separable”.

2

a)

Let 's assume a case of two inputs (x1, x2) and one output (y) for the sake of simplicity. So we have w0, w1, and w2

- And : we can encode And function as a threshold perceptron with wo = 1.5 , w1 = 1, w2 = 1
- OR: we can encode OR function as a threshold perceptron with wo = 0.5 , w1 = 1, and w2 = 1
- Exclusive-OR (XOR): we cannot encode XOR function as a threshold perceptron since it is not linearly separable. However, we can use feature mappings to make data linearly separable by creating an extra variable which is product of x1 and x2 (x1*x2).
- Iff (or XNOR). we cannot encode XOR function as a threshold perceptron since it is not linearly separable. XNOR is the opposite of XOR. So we can use feature mappings to make data linearly separable by creating an extra variable which is product of x1 and x2 (x1*x2).

b) To answer whether the train set is linearly separable. We use the weight vector from the model with 'best lambda' to identify the data which falls above or below the hyper-plane given by $\mathbf{w}^T\mathbf{x}$. if $\mathbf{w}^T\mathbf{x} > 0$, we assign 1 and otherwise 0 to a result vector. Please note that $\mathbf{w_0}$ is augmented inside **w**. Then we can compare the result vector with train labels by computing confusion matrix. If train data is linearly separable, there will not be any misclassification.

Here is the resulting confusion matrix:

```
[[468  80]
 [ 64 488]]
```

There are misclassifications, so we can conclude that the train data is not separable.