

Q1A) MIXTURE OF GAUSSIANS RESULTS**Accuracy:** 97 / 110**pi:** 0.5

mean1: [4.846 4.902 8.28 9.858 10.68 9.342 6.198 4.602 4.978 5.998 10.498 9.662 8.448 7.66
5.322 4.412 4.438 6.36 10.02 6.84 5.6 4.764 4.436 4.366 4.758 6.392 10.126 9.514 7.924 6.302
4.532 4.64 4.812 5.274 7.596 7.948 8.518 7.668 5.634 4.226 4.612 4.718 4.924 5.762 7.362
8.612 5.764 4.408 4.29 5.038 7.032 7.762 8.782 7.908 5.516 4.548 4.456 4.758 8.606 10.226
8.702 6.198 5.002 4.944]

mean2: [5.158 4.954 5.654 9.55 8.618 5.482 5.478 4.896 4.782 4.75 8.194 10.272 7.03 5.272
4.864 4.736 4.864 5.058 9.814 8.438 4.92 5.068 4.762 4.882 4.686 5.76 10.572 8.02 6.306 5.47
5.154 4.664 4.988 6.278 10.67 9.814 9.532 8.804 5.84 4.588 4.832 5.732 10.402 9.132 7.756
9.324 8.338 4.732 4.704 5.334 9.224 9.968 7.304 9.668 8.934 4.968 4.932 5.002 5.84 9.20 10.99
9.72 6.41 5.186]

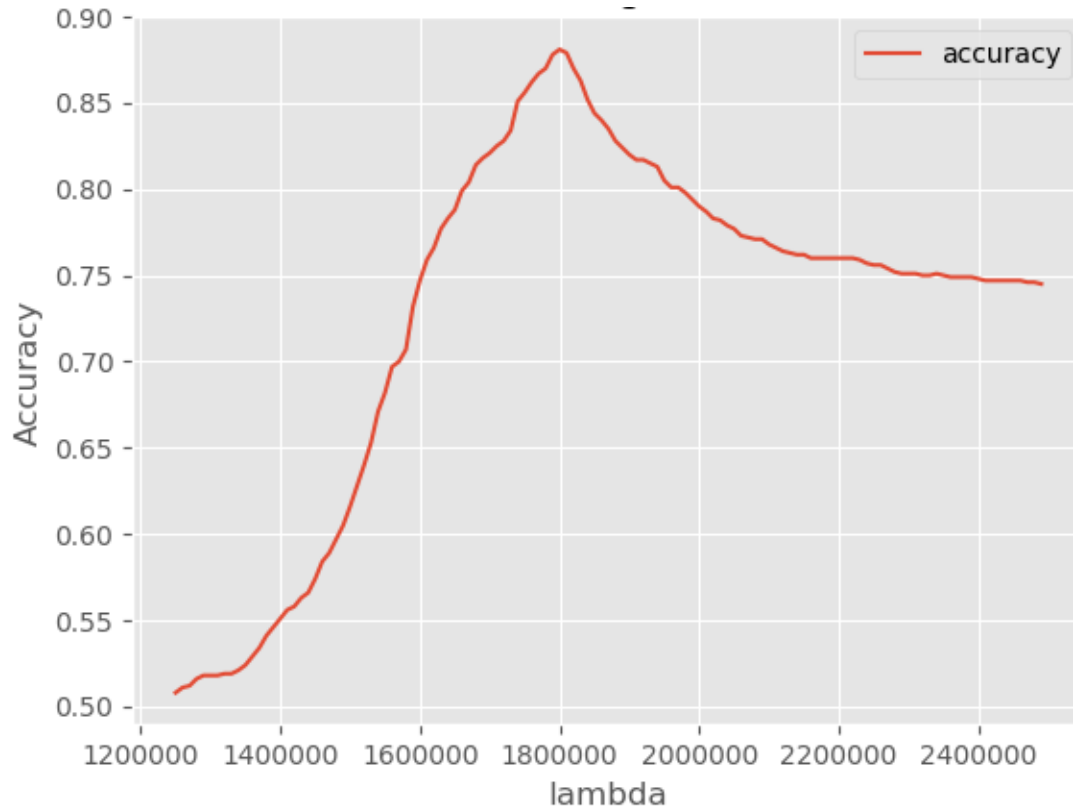
covariance matrix diagonal: [40.06966 39.82214 43.393942 42.766668 40.430838 42.431356
43.962156 35.76039 37.639996 39.744748 41.88718 42.250886 43.980198 40.399208 39.59391
37.22228 35.94983 40.826518 42.785502 46.726278 39.7868 38.79184 35.49763 36.49206 36.55942
40.446368 40.62147 46.540702 45.725294 43.487948 37.28763 36.988752 37.872256 42.13782
41.990942 44.39835 45.407326 45.28368 42.051222 32.90459 37.272616 39.713326 40.32731
44.527966 46.29971 43.68424 43.72803 35.242856 35.657142 40.4655 44.1844 44.524166
46.08703 44.612656 43.811694 38.729336 37.56372 37.560716 45.240582 42.59006 42.565548
43.998198 42.671948 40.080134]

w: [-0.01543618 0.01619483 0.03937567 -0.00457235 0.06195358 0.15364871 0.01919253
0.05518105 0.08687927 -0.01684631 0.12062344 0.00442281 0.02593515 0.15031986 -0.00108072 -
0.04183353 0.07208746 0.04531849 0.01064029 -0.02699645 -0.00676469 -0.05821995 -0.04185447
0.02763963 0.08186341 0.02268948 -0.02127659 0.01798843 0.09210444 -0.0083851 -0.07581809 -
0.00457054 0.08691497 -0.01244805 -0.06072045 -0.0338675 -0.02835648 -0.08842853 -0.04118733 -
0.03513544 -0.02525595 0.00682692 -0.21289952 -0.0724104 -0.00984092 0.03160575 0.0265646 -
0.0510064 0.02278539 -0.0958112 -0.07552381 -0.04697154 0.06153424 -0.03537481 -0.06782566 -
0.00184471 -0.04236931 -0.07346939 0.08723099 0.0529176 -0.00980751 -0.05980714 -0.00544649
0.00300349]

w₀: 0.09863431377589027

Q1B) LOGISTIC REGRESSION RESULTS

LOGISTIC REGRESSION CROSS VALIDATION ACCURACY OF LAMBDA



Best Lambda: 1805000

Accuracy: 95 / 110

Discussion:

Mixtures of gaussians requires 4 parameters to calculate \mathbf{w} and w_0 : π , μ_1 , μ_2 , and Σ . Each parameter can be calculated in $O(n)$ time. \mathbf{W} and w_0 can then be computed in $O(M^2)$ time, because $\Sigma^{-1}(\mu)$ requires $M \cdot M$ calculations.

Logistic regression requires 3 parameters to estimate \mathbf{w} and w_0 : H , R , and ∇L . H is calculated in $O(N^2M)$ time, R is calculated in $O(M^2)$ time, and ∇L takes $O(N^2)$ time. These calculations must then be repeated 10 times to arrive at \mathbf{w} and w_0 using Newton's method. Furthermore, the entire process must then be repeated hundred of times to determine the optimal lambda.

In practice, my solution for mixture of gaussians took 0.339 seconds to complete with an accuracy of 0.88, while my solution for logistic regression completed in 8.44 seconds with an accuracy of 0.863. Therefore, logistic regression requires more computation but yields less accurate results.

Mixtures of gaussians and logistic regression find a linear separator, where as KNN finds a non-linear separator. If the data is linearly separable as it is in Q1, linear separators are much simpler to implement and will perform best. However, if the data is not linearly separable, then a non-linear method is required. The classification accuracy for KNN in A1 was 0.76, while the accuracy of the linear separators in A2 are 0.86 and 0.88. It is possible that KNN is less accurate because some data points lie very close to each other while one different sides of the linear separation line (ie they belong to different classes). While mixtures of gaussians and logistic regression would classify these points correctly, KNN would not.

Q2A) PERCEPTRONS

Boolean Function	Threshold perceptron possible?	Values
And	YES	W_0 : -0.7 W_1 : 0.5 W_2 : 0.5
Or	YES	W_0 : -0.5 W_1 : 1 W_2 : 1
Exclusive-or	NO	$\Phi: (x_1 - x_2)^2$ W_0 : -0.5 W_1 : 1
iff	NO	$\Phi: (x_1 - x_2)^2$ W_0 : 0.5 W_1 : -1

Q2B) LINEAR SEPARABILITY EXPERIMENT

To determine if the training set is linearly separable using a logistic regression classifier, start by determining the best value for lambda just like in Q1b. Then, train the logistic regression model with the full set of training data. The model will arrive at a set of weights, \mathbf{w} , for the optimal linear separation between the classes. Finally, use the \mathbf{w} from the training step to classify the exact same set of data that the model was trained on. Typically, one should never train and test their model on the same data because the results would not be an accurate depiction of the model's performance; however, in this experiment we don't care about the true accuracy of the classifier. Logistic regression is guaranteed to find a linear separator if it exists; therefore, the training data is linearly separable if the predictions are 100% correct.

After running this experiment, the classification accuracy was only 89%, therefore the training set is not linearly separable.