Problem 1. (30 points) True of False. If you choose false, briefly explain.

(T) • SVMs can be used in an ensemble classifier. (True or False)

(F) • In bagging, we choose random samples from the input without replacement. (True or False)
Boosting method cannot choose random samples.

(F) • A deep neural network with a non-linear activation function in each layer is meaningless. (True or False)
Without non-linear activation functions no matter how many layers we have the network behaves like single-layer.

(F) • Linear discriminant functions always lead to non-convex decision regions. (True or False)
Decision regions for a linear machine are always convex.

(F) • To achieve better performance, we should look at the testing set when training a classifier. (True or False)
Testing set cannot be used in training phase might lead to overfitting, less effective.

(T) • When applying K-Means for clustering, the distance measure should always be differentiable.
(True or False)

(F) • Compared to SVM, Adaboost is more sensitive to and thus easier to fail on un-normalized data. (True or False)
Adaboost doesn't need to meet normalized data, it's better against unnormalized data.

(F) • A kernel matrix have to be designed to be invertible. (True or False)
Kernel matrix has to be symmetric, but sometimes cannot be invertible. E.g. 0

(F) • Using high-degree polynomial features help to reduce overfitting in a SVM classifier. (True or False)
Raising features tend to result into overfitting.

(F) • EM algorithm always finds the global optimal. (True or False)
Not necessarily global optimal, could be local optimal

(F) • Deep neural networks usually require fewer training data as compared to Adaboost. (True or False)
Deep neural networks requires more since adaboost is self-iterating

(T) • When training a random forest, the order of training the decision trees does not play a role.

(True or False)

(F) • When applying SVM, we have to explicitly specify a kernel function that maps a sample from a lower dimension to a higher dimension. (True or False)
Only need to perform when meeting non-separable case, not necessary in separable case

(F) • Deep neural networks rely on setting the weights of the networks mannually.
In fact it uses backpropagation for setting weights.

(F) • In Adaboost, the weak learner that yields a higher error rate results in a higher weight in the combined classifier.
The purpose is to discard higher rate classifier, so it gets lower weight

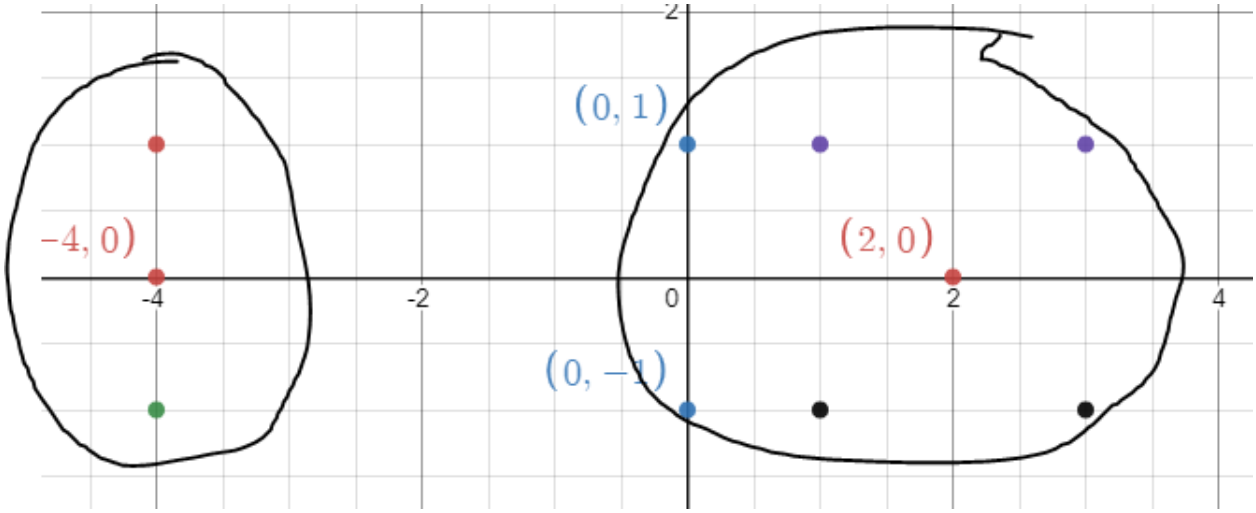
Problem 2.

(a) Increase when misclassified, decreased when correct-classified.

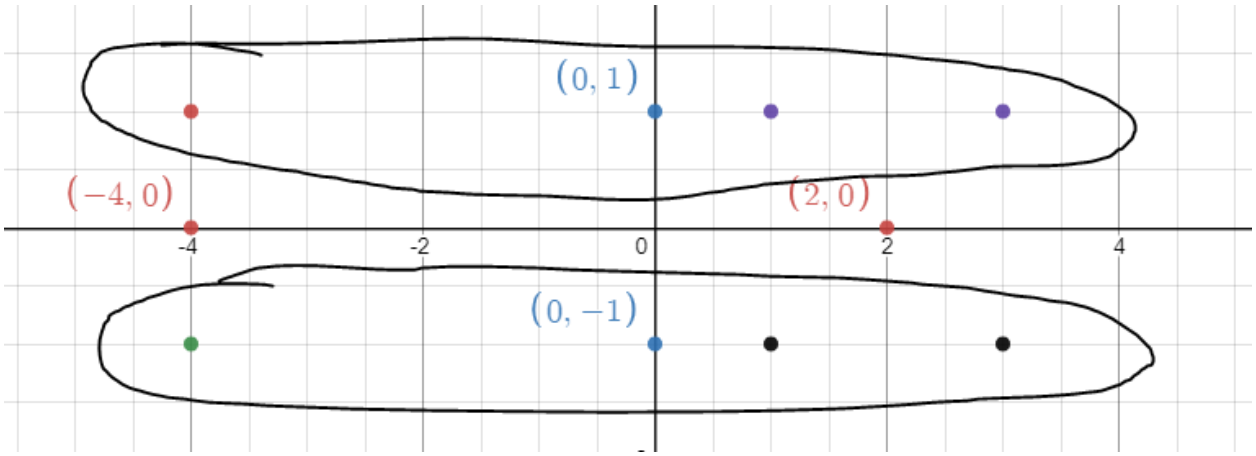
(b) The high error-rate classifier gets low weight.

(c)Second sample set {0:1538 0:2308 0:2308 0:1538 0:2308}
Since three out of five are misclassified according to three raised weight samples, the error rate of the classifier is above 0.5

Problem 3.



Sol1



Sol2

By calculating average distance between points and center in a cluster. Sol1 is preferable. In Sol1, Average distances are 2 and 1.69. In Sol2, Average distances are 5.33. Sol1 gets lower distances thus gets better result.

Problem 4.
Step2:
Computing the means is to tell the center points. If most of the points are close to the center then the cluster is dense, giving high (μi), thus minimizing the function.

Step3:
The distances between points and centers represent closeness. To minimize the function we need to minimize the distances. If we reassign points it gets closer to center it further minimizes the function.

Problem 5.
Step1: subtract mean from each row of the data.
Step2: compute the covariance matrix

Step3: Compute the eigenvectors and eigenvalues of the covariance matrix
Step4: Acquire k eigenvectors with highest eigenvalues
Step5: acquire product of raw data and k eigenvectors to obtain the reduced-dimension data.

Problem 6.

(a) Perceptron costs more since it considers all the points and SVM only considers support vectors. And also some points are very close to the hyperplane in perceptron; it might not be an ideal classification.

(b) MLE with Gaussian distribution is not accurate if the raw data does not follow Gaussian distribution. MLE is not giving correct result thus we need iterations to optimize result.