Name:

NetID:

1.) Given below is a Perceptron. Using the Step function between -1 and 1 as the activation function, classify the three points given below. Then update the Perceptron weights for each misclassified sample and record the new weights. (4 pts.)

ω = (1.2, 0.7, 0.2, 0.5), η = 0.2

Samples	X_1	X_2	X_3	Y
S_1	2	3	-1	1
S_2	1	1	1	1
S_3	-2	-4	3	-1

Classify: 1.2 1.4 0.6 -0.5

S_1: $1.2(1)+0.7(2)+0.2(3)+0.5(-1)=2.7 \rightarrow 1$ ✓

S_2: $1.2(1)+0.7(1)+0.2(1)+0.5(1)=2.6 \rightarrow 1$ ✓

S_3: $1.2(1)+0.7(-2)+0.2(-4)+0.5(3)=0.5 \rightarrow 1$ ✗

Weight Updates:

$w' = w + \eta \times y$

$w_0' = 1.2 + 0.2(1)(-1) = 1.0$

$w_1' = 0.7 + 0.2(-2)(-1) = 1.1$

$w_2' = 0.2 + 0.2(-4)(-1) = 1.0$

$w_3' = 0.5 + 0.2(3)(-1) = -0.1$

Final Weights:

ω = (ω_0: 1.0 ,ω_1: 1.1 ,ω_2: 1.0 ,ω_3: -0.1)

2.) What is the difference between a regression task and a classification task? (2 pts.)

Regression tasks yield continuous real numbers while classification gives discrete class labels

3.) What is the difference between a Perceptron and Maximal Margin Classifier? (1 pt.)

A perceptron finds a decision boundary, a MMC finds the boundary that maximizes the margin

4.) What are the support vectors? Why are these so important? (1 pt.)

Support vectors are the set of points that influence the margin of a margin based classifier. If they move the boundary moves

5.) What is the kernel trick? (1 pt.)

A method SVMs use to implicitly project the data (relationships) into higher dimensions

6.) Explain how an SVM could be used for multiclass classification (1 pt.)

We could train as many SVMs as there are classes in a one-vs-rest manner and then combine the decision boundaries

4.) Why can't we use gradient descent for the step function? (1 pt.)

It's not differentiable

5.) What is the gradient we're descending when we use gradient descent? What are we trying to optimize and what do we take the partial derivatives with respect to to do so? (2 pts.)

We're descending the weight manifold to try and minimize error.
We take the partial derivative with respect to every weight

6.) What are the differences between supervised and unsupervised learning? (1 pt.)

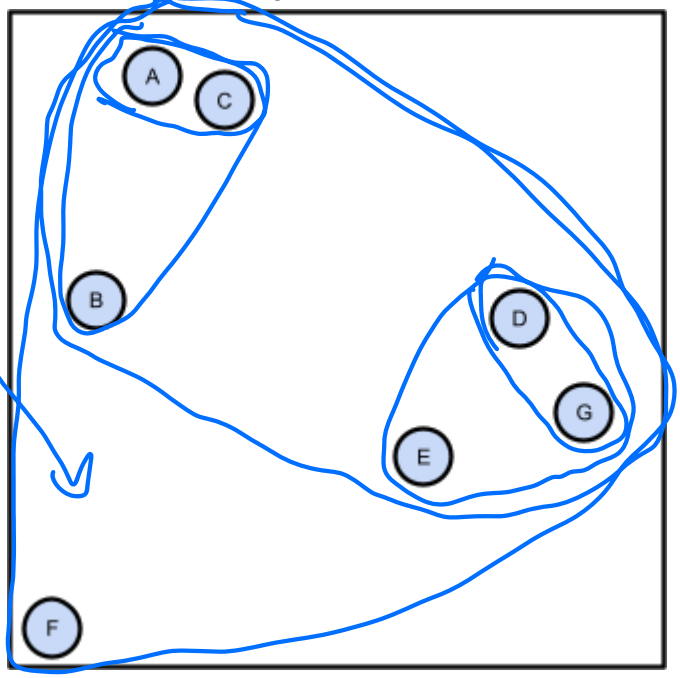
Supervised has labels/ground truth, Unsupervised does not

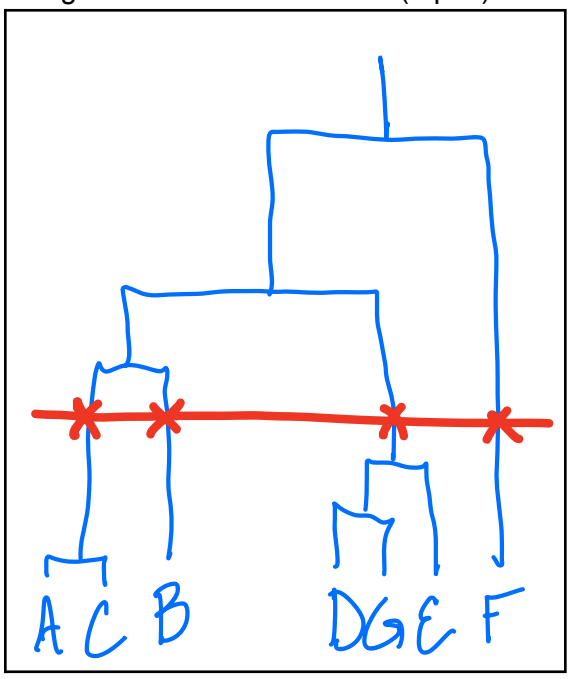
7.) What are centroids in k-means clustering? (1 pt.)

They are the "middle" of each cluster. May not actually be points in our dataset

kind of a judgment call sorry :\

8.) Given the data points, draw the dendrogram that would be created using agglomerative hierarchical clustering and then draw a line on the dendrogram to create 4 clusters. (3 pts.)

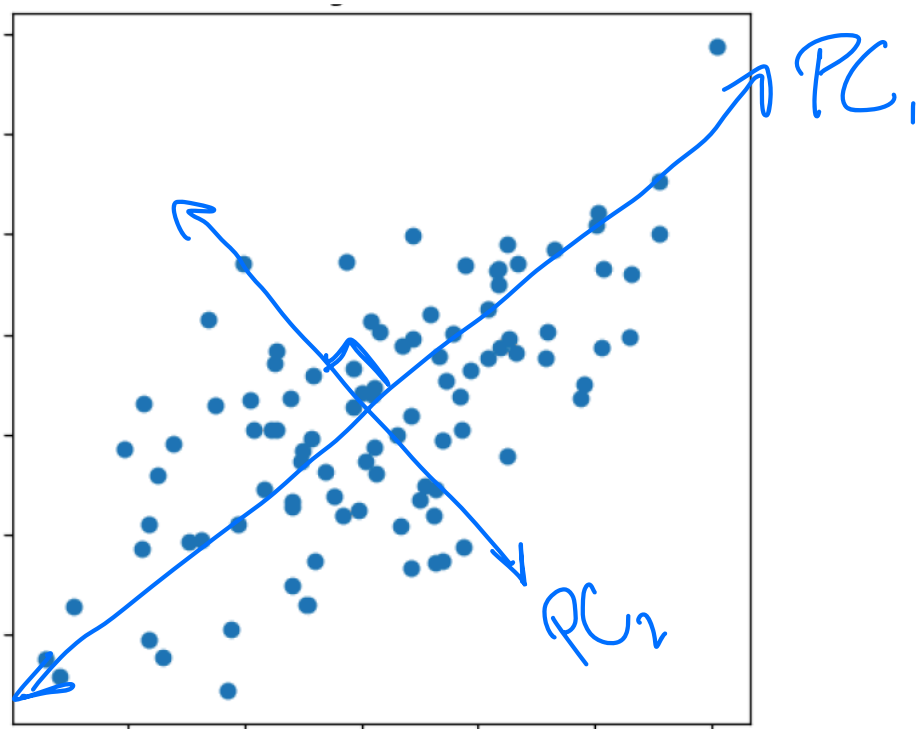




9.) In your own words, what is the curse of dimensionality? (1 pt.)

As we add dimensions our space grows more sparse. We need way more training data to "cover" the same amount of space

10.) Draw (approximately) the two principal components on the plot and label them (1 pt.)



11.) What do the values of the eigenvalues represent when we do PCA? (1 pt.)

The amount of variance captured by each principal component

12.) Given the following experiments, which of the metrics do you think would be most useful for measuring task performance. Select only one. (Multiple Choice) (4 pts.)

a.) An imbalanced multiclass classification task

- Precision
- Recall
- F1 Score
- MSE

b.) Deciding whether to give someone a loan

- Precision
- MSE
- Accuracy
- Silhouette Score

False positives are bad! give loan to someone who can't repay it

c.) A regression task

- Precision
- F1 Score
- MSE
- Laplacian Difference

d.) A clustering task

- MSE
- Precision
- Silhouette Score
- F1 Score

13.) What is grid search and what do we use it for? (1 pt.)

Systematically testing hyperparameters for the "best" set

14.) Write **psuedo-code** for setting up a multiclass classification task on the iris dataset using a Naive Bayes classifier and giving a classification report for a test set (3 pts.)

data = load('iris_dataset')

1.) Model = NB(gaussian)

2.) X_train, y_train, X_test, y_test = split(data)

3.) Model.fit(X_train, y_train)

4.) y_pred = Model.predict(X_test)

5.) report(y_test, y_pred)

15.) Write the following statement (1 pt.)

"I must always split my data into training and testing and must not train on the testing data"

I must always split my data into training and testing and must not train on the testing data

Bonus.) Which homework assignment (if any 😛) have you enjoyed or learned from? Why do you think that was? (1 bonus pt.)

None of them :(