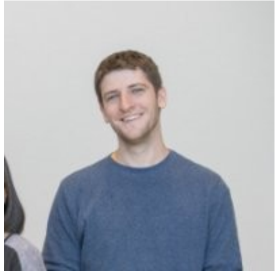
Final Project - Team Cortes

UCB Extension: COMPSCIX433.6-004

Corinna Cortes, Danish computer scientist and competitive runner

Group Members

Graham Gelwicks	Prathyusha Kanala	Kyle McGovern
Ajey Patil	Abdul Hadi Sediqi	Andy Tse



Dataset

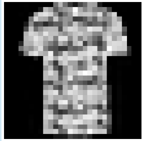
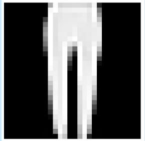
Fashion MNIST is a data set created by Zalando, a German fashion ecommerce company, as a direct drop-in replacement for the original MNIST dataset. Instead of hand-written numbers, the images are articles of clothing.

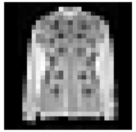
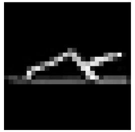
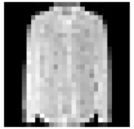
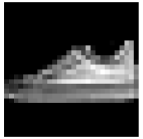
Feature vectors: 28*28 (784) grayscale (each feature is a # from 0-255)
Classes: 10 (0-9)
Training dataset: 60,000 feature vectors and labels
Testing dataset: 10,000 feature vectors and labels

The team wanted to focus on image classification, and considered a variety of other datasets including the University of Toronto’s CIFAR-10 (32*32 color images of animals). Fashion MNIST was chosen to focus more on trying different methods.

Dataset

Class	0	1	2	3	4	5	6	7	8	9
Name	T-shirt/top	Trouser	Pullover	Dress	Coat	Sandal	Shirt	Sneaker	Bag	Ankle boot
Tra										

*T-shirt, pullover, coat and shirt classes are particularly similar

Objective

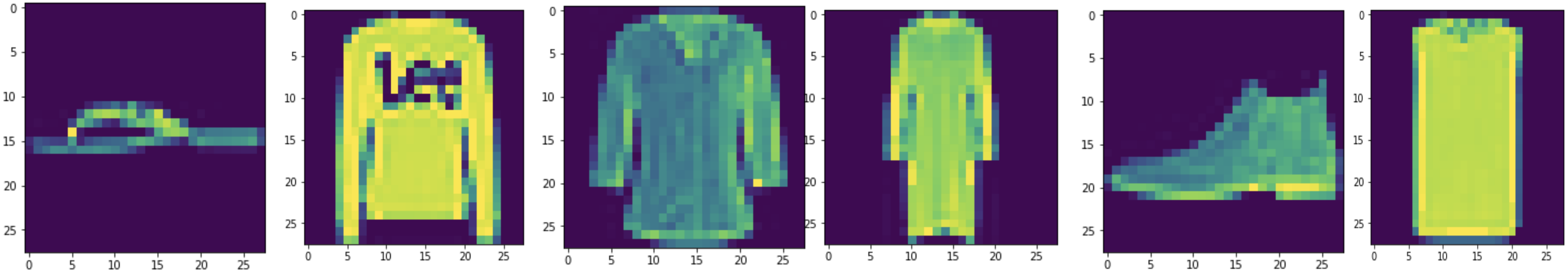
- Investigate and apply multiple ML techniques that can be used for image classification including:
    - The relatively simple: our beloved histogram using PCA
    - The complex: convolutional neural networks
    - And several in-between
- Compare accuracy across the different algorithms to understand which performs the best
- Expand our toolkit for future machine learning projects

Background of Naive Bayes Algorithm

- Algorithm is based off of Bayesian Theorem for having an assumption based off of independence and predictors.
- Simple classification problems in a particular feature class that is unrelated to another feature
- Often used for supervised learning
- Applications of its use:
  - Text classification
  - Spam filtering (Difference between spam and ham)
  - Determining word frequencies

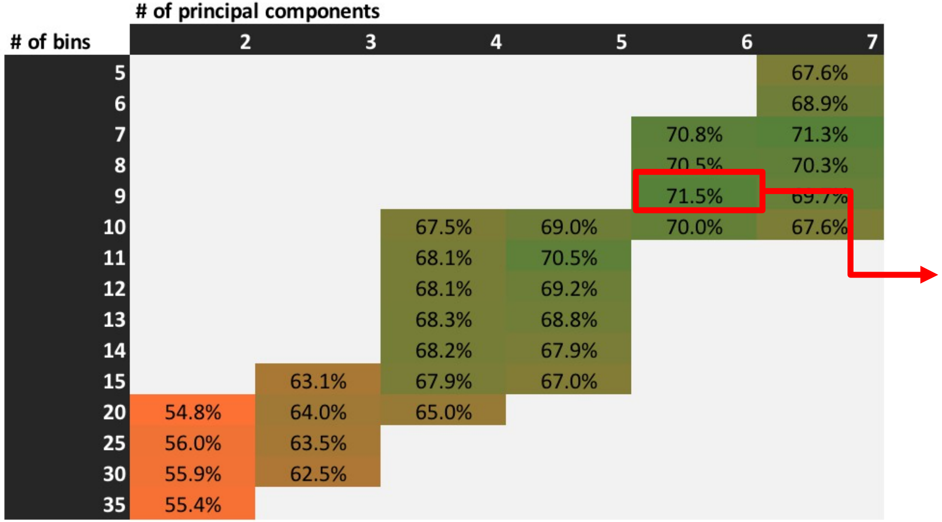
Naive Bayes Algorithm

- By using the Naive Bayes Algorithm, the accuracy result came up to 59%.
- Naive Bayes is the least effective method in comparison to many methods.
- These images below are misclassified as follows:
    - Predicted Class: Sneaker, Coat, Coat, Trouser, Sneaker, Dress
    - True Class: Sandal, Pullover, Shirt, Dress, Ankle Boot, T-Shirt



Histogram using Principal Components

Accuracy Table (%)



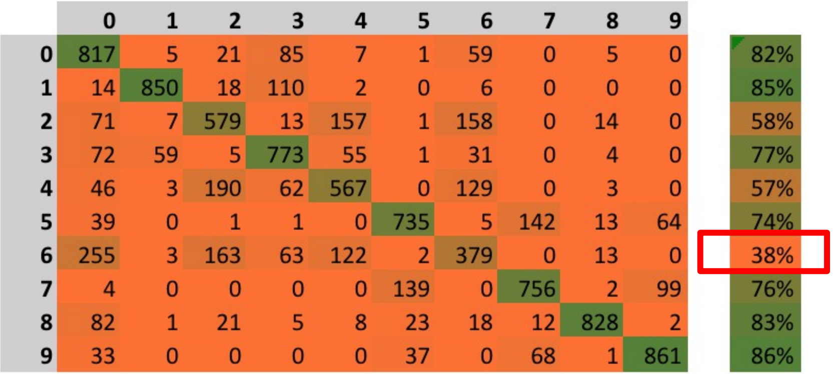
# of bins	2	3	4	5	6	7
5						67.6%
6						68.9%
7					70.8%	71.3%
8					70.5%	70.3%
9					71.5%	69.7%
10			67.5%	69.0%	70.0%	67.6%
11			68.1%	70.5%		
12			68.1%	69.2%		
13			68.3%	68.8%		
14			68.2%	67.9%		
15		63.1%	67.9%	67.0%		
20	54.8%	64.0%	65.0%			
25	56.0%	63.5%				
30	55.9%	62.5%				
35	55.4%					

# of principal components

- Just like homework 3 except it classified all 10 and used a varying number of Principal Components and bins
- Varied Principal Components from 2 - 7
- Varied # of bins from 5 - 35
- Best results: **6** principal components, **9** bins = **71.5%**

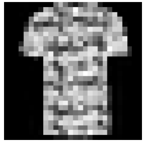
Confusion Matrix

Sensitivity



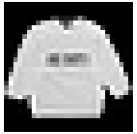
	0	1	2	3	4	5	6	7	8	9	Sensitivity
0	817	5	21	85	7	1	59	0	5	0	82%
1	14	850	18	110	2	0	6	0	0	0	85%
2	71	7	579	13	157	1	158	0	14	0	58%
3	72	59	5	773	55	1	31	0	4	0	77%
4	46	3	190	62	567	0	129	0	3	0	57%
5	39	0	1	1	0	735	5	142	13	64	74%
6	255	3	163	63	122	2	379	0	13	0	38%
7	4	0	0	0	0	139	0	756	2	99	76%
8	82	1	21	5	8	23	18	12	828	2	83%
9	33	0	0	0	0	37	0	68	1	861	86%

0



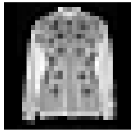
T-shirt

2



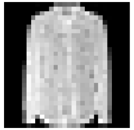
Pullover

4



Coat

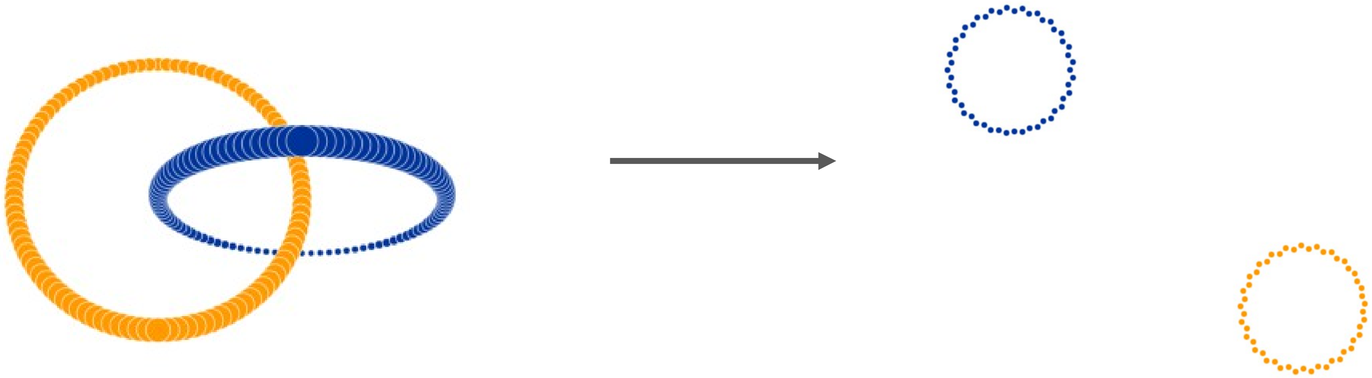
6



Shirt

t-SNE

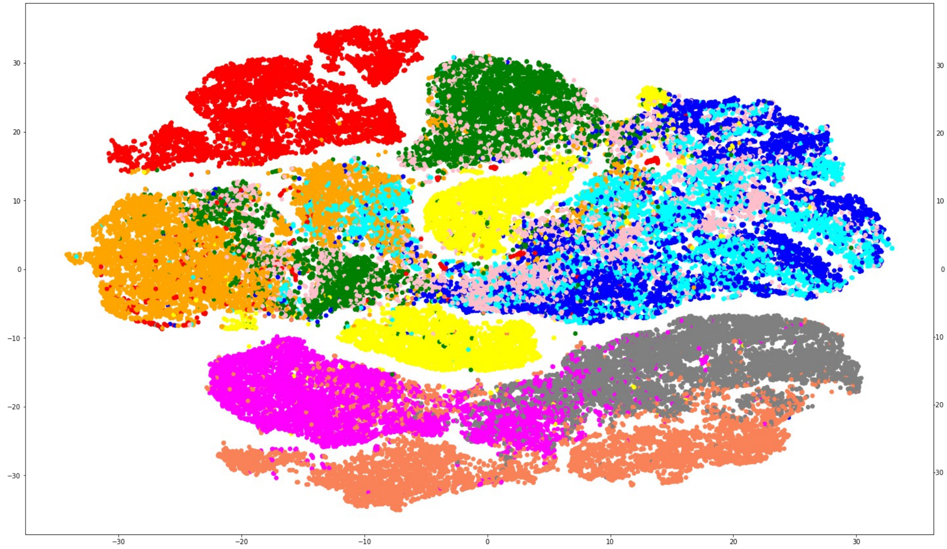
Like PCA, t-SNE reduces the number of dimensions. While PCA uses a linear transformation, t-SNE is a nonlinear algorithm that tries to maintain the distances between neighboring points as the number of dimensions are reduced.

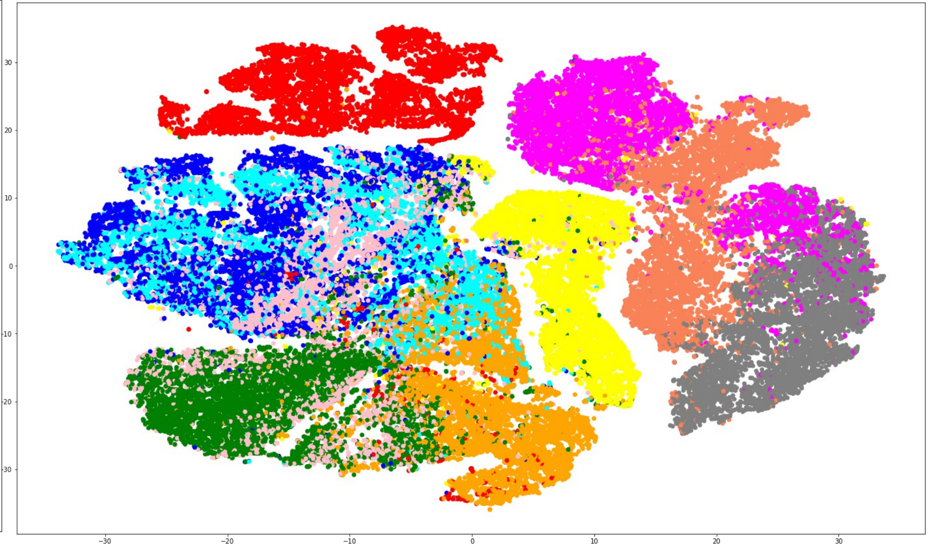


Source: https://distill.pub/2016/misread-tsne/

t-SNE

t-SNE uses some random numbers, so we need to run the algorithm on both the test and training data together. The output looks different on separate runs. Notice that the distances between clusters and their density are not maintained in different runs of t-SNE with the exact same data.

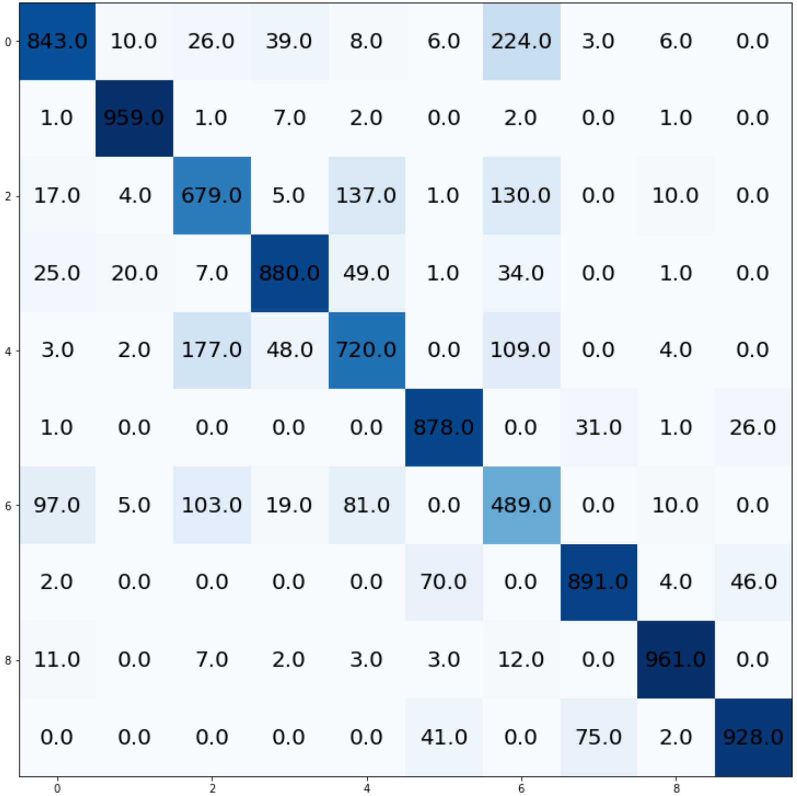




t-SNE - Histogram Classifier w/ 82 bins

Results: Accuracy ~ 82%

- Labels pull-over, coat and shirt have the poorest accuracy.



843.0	10.0	26.0	39.0	8.0	6.0	224.0	3.0	6.0	0.0
1.0	959.0	1.0	7.0	2.0	0.0	2.0	0.0	1.0	0.0
17.0	4.0	679.0	5.0	137.0	1.0	130.0	0.0	10.0	0.0
25.0	20.0	7.0	880.0	49.0	1.0	34.0	0.0	1.0	0.0
3.0	2.0	177.0	48.0	720.0	0.0	109.0	0.0	4.0	0.0
1.0	0.0	0.0	0.0	0.0	878.0	0.0	31.0	1.0	26.0
97.0	5.0	103.0	19.0	81.0	0.0	489.0	0.0	10.0	0.0
2.0	0.0	0.0	0.0	0.0	70.0	0.0	891.0	4.0	46.0
11.0	0.0	7.0	2.0	3.0	3.0	12.0	0.0	961.0	0.0
0.0	0.0	0.0	0.0	0.0	41.0	0.0	75.0	2.0	928.0

0
2
4
6
8
0
2
4
6
8

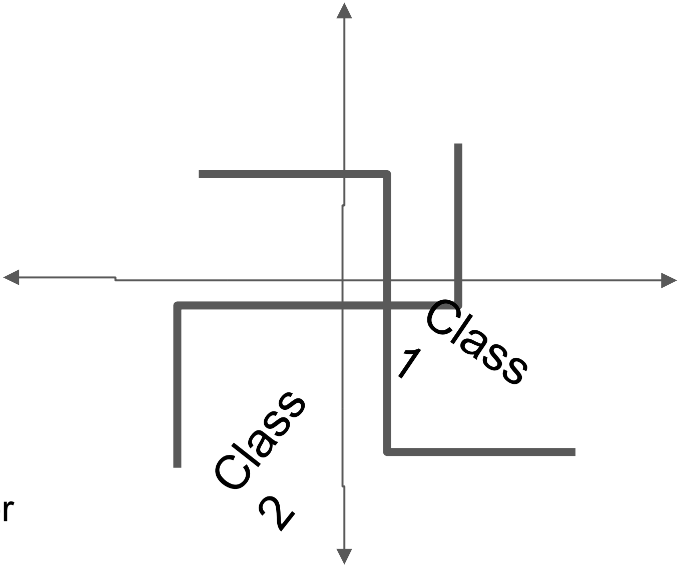
Subspace-based Classification - Overview

Assumption
The data from different classes lie in different subspaces

Classification Model
C subspaces - one for each class
(each subspace is computed with PCA
using training data from the corresponding class)

Prediction on a given test data

1. Compute reconstruction error with C subspaces
2. Predicted class is the one with smallest reconstruction error



Subspace-based Classification - Details

Prediction Details

For x in test data, compute the following for all the 10 classes (C)

zc = x - muc

pc = zc.vcT

rc = pc.vc (reconstruction)

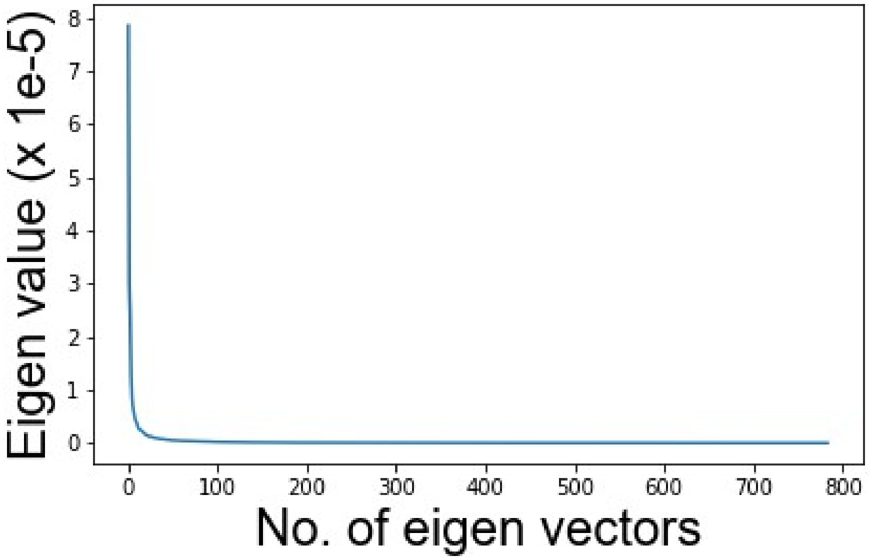
rc_err = norm(zc - rc)

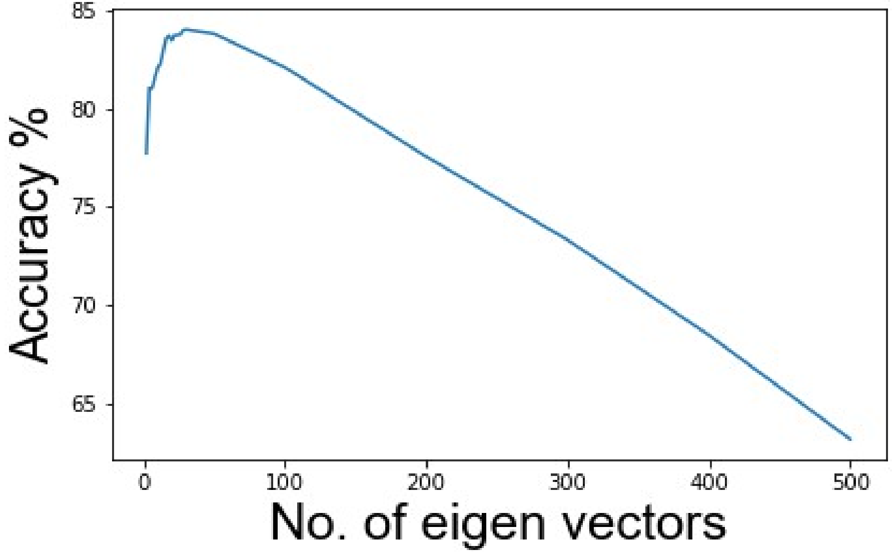
r_err_arr = [r1_err, r2_err, r3_err,...,r10_err]

C_pred = argmin(rc_err_arr)

Subspace-based Classification - Results

- The best accuracy is **84%** with 28 dimensions
- When subspace dimension is small, the accuracy is low because it does not represent the data well
- When subspace dimension is large, a given test data gets reconstructed well with all the subspaces and hence, the classification accuracy reduces
- There is an optimal subspace dimension as seen in the accuracy plot





Classification Using MLP Neural Networks

- Create MLP Neural Network using Stochastic Gradient Descent Algorithm
- Input is 784 pixels
- 3 hidden layers of 784 neurons each
- 10 outputs for classifications for 10 types of data classes
- Compared neural network performance using different activation functions
- Activation functions used - sigmoid, relu and tanh
- Observations: Relu activation function yields maximum accuracy (88%) compared to sigmoid (76%) and tanh (72%)
- Conclusions: Training of neural network takes time and resources but once the network is trained, the weights and biases parameters can be saved and can be used for fast prediction without need for retraining
- Best accuracy achieved: 88%

Convolutional Neural Network Background

- Machine Learning/Deep Learning Algorithm that is used to visualize images by having the convolution, pooling, fully connected and normalization networks.
- Consists of inner and outer layers for feed forward of an Artificial Neural Network.
- Made of neurons, learnable weights, and biases.
- A design for minimal preprocessing by creating multilayer perceptrons.
- Weights are used to change the features or fields.
- The use of Convolutional Neural Network is demonstrated on the next slide.

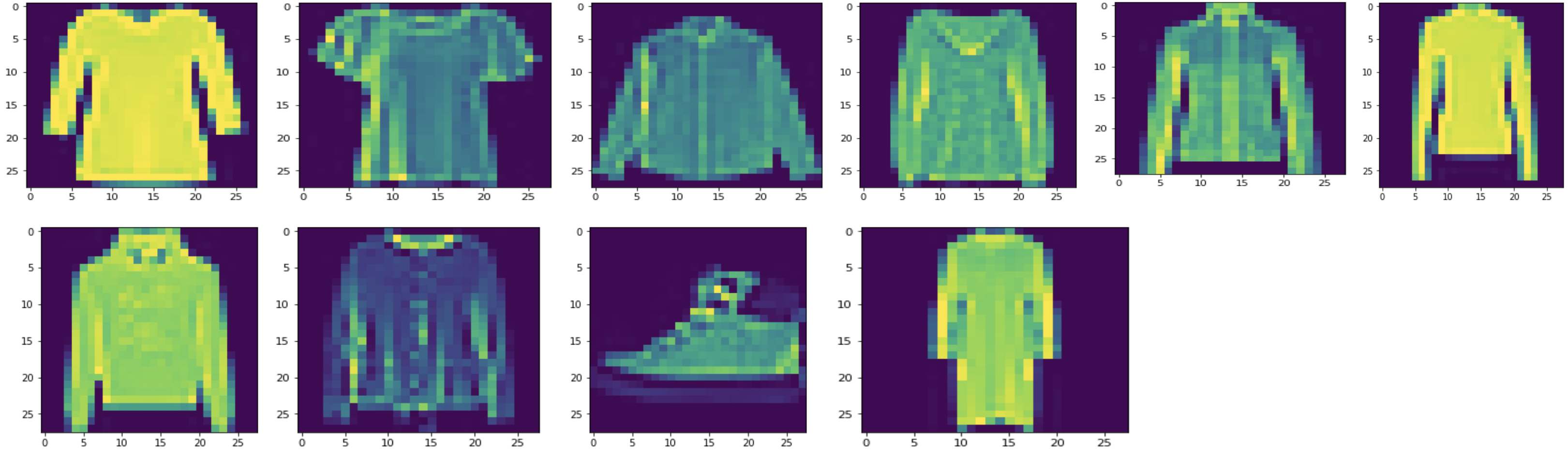
Convolutional Neural Network Example

- By utilizing the Convolutional Neural Network algorithm, the variable sparse accuracy is much higher in comparison to the Naive Bayes Algorithm at 89.36%, using the base features to perform the Validation Accuracy.
- However, there are many features that could be changed in order to make adjustments for the weights whether it could be higher accuracy or lower accuracy.
- Example is demonstrated below:

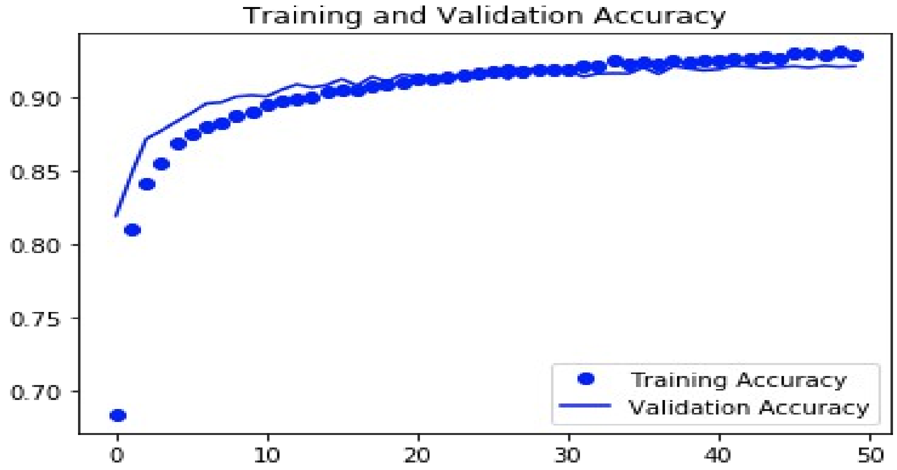
    -
      ```
      Epoch 19/20
      60000/60000 [==============================] - 121s 2ms/step - loss: 0.2101 -
      sparse_categorical_accuracy: 0.9239 - val_loss: 0.3031 - val_sparse_categorical_accuracy: 0.8943
      Epoch 20/20
      60000/60000 [==============================] - 123s 2ms/step - loss: 0.2017 -
      sparse_categorical_accuracy: 0.9271 - val_loss: 0.3080 - val_sparse_categorical_accuracy: 0.8936
      ```

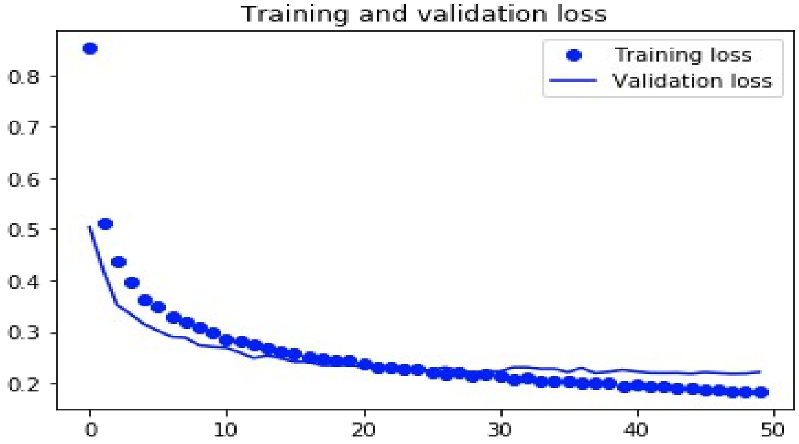
Predicted Images on Convolutional Neural Network

- These images below are misclassified as follows:
  - Predicted Class: Shirt, T-Shirt, Coat, Shirt, Coat, Shirt, Coat, Coat, Sneaker, Dress
  - True Class: T-Shirt/Top, Shirt, Shirt, Pullover, Shirt, Pullover, Pullover, Pullover, Ankle Boot, Coat



Area Under the Curve Visualizations





By splitting the different datasets to training, testing, and validation, the accuracy losses matches up to what is described in these two plots, where it measures to 92.7% accuracy. The loss is at 20%.

Results Summary

Method	Accuracy (%)
Naive Bayes	59%
PCA Histogram (6 principal components, 9 bins)	72%
t-SNE Histogram (2 dimensions, 82 bins)	82%
Subspace-based Classification	84%
MLP Neural Network with relu activation function (3 hidden layers)	88%
Convolution Neural	

Closing thoughts

I enjoyed having the chance to genericize problem set 3 to support all 10 classes and varying bin and principal components counts. I also enjoyed seeing how much my teammates were able to improve on accuracy with more complex models. I would not have known even where to begin on building a classifier without this class. - Graham

I have learned much more insights by using algorithms that I am not familiar with and to utilize new methods to do further analyses. I also have enjoyed working with different group members to tackle the challenging tasks. The complex process on running the algorithms has been the surprise as I am not aware how long it has taken to run it. As a result, I have gone on to concentrate on other parts of the project. I could not have done it without attending the class in order to ask for clarification. - Andy

I learned that training of neural network is resource intensive and time consuming process. But the accuracy achieved using neural network is high.
learni
highe
attending the class - Ajoy

- What did you enjoy about your project? Did you gain any new insights?
- What, if anything, took you by surprise? What did you learn as a result of this?
- Could you have done this project without attending this class?