

Parker Williamson  
2/13/18  
Springboard

### Inferential Statistics

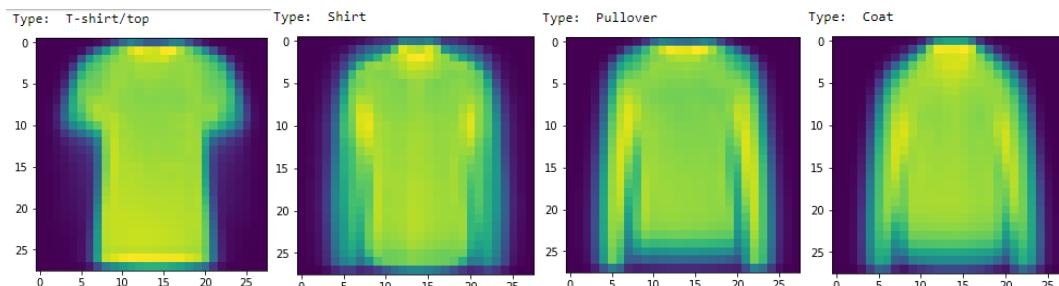
In order to compare how correlated the categories of clothing are I compared the means of the data. The most similar means would be next to each other, so I ordered the means from smallest to largest and then did a z-test to determine if the different categories have the same means. A z-test assumes normal distribution and whether two different samples have the same mean. For a z-test the sample sizes should be larger than 30 (here it is 6000). Z-tests use the mean and standard deviation. The ordered means of the photos have no statistically significant correlation, because all of the p-scores of the closest means are below .05. A p-score shows the results of the null hypothesis test determining where or not the means are correlated. It is not a probability, but you can see how far away the results are from being above .05.

```
p-score 0-1(Z): 0.0
p-score 1-2(Z): 0.0
p-score 2-3(Z): 0.0
p-score 3-4(Z): 0.0
p-score 4-5(Z): 0.0
p-score 5-6(Z): 0.0001
p-score 6-7(Z): 0.0
p-score 7-8(Z): 0.0
p-score 8-9(Z): 0.0
```

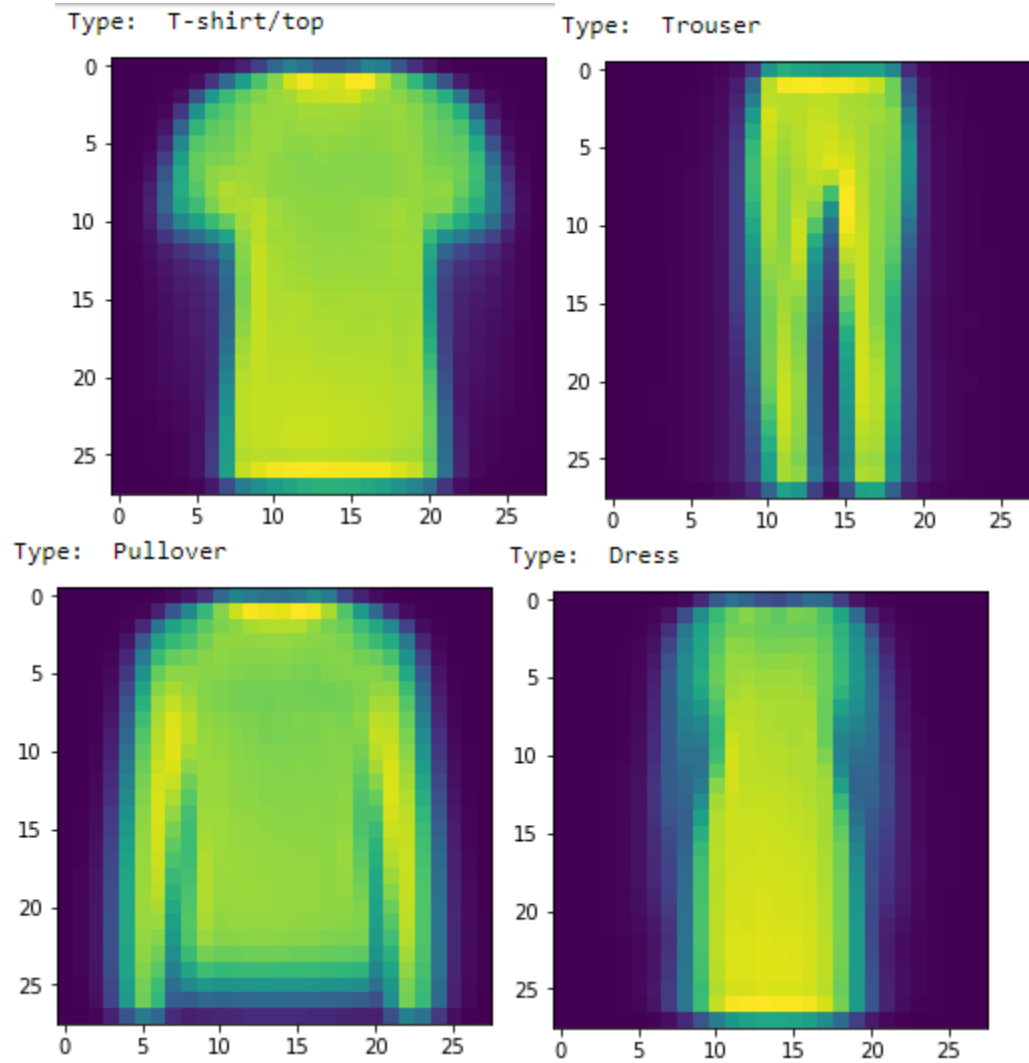
Another way I compared the relationship between the categories was to analyze the results of the classification. After using standardization and the MLP classification I calculated the 2.5-97.5% percentile range of the total errors of each type of misclassification (for example the total number of shirts classified as sandals). 2.5% of the data is at 0 and 97.5% of the error is 110.4 errors per misclassification. Anything outside that percentile range is considered an error.

```
95% percentiles: [ 0. 110.4]
outlier counts: [ 114. 201. 152.]
T-shirt/top is commonly classified as a Shirt
Pullover is commonly classified as a Coat
Shirt is commonly classified as a T-shirt/top
```

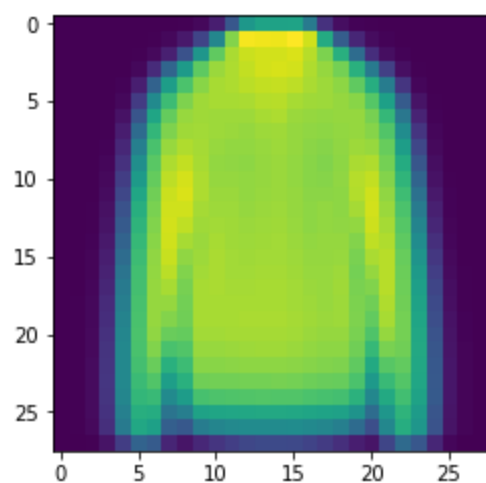
Based on the number of misclassifications there were 3 types that were unusually common. T-shirts/tops were commonly classified as shirts, Pullovers were commonly classified as coats and shirts were commonly classified as T-shirts/tops. Those are some of the categories that would be hard for a human to classify as well as you can see when looking at the average image of those types.



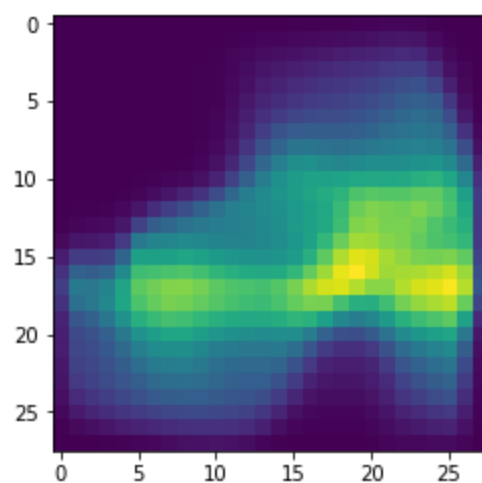
The average images of each different class:



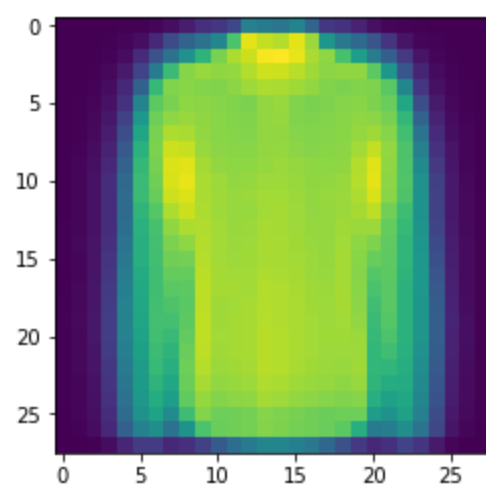
Type: Coat



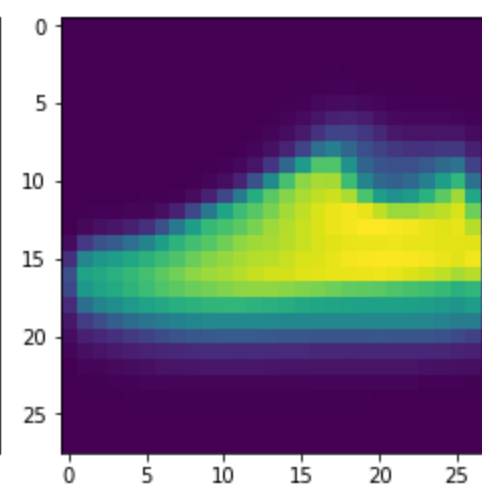
Type: Sandal



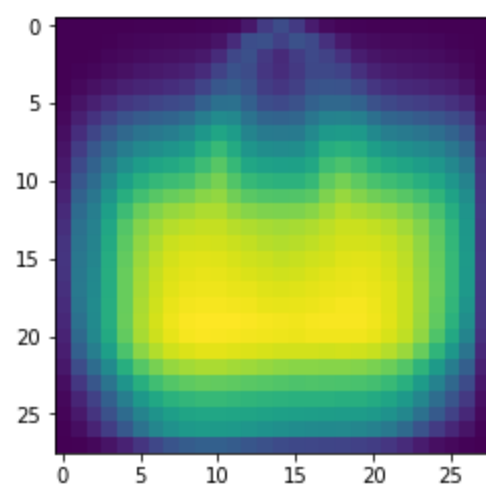
Type: Shirt



Type: Sneaker



Type: Bag



Type: Ankle boot

