COMP300027 Machine Learning, 2021 Semester 1

Project 1: Pose classification with naïve Bayes

|  |  |
| --- | --- |
| 1085020 | Mihai Blaga |
| 963632 | Kai Stevens-Noguchi |

# **Question 1**

Nine different techniques for calculating scores are evaluated and outputted in the evaluate() function, but for this question we will focus on *macro precision* (MP) and *weighted average precision* (WAP).

The critical difference between MP and WAP is in the weight each class is given for the final calculation. In MP, each class is treated equally, so each class' precision contributes an equal amount to the overall MP calculation. However, in WAP, as the name implies, each class contributes an amount proportional to its occurrence in the test set.

An assumption made when choosing the test and training data is that both sets of data are representative of the real world. As such, it is expected that both sets will have similar distributions of class labels, i.e. a class which is rarely seen in the training data is also rarely seen in the test data.

Further, it is generally true that the system *learns* more about labels that are seen more frequently. As such, it is expected that frequently occurring labels will score higher than labels which are not seen as frequently.

As MP treats all classes equally, this expected correlation between frequency and score is lost while in WAP, it impacts the final outcome. As seen in the table labelled "Output table for Q1", if the data is separated into two halves based on precision, the higher scoring half collectively occurs 74 times while the lower scoring half occurs 42 times, supporting the expected correlation.

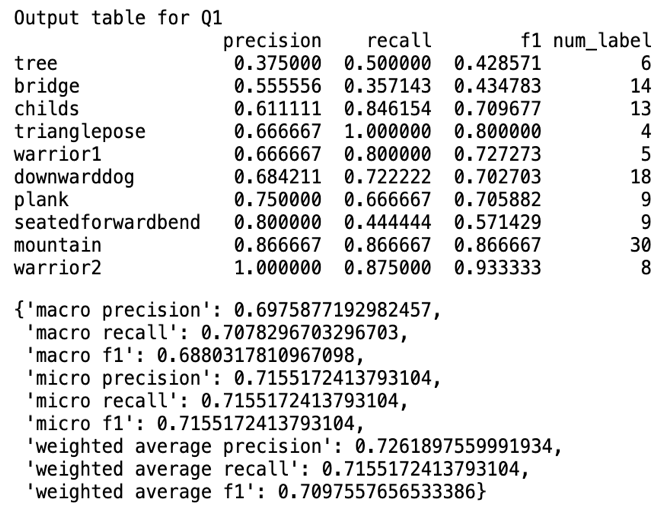


fig 1 Output table for Q1

# **Question 2**

The assumption that the numerical attributes of the set come from a Gaussian distribution is sound most of the time. However, there exist cases, like in the tree pose where this is not the case.

The issue with assuming that the attributes come from a Gaussian distribution is when, within a class, there are multiple subclasses from which it is composed. This is most obvious in the tree pose where the hands are either above the head or around the chest area (these can be thought of as classes 'tree with arms above' and 'tree with arms below'). As can be seen in both the histograms and scatterplots from the figures for Q2, attributes like (representing the head) and , (representing the arms) for tree pose have two peaks and can be separated into quite distinct clusters visually.

The effect that this has on the overall performance is evident as the tree pose (in which this double peak is most evident) has the worst performance of any other class (see in split test scores for Q1). The next lowest scoring class is bridge, within which a similar effect can be seen in attributes like , and .

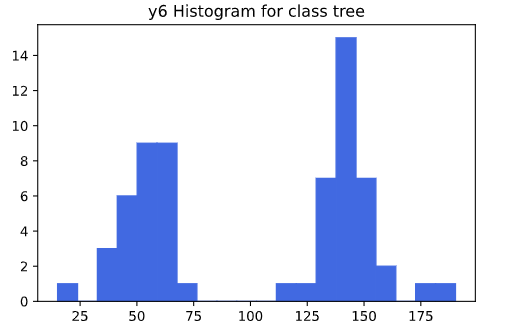
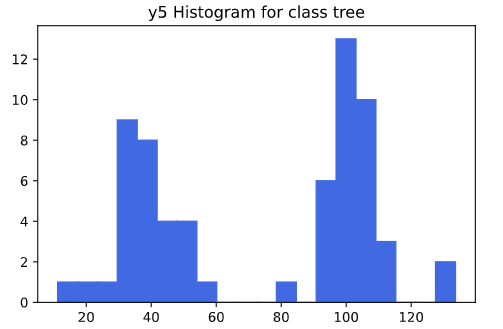


fig 3 Histogram for y5 and y6

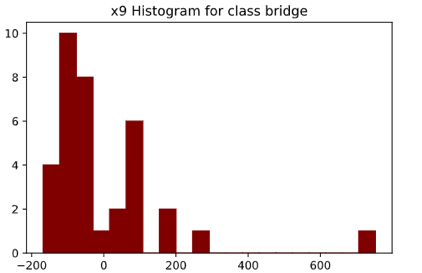
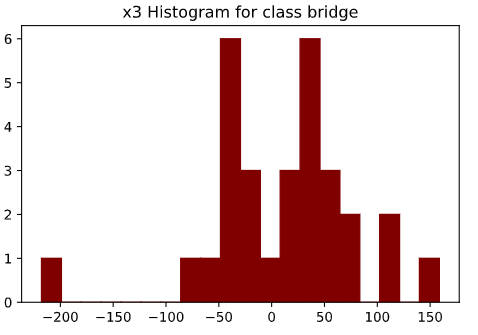
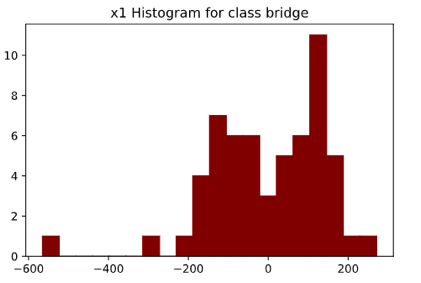


fig 4 x1, x3 and x9 for class bridge

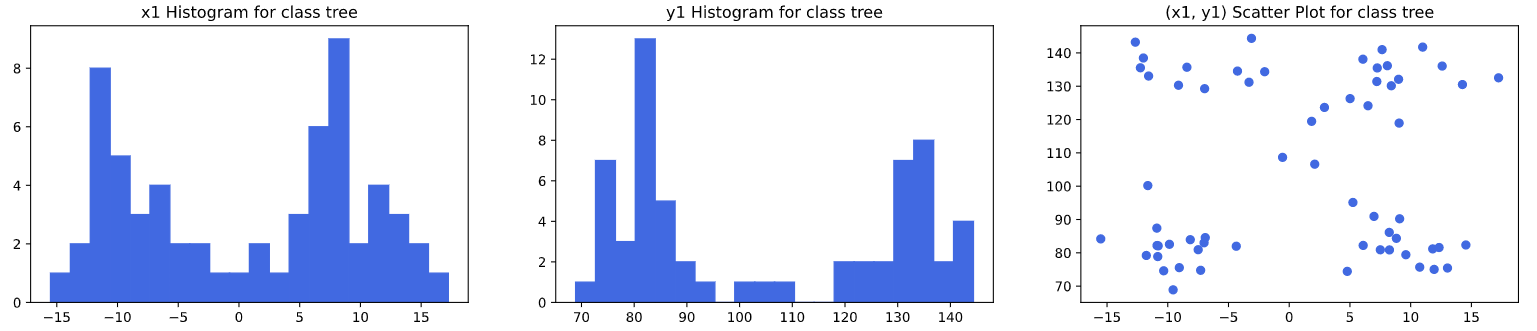


fig 2 Histograms and scatter plot for head

# **Question 5**

Overall, using missing value as part of information caused a slight decrease in scores except for *weighted average precision* (WAP). Looking at individual poses, such as the triangle pose, downwardsdog and mountain, an increase in score was observed. However, score decrease in poses such as bridge, child, plank and seatedforwardbend, were also observed.

An increase in WAP was caused as information on missing values increased precision for data that occurred more frequently in test sets such as mountain pose. It seems that adding missing value works well for cases when there are fewer missing values in both train and test data (triangle pose and mountain) and/or there are more samples in train data (downwardsdog).

There was a significant drop in the precision of seatedfarwardbend and recall of plank. In both cases, while one score changed dramatically, the other score stayed the same, which implies missing value information has caused an increase in false positives or an increase in false negatives. In either case, the seatedfarwardbend has a relatively high proportion of missing values in both train and test data and plank has fewer sample in both train and sample data.

With the provided train and test data, incorporating missing value information did not have a significant effect on classification results.

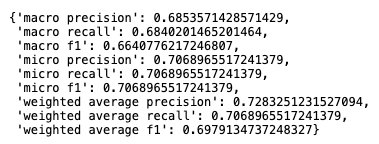


fig 7 Overall score of Naïve Bayes using missing value information

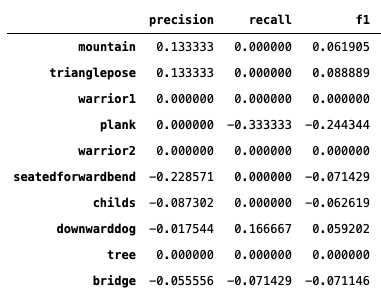


fig 5 Percentage change in scores

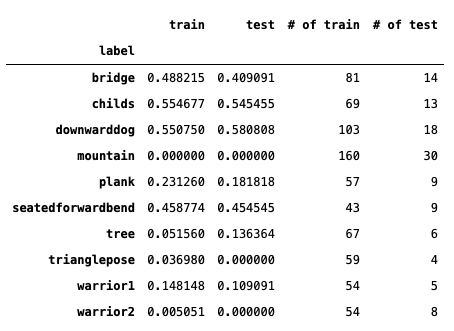


fig 6 Class count in train and test data and missing value proportions

# **Question 6**

The engineered pose features chosen were horizontal, vertical and Euclidean distance between the head (1) and select key points (hands (4, 6) and feet (9,11)). As these points could be missing or otherwise contain null values, if either value was null, the Q5 method of null representation (NR) was used.

We chose to employ the NR of Q5 because, in testing, a strong correlation was found between the number of missing values and a lack of performance in the engineered pose features which was addressed under the new NR. When comparing to the standard (x,y) features, we compare to the output of Q5 since this uses the same NR and is thought to be a more accurate comparison.

Overall, the performance was lower than that of the standard features scoring ~0.07 less in most metrics. The only metric which scored higher was WAP and macro precision, but the difference is negligible.

Interestingly, it appears that the engineered features are more sensitive to missing data when compared to the standard features. The three classes which had the largest fall in prediction quality were "seatedforwardbend", "trianglepose" and "childs" all of which had missing proportions > 0.3. Similarly, the classes which saw the greatest improvement in prediction "tree", "warrior1" and "plank", all had missing proportions < 0.2.

テーブル

中程度の精度で自動的に生成された説明As such, these new features can be thought of as a dimension reduction at the cost of some performance. This performance difference is likely due to a higher conditional dependance between the new features.

fig 9 Per class scores

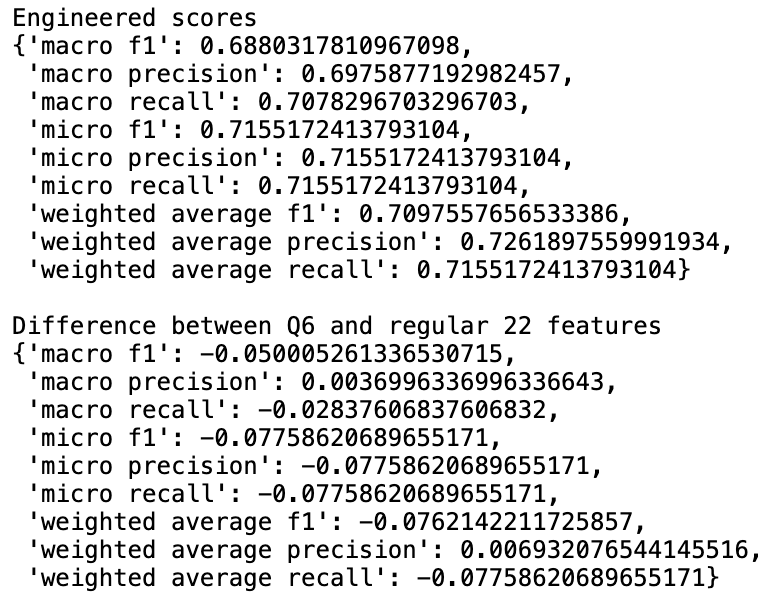


fig 8 Overall scores

テーブル

自動的に生成された説明テーブル

自動的に生成された説明

fig 11 Proportion of missing features per class

fig 10 Score change between 22 features and Q6 features for individual class