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.

Output table for Q1

	precision	recall	f1	num_label
tree	0.375000	0.500000	0.428571	6
bridge	0.555556	0.357143	0.434783	14
childs	0.611111	0.846154	0.709677	13
trianglepose	0.666667	1.000000	0.800000	4
warrior1	0.666667	0.800000	0.727273	5
downwarddog	0.684211	0.722222	0.702703	18
plank	0.750000	0.666667	0.705882	9
seatedforwardbend	0.800000	0.444444	0.571429	9
mountain	0.866667	0.866667	0.866667	30
warrior2	1.000000	0.875000	0.933333	8

```
{'macro precision': 0.6975877192982457,
 'macro recall': 0.7078296703296703,
 'macro f1': 0.6880317810967098,
 'micro precision': 0.7155172413793104,
 'micro recall': 0.7155172413793104,
 'micro f1': 0.7155172413793104,
 'weighted average precision': 0.7261897559991934,
 'weighted average recall': 0.7155172413793104,
 'weighted average f1': 0.7097557656533386}
```

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 (x_1, y_1) (representing the head) and y_5, y_6 (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 x_1, x_3 and x_9.

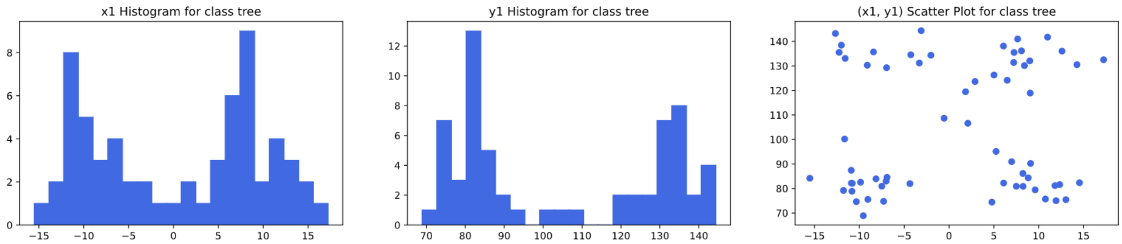

fig 2 Histograms and scatter plot for head

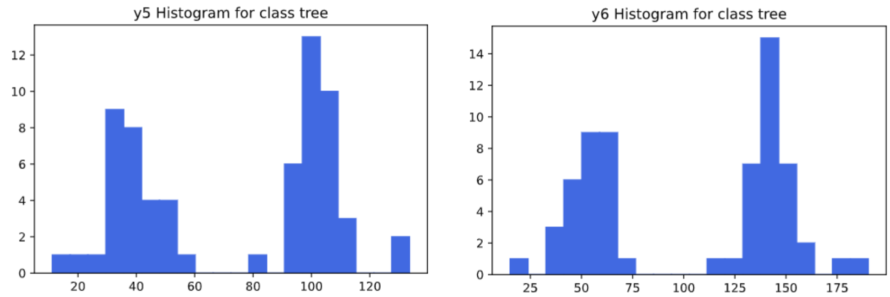

fig 3 Histogram for y5 and y6

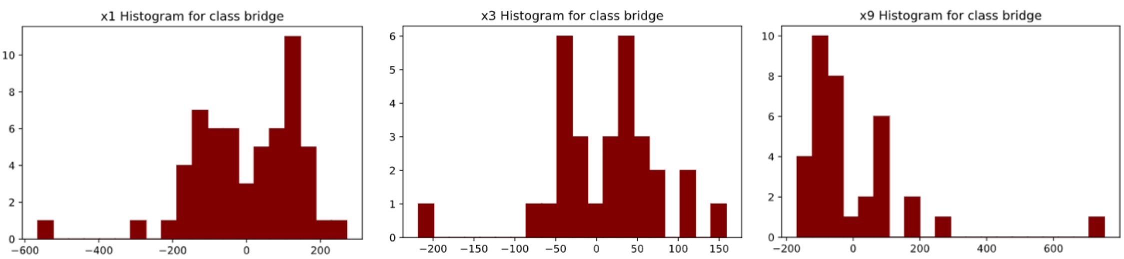

fig 4 x1, x3 and x9 for class bridge

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.

	precision	recall	f1
mountain	0.133333	0.000000	0.061905
trianglepose	0.133333	0.000000	0.088889
warrior1	0.000000	0.000000	0.000000
plank	0.000000	−0.333333	−0.244344
warrior2	0.000000	0.000000	0.000000
seatedforwardbend	−0.228571	0.000000	−0.071429
childs	−0.087302	0.000000	−0.062619
downwarddog	−0.017544	0.166667	0.059202
tree	0.000000	0.000000	0.000000
bridge	−0.055556	−0.071429	−0.071146

fig 5 Percentage change in scores

label	train	test	# of train	# of test
bridge	0.488215	0.409091	81	14
childs	0.554677	0.545455	69	13
downwarddog	0.550750	0.580808	103	18
mountain	0.000000	0.000000	160	30
plank	0.231260	0.181818	57	9
seatedforwardbend	0.458774	0.454545	43	9
tree	0.051560	0.136364	67	6
trianglepose	0.036980	0.000000	59	4
warrior1	0.148148	0.109091	54	5
warrior2	0.005051	0.000000	54	8

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

```
{'macro precision': 0.6853571428571429,
 'macro recall': 0.6840201465201464,
 'macro f1': 0.6640776217246807,
 'micro precision': 0.7068965517241379,
 'micro recall': 0.7068965517241379,
 'micro f1': 0.7068965517241379,
 'weighted average precision': 0.7283251231527094,
 'weighted average recall': 0.7068965517241379,
 'weighted average f1': 0.6979134737248327}
```

fig 7 Overall score of Naïve Bayes using missing value information

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.

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.

```
Engineered scores
{'macro f1': 0.6880317810967098,
 'macro precision': 0.6975877192982457,
 'macro recall': 0.7078296703296703,
 'micro f1': 0.7155172413793104,
 'micro precision': 0.7155172413793104,
 'micro recall': 0.7155172413793104,
 'weighted average f1': 0.7097557656533386,
 'weighted average precision': 0.7261897559991934,
 'weighted average recall': 0.7155172413793104}

Difference between Q6 and regular 22 features
{'macro f1': -0.050005261336530715,
 'macro precision': 0.0036996336996336643,
 'macro recall': -0.02837606837606832,
 'micro f1': -0.07758620689655171,
 'micro precision': -0.07758620689655171,
 'micro recall': -0.07758620689655171,
 'weighted average f1': -0.0762142211725857,
 'weighted average precision': 0.006932076544145516,
 'weighted average recall': -0.07758620689655171}
```

fig 8 Overall scores

Individual class performance

	precision	recall	f1	num_label
seatedforwardbend	0	0	0	9
downwarddog	0.380952	0.888889	0.533333	18
bridge	1	0.285714	0.444444	14
plank	1	0.555556	0.714286	9
tree	0.384615	0.833333	0.526316	6
warrior2	1	0.875	0.933333	8
trianglepose	0.5	1	0.666667	4
mountain	1	0.733333	0.846154	30
warrior1	1	1	1	5
childs	0.625	0.384615	0.47619	13

fig 9 Per class scores

	precision	recall	f1
seatedforwardbend	−0.571429	−0.444444	−0.5
downwarddog	−0.285714	0	−0.228571
bridge	0.5	0	0.0808081
plank	0.25	0.222222	0.252747
tree	0.00961538	0.333333	0.0977444
warrior2	0	0	0
trianglepose	−0.3	0	−0.222222
mountain	0	−0.133333	−0.0824176
warrior1	0.333333	0.2	0.272727
childs	0.10119	−0.461538	−0.170868

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

Proportion of missing features

	train	test	# of train	# of test
label				
bridge	0.388889	0.350649	81	14
childs	0.450593	0.409091	69	13
downwarddog	0.476611	0.477273	103	18
mountain	0.000000	0.000000	160	30
plank	0.188995	0.166667	57	9
seatedforwardbend	0.374207	0.333333	43	9
tree	0.050882	0.113636	67	6
trianglepose	0.032357	0.000000	59	4
warrior1	0.123737	0.081818	54	5
warrior2	0.007576	0.000000	54	8

fig 11 Proportion of missing features per class