1 Exercise 1

1.1 Cross Entropy

H(y,g)= -(0*log(0.25)+1*log(0.6)+0*log(0.15))=0.222

1.2 Mean Squared Error Loss

MSE(y,g)= $\frac{1}{3}(0.25^2 + (-0.4)^2 + 0.15^2)$=0.0817

1.3 Hinge Loss

SVM(y,j)=max(0,1.25)+max(0,0.6)max(0,1.15)=3

2 Task A

2.1 Accuracy

$$ACC = \frac{TP + TN}{P + N}$$

, high accuracy of predicting predominant class is easily achieved and misleading. If we have highly unbalanced data the problem of accuracy as a metric is quite apparent: Let's consider an example with medical data: normally only a few patients have a certain disease. If our dataset consists of 200 patients, of which only 8 have cancer, than a classifier which predicts that no patient has cancer would yield 96 percent accuracy; this prediction however would be deadly for the patients. If we also consider false negatives, than we see that in every case of a cancer patient, the classifier predicted no cancer and thus false negatives are 100 percent.The ability to capture misclassifications in a satisfying way is what is lacking in the accuracy metric.

In a muti-class setting, exact matching does not distinguish between partially correct and completely incorrect labeling. The denominator, **P+N** which is equivalent to number of samples(N) is constant. It is thus sensitive to only how many labels correctly identified with respect to N

For a single class C Precision, Recall and $F1$ are defined as follows:

$$F1_c = \frac{2|T_c \cap P_c|}{|P_c| + |T_c|}$$

The hamonic mean of precision $Precision = \frac{|T_c \cap P_c|}{|T_c|}$ and recall $Recall = \frac{|T_c \cap P_c|}{|P_c|}$

$$Jacc_c = \frac{|T_c \cap P_c|}{|T_c \cup P_c|}$$

which can e rewritten as

$$Jacc_c = \frac{|T_c \cap P_c|}{|P_c| + |T_c| - |T_c \cap P_c|}$$

$$Jacc_c \leq F1_c$$

Jaccard metric penalize bad classification more than the F score "https://stats.stackexchange.com/questions/27 dice-score-vs-iou"

2.2 Task B

GT Class	#	Freq
B	4	0.5
TD	4	0.5
all other classes$(B^*C^*D^*T^*)\backslash(B, TB)$	0	0

Evaluation of class B

$$|T_b| = |4B| = 4$$

$$|P_b| = |2B, 2T, 2D| = 6$$

$$Jacc_b = \frac{|4B \cap (2B, 2T, 2D|)}{|P_b| + |T_b| - |(4B \cap 2B, 2T, 2D)|}$$

$$Jacc_b = \frac{|2B|}{4 + 6 - |2B|}$$

$$Jacc_b = \frac{2}{4 + 6 - 2} = \frac{1}{4}$$

Following the definition :

$$Precision_b = \frac{2}{6} = \frac{1}{3}$$

$$Recall_b = \frac{2}{4} = \frac{1}{2}$$

$$F1 = \frac{2(Precision \times Recall)}{Precision + Recall} = \frac{2}{5}$$

Evaluation of class TD

$$|T_{td}| = |4T, 4D| = 8, |P_{td}| = |3T, 2C, B, D| = 7$$

$$Jacc_b = \frac{|(4T, 4D) \cap (3T, 2C, B, D|)}{|P_{td}| + |T_{td}| - |(4T, 4D) \cap (3T, 2C, B, D|)|}$$

$$Jacc_{td} = \frac{|(3T, D)|}{8 + 7 - |(3T, D)|}$$

$$Jacc_{td} = \frac{4}{8 + 7 - 4} = \frac{4}{11}$$

$$Precision_{td} = \frac{4}{7}$$

$$Recall_{td} = \frac{4}{8} = \frac{1}{2}$$

$$F1 = \frac{2(Precision \times Recall)}{Precision + Recall} = \frac{8}{15}$$

Mean Computations

class balance:

$$Jacc_{cb} = \frac{1}{n}\sum_{i=1}^{n}\frac{|T_i \cap P_i|}{|T_i \cup P_i|} = \frac{1}{2}\left(\frac{1}{4} + \frac{4}{11}\right) = 0.3068$$

$$Precision_{cb} = \frac{1}{n}\sum_{i=1}^{n}\frac{|T_i \cap P_i|}{|T_i|} = \frac{1}{2}\left(\frac{1}{3} + \frac{4}{7}\right) = 0.4\overline{523809}$$

$$Recall_{cb} = \frac{1}{n}\sum_{i=1}^{n}\frac{|T_i \cap P_i|}{|P_i|} = \frac{1}{2}\left(\frac{1}{2} + \frac{1}{2}\right) = 0.5$$

$$F1_{cb} = \frac{1}{n}\sum_{i=1}^{n}\frac{2|T_i \cap P_i|}{|P_i| + |T_i|} = \frac{1}{2}\left(\frac{2}{5} + \frac{8}{15}\right) = 0.4\overline{6}$$

class frequency:

Because we only have two classes and both have a freq of 0.5 the results are the same as with the class balance.

exact match:

$$ExMatch = \frac{\#correctInstances}{\#instances} = \frac{2}{8} = 0.25$$

Questions

1. Intuitively the Exact Match would be the strictest metric possible. However,it might not be the lowest number: how come?

Because the absolute values of these metrics are not comparable. A value of 1 in the jaccard metric does not mean the same as a $F1$ value of 1

2. Can you compute the global accuracy on this example? Justify your answer.

3. What is the main issue going from multi-class to multi-label setting?

I a multi-class setting a prediction is eighter correct of not. In a multi-label setting we can have partially correct labels. Under these conditions we need different definition for the metrics.