50.021 – AI

Alex

Week 01: Discriminative ML - quick intro

[The following notes are compiled from various sources such as textbooks, lecture materials, Web resources and are shared for academic purposes only, intended for use by students registered for a specific course. In the interest of brevity, every source is not cited. The compiler of these notes gratefully acknowledges all such sources.]

Due: week2 Thursday, 30th of May, 6pm

Understand the behaviour of performance measures under baselines which do not look at the data. How good can you get without learning? Understand the difference between these two measures.

- Consider the following two measures vanilla Accuracy

$$\frac{1}{n}\sum_{i=1}^{n} 1[f(x_i) == y_i]$$

  class-wise averaged accuracy:

$$\begin{aligned}
A =& \frac{1}{C}\sum_{c=1}^{C} a_c \\
a_c =& \frac{1}{\sum_{i=1}^{n} 1[y_i == c]} \sum_{i=1}^{n} 1[y_i == c] 1[f(x_i) == c] \\
=& \frac{1}{\sum_{(x_i,y_i):y_i=c} 1} \sum_{(x_i,y_i):y_i=c} 1[f(x_i) == c]
\end{aligned}$$

- consider a three class problem with the following frequencies of test labels: p_1, p_2, p_3 with a sample size of n of the test set
- suppose you do baselines predictions which do not look at the data to arrive at a prediction. Compute these two accuracies obtained when
  - you predict constantly the most frequent class $f(x_i) = \text{argmax}_c p_c$. For notation it can be helpful to abbreviate $\text{argmax}_c p_c = c_\uparrow$, $\max_c p_c =$

$p_\uparrow$. Note that the argmax returns the input argument – the class, while the max returns the value, here the probability.

Hint: Decompose $\frac{1}{n}\sum_{i=1}^{n} 1[f(x_i) == y_i]$ into the sum of losses for the three label classes weighted by the correct weights.

- you predict constantly the least frequent class $f(x_i) = \text{argmin}_c p_c$. For notation it can be helpful to abbreviate $\text{argmax}_c p_c = c_\downarrow$, $\max_c p_c = p_\downarrow$
- compute the expected values of these two accuracies when you predict for each sample x the class c with probability q_c where q_1, q_2, q_3 is some given probability $q_c > 0, \sum_c q_c = 1$.
- compute the expected values of these two accuracies when you predict for each sample x the class c with probability p_c equal to the above frequencies of the test labels, that is when $q_c = p_c$.
- Suppose you predict constantly the most frequent class $f(x_i) = \text{argmax}_c p_c$. Find the set of test set frequencies p_1, p_2, p_3 such that (A) the vanilla accuracy will be equal to the class-wise averaged accuracy. Show a set of test set frequencies p_1, p_2, p_3 where (B) the vanilla accuracy will be much higher than the class-wise averaged accuracy (for example, provide a p_1, p_2, p_3, such that the vanilla accuracy is 0.95, but the class-wise accuracy is $\frac{1}{3}$ only). I hope that shows you the difference between these two measures.

Show your work when calculating.