CSDS 440: Assignment 3

Shaochen (Henry) ZHONG, `sxz517`

Mingyang TIE, `mxt497`

Due on 09/25/2020, submitted early on 09/18/2020

Problem 10

It depends on the task and how is the performance of the model being "worsen" on test data than training data. Say we have a task to detect fire of a building so that less people get hurt, and the model performance is lower on test data than training data due to having a lot of false positives. Such model may still be beneficial as the cost of having a false negative is a lot more expensive than having a false positive in this particular task. And even though the model might be overfitting by definition, it might perform better than a model that is less overfit but has more false negative.

Problem 11

Problem 12

No. This is not a goth methodology due to an effective concept usually takes a reasonably large amount of training to learn. However by having an equal-size training and evaluation sets, the training set is likely not large enough – or at least not as effective having a larger training set.

Also because of the equal-sized division, the examples in the training set of during an iteration can be very different to another iteration. This inconsistency will increase the difficuty for person X to analyse wheather it is the problem on training data or the model itself, should there ever be any undesired/unstable performance measures.

Problem 13

Because ROC graph is ploting TP Rate $= \frac{TP}{TP+FN}$ against FP Rate $= \frac{FP}{FP+TN}$. As we are lowering the classification threshold, more examples will be classfied as *Positive*. This implies there will be more TP and FP (for a typical model) as the threashold being lower – and since both $TP + FN$ and $FP + TN$ are constant for all time – we will have a larger numerators on the same denominators. Which will result in an increase on both axises and the statament is therefore proven.

Problem 14

Let R denotes the examples that are being classfied as *Positive*, and T denotes the true posive cases. We have:

$$
\begin{aligned}
P(R) &= P(R \mid T)P(T) + P(R \mid T^c)P(T^c) \\
&= P(R \mid T)(1 - P(T^c) + P(R \mid T^c)P(T^c) \\
&= P(R \mid T) - P(R \mid T)P(T^c) + P(R \mid T^c)P(T^c) \\
&= P(R \mid T) + [P(R \mid T^c) - P(R \mid T)] \cdot P(T^c) \\
P(R) - P(R \mid T) &= [P(R \mid T^c) - P(R \mid T)] \cdot P(T^c)
\end{aligned}
$$

We know that there must be $P(R) = P(R \mid T)$ as random guessing is an independent variable. We also know we should have $P(T^c) > 0$ for being a meaningful task. Subsituting these into the above equation, we have $0 = P(R \mid T^c) - P(R \mid T) \implies P(R \mid T^c) = P(R \mid T)$. This implies that the TP Rate is the same as the FP Rate, and the ROC graph for a random gussing classifier will therefore be a diagonal line.