Problem Set 5

Q1 – Learning from Noisy Labels

Overview

In this problem, you will build on the proofs for learning from noisy labels we worked through during class. Specifically, you will be investigating the anchor-and-learn approach as well as how to leverage noisy labels at prediction time, if they're available.

Learning Goals

Medical data is messy and true gold labels are often hard or prohibitively expensive to obtain in bulk. As a result, it is often necessary to leverage noisy labels in the training process. In this problem, we hope you:

- Solidify your understanding of how training from noisy labels works,
- Recognize scenarios in which noisy labels may be available at prediction time,
- And understand both theoretically and intuitively why noisy labels at prediction time can lead to more accurate performance.

Background

In medical data, truly gold data labels are often hard and/or expensive to come by, since there is a limited pool of qualified experts. Even with experts, labels can be imperfect. As a result, we often have to train on labels with some noise. In class, we examined the problem as training on a set of labels $(X, \tilde{Y})$ where $\tilde{Y}$ are our noisy labels.

[10 points] 1.1

In the anchor-and-learn work, the authors make the assumption that if an anchor is present, the true label must be 1. For 2.1, we lift that assumption. For example, a patient may be taking Metformin for one of its other indications (e.g. PCOS) or 'DM2' might be mentioned in the notes because a patient's parent has diabetes. Therefore, $p(\tilde{Y} = 1|Y = 0)$ can be nonzero, like in the noisy label example we did in class. Furthermore, when deploying this model, we do in fact have $\tilde{Y}$ available to us, in addition to X, since we can search back in the EHR. As a result, we want you to show how you would leverage the knowledge of the noisy label $\tilde{Y}$ to improve your prediction. Concretely, we want you to show how you would estimate $p(Y = 1|X, \tilde{Y})$ instead of just $p(Y = 1|X)$. In particular, under the class-conditional independence assumption, $\tilde{Y} \perp X|Y$, do the following:

a) [5 points] Express $p(Y = 1|X, \tilde{Y})$ in terms of $p(Y|X)$, $p(\tilde{Y}|Y)$, and $p(\tilde{Y}|X)$.

b) [5 points] Explain how you would estimate each of these quantities: $p(Y|X)$, $p(\tilde{Y}|Y)$, and $p(\tilde{Y}|X)$. Do any require gathering more data? Are these extra estimation tasks feasible? (Hint: review your notes from lecture.)

[10 points] 1.2

Now let us again assume that the presence of an anchor does mean that $Y = 1$, as in the anchor-and-learn paper. Concretely, that means $p(Y = 1|\tilde{Y} = 1) = 1$. However, as before, an anchor is only present in true cases part of the time, e.g. $p(\tilde{Y} = 1|Y = 1) = \alpha$. This is identical to the scenario provided in the Learning Classifiers from Only Positive and Unlabeled Data paper. With this new information about $p(\tilde{Y}|Y)$, plug those values into your answer for Part 1, and provide the probabilities for $p(Y = 1|X, \tilde{Y} = 1)$ and $p(Y = 1|X, \tilde{Y} = 0)$. These should be expressed only in terms of $p(Y|X)$ and α.