CSE 250B Winter 2020 - Quiz 1

Jan 24, 2020 - 30 Minutes **Name:**

---

(1) This is a closed book, closed notes exam. Switch off your cell phone and do not communicate with anyone other than an exam proctor.
(2) Start writing when instructed. Stop writing when your time is up.
(3) Remember that your work is graded on the quality of your writing and explanation as well as the validity of the mathematics.

(1) (5 points) Write down the Bayes Optimal classifier for the following data distribution, and calculate its risk R^*. The instance space $\mathcal{X} = [0,1]$, and the label space is $\{0,1\}$. The marginal μ over the unlabeled data is uniform on $[0,1]$, and $\eta(x) = p(y=1|x) = 2\left|x - \frac{1}{2}\right|$.

Solution.

Since the marginal μ is uniform on $[0,1]$, we have

$$\mu(x) = 1, \forall x \in [0,1].$$

The Bayes optimal classifier h^* has the following classification rule

$$h^*(x) = \begin{cases} 1 & \text{if } \eta(x) \geq 1/2, \\ 0 & \text{otherwise.} \end{cases}$$

Solving the inequality

$$\eta(x) = 2\left|x - \frac{1}{2}\right| \geq \frac{1}{2},$$

we find that the $\eta(x) \geq 1/2$ for $x \in [0, 1/4] \bigcup [3/4, 1]$.

Therefore, the Bayes optimal classifier is

$$h^*(x) = \begin{cases} 1, & \forall x \in [0,1/4] \bigcup [3/4,1], \\ 0, & \forall x \in (1/4, 3/4), \end{cases}$$

and the risk is

$$\begin{aligned}
R^* &= \mathbb{E}_{x\sim\mu}\left[\min(\eta(x), 1-\eta(x))\right] \\
&= \int_0^{1/4} \mu(x)(1-\eta(x))dx + \int_{1/4}^{3/4} \mu(x)\eta(x)dx + \int_{3/4}^{1} \mu(x)(1-\eta(x))dx \\
&= \int_0^{1/4} 1 - \left(2\left(\frac{1}{2} - x\right)\right) dx + \int_{1/4}^{3/4} 2\left|x - \frac{1}{2}\right| dx + \int_{3/4}^{1} 1 - \left(2\left(x - \frac{1}{2}\right)\right) dx \\
&= \int_0^{1/4} 2x dx + \int_{1/4}^{3/4} \left|2x - 1\right| dx + \int_{3/4}^{1} (2-2x) dx \\
&= x^2\Big|_0^{1/4} + (x - x^2)\Big|_{1/4}^{1/2} + (x^2 - x)\Big|_{1/2}^{3/4} + (2x - x^2)\Big|_{3/4}^{1} \\
&= 1/16 + 1/16 + 1/16 + 1/16 = 1/4.
\end{aligned}$$

(2) (5 Points) Does the Bayes optimal classifier change when the marginal distribution $\mu(x)$ changes? If yes, provide an example of such a change; if no, justify your answer.

Solution.

No, the Bayes optimal classifier does not change when only $\mu(x)$ changes. The classification rule only depends on $\eta(x)$.

However, the risk may vary when $\mu(x)$ changes.

(3) (5 Points) Remember that in lecture, we talked about the statistical learning framework, where training and test data are all independent samples drawn from the same underlying data distribution. State whether the statistical learning framework assumption applies in the following case. If your answer is yes, explain why. If your answer is no, explain what is different between training and test data μ (the marginal over x), η (the conditional distribution of $y|x$), or something else and justify your answer.

> Alice wants to train a classifier to predict if a prisoner will reoffend when released early. She collects historical data on prisoners who were released early and whether they reoffended, and uses it to train a classifier. The classifier is tested on all new prisoners.

Solution.

No, the statistical learning framework assumption does not apply in this example most of the time.

The marginal distribution μ may be different between training and test data. The training examples only involve prisoners who have been released early, while the test examples are on all new prisoners. The decision of early releasing may already depends on information encoded in the input vector x, and therefore causes change in μ. For example, if the feature vector x includes the length of initial sentence, and the prisoners are more likely to be early released if the sentence period is short, then it is also likely that $\mu_{train}(x) > \mu_{test}(x)$ for x with shorter sentence period. In practice, it is rare that the early release decision is independent of x, and thus μ is likely to change.

The conditional distribution of $y|x$, η, may also be different. The composition of prisoners with the same x may be different between training and test data as μ changes, and such difference can cause selection bias. For example, prisoners with long initial sentence period may need exceptionally positive records to be released early and also may treasure this opportunity more than average. As a result, $\Pr(\text{reoffend}|\text{long initial sentence period})$ could be low among the prisoners who are early released. However, the conditional reoffend probability may increase if the distribution now includes prisoners who have less positive records and do not meet the original early release criteria.

Another important factor here is time. The training data represent cases in a past time period, while the test is on new prisoners. A lot of factors, such as percentage of different types of crime and average sentence period length, can change over the period of time. In general, μ and η are different to various extent across different time periods.