DSC 140A - Quiz 04

Name:
PID:

Quiz 04 Instructions. All technology is prohibited. Pen and paper only. You may not discuss with any classmates. Any prohibited use of notes or technology will result in a 0 and is a violation of Academic Integrity. You will have 50 minutes to complete your quiz. If you have any questions, please raise your hand and a TA will come assist you. Make sure to write your name and PID at the top of this page as clearly as possible. Explain/show your work.

Notation: Bold upper case letters $\mathbf{A}$ represent matrices.
Bold lower case letters $\mathbf{a}$ represent vectors.
Non-bolded lower case letters a represent scalars.
For a matrix $\mathbf{A}$, $\mathbf{a}_i$ indicates the ith row vector of matrix $\mathbf{A}$.

Problem 1. (1 point)

Suppose the Bayes classifier achieves an error rate of 15% on a particular data distribution. True or False: It is impossible for any classifier trained on data drawn from this distribution to achieve better than 85% accuracy on a finite test set that is drawn from this distribution.

○ True

● False

> **Solution:** False. Baye's error/accuracy is the best in **expectation**. For any finite set of data, there is a non-zero probability that the test set is a group of samples that is more optimal for your given classifier than the Bayes rate suggests.

Problem 2. (2 points)

a) Suppose a particular probability distribution has the property that, whenever data are sampled from the distribution, the sampled data are guaranteed to be linearly separable. True or False: the Bayes error with respect to this distribution is 0%.

● True

○ False

> **Solution:** True.

b) Now consider a different probability distribution. Suppose the Bayes classifier achieves an error rate of 0% on this distribution. True or False: given a finite data set sampled from this distribution, the data must be linearly separable.

○ True

● False

> **Solution:** False.

Problem 3. (1 point)

Let $(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)$ be a set of n points in a binary classification problem, where $\vec{x}_i \in \mathbb{R}^2$ and $y_i \in \{0, 1\}$.

Suppose a classifier is trained by estimating the class-conditional densities with histograms using rectangular bins and applying the Bayes classification rule.

True or False: it is always possible to achieve a 100% training accuracy with this classifier by choosing the rectangular bins to be sufficiently small. You may assume that no two points $\vec{x}_i$ and $\vec{x}_j$ are identical.

● True

○ False

> **Solution:** True.
>
> Video explanation: https://youtu.be/A1fBjOnjs5E

Problem 4. (1 point)

Suppose data points $\vec{x}_1, \ldots, \vec{x}_n$ are drawn from an arbitrary, unknown distribution with density f.

True or False: it is guaranteed that, given enough data (that is, n large enough), a Gaussian fit to the data using the method of maximum likelihood will approximate the true underlying density f arbitrarily closely.

- ○ True
- ● False

Solution: False.

Problem 5.

A data set contains the following observed pairs (x, y), where Y is binary.

Observation	1	2	3	4	5	6	7	8
x	a	a	b	b	c	c	d	d
y	0	1	0	0	1	1	0	1

Treat this data set as the full population distribution, where each observation is equally likely.

a) (1 point) Compute $\mathbb{P}(Y = 0)$ and $\mathbb{P}(Y = 1)$.

Solution:

$$\mathbb{P}(Y = 0) = \frac{4}{8} = \frac{1}{2}, \qquad \mathbb{P}(Y = 1) = \frac{4}{8} = \frac{1}{2}.$$

b) (1 point) Compute $\mathbb{P}(Y = 1 \mid X = x)$ for each $x \in \{a, b, c, d\}$.

Solution:

x	a	b	c	d
$\mathbb{P}(Y = 1 \mid X = x)$	1/2	0	1	1/2

c) (1 point) Suppose instead that the classifier is only allowed to use the feature

$$Z = \begin{cases} 0, & X \in \{a, b\} \\ 1, & X \in \{c, d\}. \end{cases}$$

For which values of z does the Bayes classifier predict $Y = 1$?

Solution: If $Z = 0$, then $X \in \{a, b\}$. The corresponding labels are

$$0, 1, 0, 0.$$

Hence,

$$\mathbb{P}(Y = 1 \mid Z = 0) = \frac{1}{4}.$$

So the Bayes classifier predicts $Y = 0$ when $Z = 0$.

If $Z = 1$, then $X \in \{c, d\}$. The corresponding labels are

$$1, 1, 0, 1.$$

Hence,

$$\mathbb{P}(Y = 1 \mid Z = 1) = \frac{3}{4}.$$

So the Bayes classifier predicts $Y = 1$ when $Z = 1$.

d) (2 point) Is the Bayes error rate using Z larger than, smaller than, or equal to the Bayes error rate using X? Show why.

Solution: Using X, the Bayes classifier predicts the majority label for each value of X. We have

$$P(Y = 1 \mid X = a) = \frac{1}{2}, \qquad P(Y = 1 \mid X = b) = 0, \qquad P(Y = 1 \mid X = c) = 1, \qquad P(Y = 1 \mid X = d) = \frac{1}{2}.$$

Thus, the only unavoidable errors occur when $X = a$ and when $X = d$. Since

$$P(X = a) = \frac{2}{8} \quad \text{and} \quad P(X = d) = \frac{2}{8},$$

the Bayes error using X is

$$\frac{2}{8} \cdot \frac{1}{2} + \frac{2}{8} \cdot \frac{1}{2} = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Now using Z, if $Z = 0$, then $X \in \{a, b\}$, so the corresponding labels are

$$0, 1, 0, 0.$$

Therefore,

$$P(Y = 1 \mid Z = 0) = \frac{1}{4},$$

so the Bayes classifier predicts $Y = 0$ when $Z = 0$, making an error with probability $\frac{1}{4}$.

If $Z = 1$, then $X \in \{c, d\}$, so the corresponding labels are

$$1, 1, 0, 1.$$

Therefore,

$$P(Y = 1 \mid Z = 1) = \frac{3}{4},$$

so the Bayes classifier predicts $Y = 1$ when $Z = 1$, making an error with probability $\frac{1}{4}$.

Since

$$P(Z = 0) = \frac{1}{2} \quad \text{and} \quad P(Z = 1) = \frac{1}{2},$$

the Bayes error using Z is

$$\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Therefore, the Bayes error rate using Z is equal to the Bayes error rate using X:

$$\boxed{\text{Bayes error using } Z = \text{Bayes error using } X = \frac{1}{4}.}$$

Although replacing X with Z loses information, in this particular example the loss of information does not increase the Bayes error.