Homework 2

Problem 1. Let $p(x|w_i) \sim \mathcal{N}(\mu_i, \sigma^2)$ for a two-category one-dimensional problem with $P(\omega_1) = P(\omega_2) = \frac{1}{2}$.

(a) Show that the minimum probability of error is given by

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-u^2/2} du$$

Where $a = \frac{|\mu_1 - \mu_2|}{2\sigma}$.

(b) Use the inequality

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-t^2/2} dt \leq \frac{1}{\sqrt{2\pi} a} e^{-a^2/2}$$

to show that P_e goes to zero as $\frac{|\mu_1 - \mu_2|}{\sigma}$ goes to infinity.

Problem 2. Consider a two-category classification problem in two dimensions with

$$p(\boldsymbol{x}|\omega_1) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}), p(\boldsymbol{x}|\omega_2) \sim \mathcal{N}(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \boldsymbol{I})$$

and $P(\omega_1) = P(\omega_2) = \frac{1}{2}$,

(a) Calculate the Bayes decision boundary.

(b) Calculate the Bhattacharyya error bound.

(c) Repeat the above for the same prior probabilities, but

$$p(\boldsymbol{x}|\omega_1) \sim \mathcal{N}(\mathbf{0}, \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}), p(\boldsymbol{x}|\omega_2) \sim \mathcal{N}(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix})$$

Problem 3. Suppose that we have three categories in two dimensions with the following underlying distributions:

- $p(\boldsymbol{x}|\omega_1) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$
- $p(\boldsymbol{x}|\omega_2) \sim \mathcal{N}(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \boldsymbol{I})$
- $p(\boldsymbol{x}|\omega_3) \sim \frac{1}{2}\mathcal{N}(\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \boldsymbol{I}) + \frac{1}{2}\mathcal{N}(\begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}, \boldsymbol{I})$

with $P(\omega_i) = \frac{1}{3}$, $i = 1, 2, 3$.

(a) By explicit calculation of posterior probabilities, classify the point $\boldsymbol{x} = \begin{pmatrix} 0.3 \\ 0.3 \end{pmatrix}$ for minimum probability of error.

(b) Suppose that for a particular test point the first feature is missing. That is, classify $\boldsymbol{x} = \begin{pmatrix} * \\ 0.3 \end{pmatrix}$.

(c) Suppose that for a particular test point the second feature is missing. That is, classify $\boldsymbol{x} = \begin{pmatrix} 0.3 \\ * \end{pmatrix}$.

(d) Repeat all of the above for $\boldsymbol{x} = \begin{pmatrix} 0.2 \\ 0.6 \end{pmatrix}$.

Problem 4. Consider two normal distributions with arbitrary but equal covariances. Prove that the Fisher linear discriminant, for suitable threshold, can be derived from the negative of the log-likelihood ratio.

Problem 5. It is easy to see that the nearest-neighbor error rate P can equal the Bayes rate P^* if $P^* = 0$ (the best possibility) or if $P^* = \frac{c-1}{c}$ (the worst possibility). One might ask whether or not there are problems for which $P = P^*$ when P^* is between these extremes.

(a) Show that the Bayes rate for the one-dimensional case where $P(\omega_i) = \frac{1}{c}$ and

$$p(x|\omega_i) = \begin{cases} 1, & 0 \le x \le \frac{cr}{c-1} \\ 1, & i \le x \le i + 1 - \frac{cr}{c-1} \\ 0, & \text{elsewhere} \end{cases} \tag{1}$$

is $P^* = r$.

(b) Show that for this case the nearest-neighbor rate is $P = P^*$.

Problem 6. Prove that the computational complexity of the basic nearest-neighbor editing algorithm for n points in d dimension is $O(d^3 n^{\lfloor \frac{d}{2} \rfloor} \ln n)$.

Nearest-Neighbor Editing Algorithm

```
begin initialize j ← 0, D ← data set, n ← #prototypes
    construct the full Voronoi diagram of D
    do j ← j + 1, for each prototype x'_j
        find the Voronoi neighbors of x'_j
        if any neighbor is not from the same class as x'_j, then mark x'_j
    until j = n
    discard all points that are not marked
    construct the Voronoi diagram of the remaining (marked) prototypes
end
```

Problem 7. Consider classifiers based on samples with priors $P(\omega_1) = P(\omega_2) = 0.5$ and the distributions

$$p(x|\omega_1) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases} \tag{2}$$

and

$$p(x|\omega_2) = \begin{cases} 2 - 2x, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases} \tag{3}$$

(a) What is the Bayes decision rule and the Bayes classification error?

(b) Suppose we randomly select a single point ω_1 and a single point from ω_2, and create a nearest-neighbor classifier. Suppose too we select a test point from one of the categories (ω_1 for definiteness). Integrate to find the expected error rate $P_2(e)$.

(c) Repeat with two training samples from each category and a single test point in order to find $P_2(e)$.

(d) Generalize to show that in general,

$$P_n(e) = \frac{1}{3} + \frac{1}{(n+1)(n+3)} + \frac{1}{2(n+2)(n+3)}$$

Confirm this formula makes sense in the $n = 1$ case.

(e) Compare $\lim_{n\to\infty} P_n(e)$ with the Bayes error.

———