EE E6887 (Statistical Pattern Recognition)
Solutions for homework 4

P.1 In this problem, we would like to get familiar with the procedure of computing the error probability of 1-nearest neighbor. Consider data samples from the following two distributions. Assume the two classes have equal priors, i.e., $P(\omega_1) = P(\omega_2) = 0.5$

$$p(x|\omega_1) = \begin{cases} 2x & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad p(x|\omega_2) = \begin{cases} 2(1-x) & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Derive the Bayesian decision rule and its probability of classification

(b) Suppose we have one single training sample from class ω_1 and one single training sample from class ω_2. Now given a randomly selected test sample, we would like to use 1-nearest neighbor classifier to classify the test data. What is the probability of classification error of such 1-NN classifier?

Answer:

(a) since $p(\omega_1) = p(\omega_2) = 0.5$, the discriminant function turns to:

$$\begin{aligned} g_1(x) &= p(x|\omega_1) \\ g_2(x) &= p(x|\omega_2) \end{aligned}$$

When $g_1(x) > g_2(x)$, we classify x to ω_1, when $g_1(x) < g_2(x)$, we classify x to ω_2. That is:

$$\begin{aligned} x \in \omega_1, &\qquad \text{when } \frac{1}{2} < x \le 1 \\ x \in \omega_2, &\qquad \text{when } 0 \le x < \frac{1}{2} \end{aligned}$$

In such case, the classification error is given by:

$$\begin{aligned} P^*(e) &= \int_0^1 \min[p(\omega_1|x), p(\omega_2|x)]p(x)dx \\ &= \frac{1}{2}\int_0^{1/2} 2x dx + \frac{1}{2}\int_{1/2}^1 2(1-x)dx = \frac{1}{4} \end{aligned}$$

(b) Suppose x_1 and x_2 are the training samples from ω_1 and ω_2 respectively. For a given test image x, We classify x to ω_1 if $|x-x_1| < |x-x_2|$, and to ω_2 otherwise.

Therefore the probability of error is given by:

$$P(e) = \int_{x_1}\int_{x_2}\int_{x} p(e|x_1,x_2,x)p(x_1,x_2,x)dx_1dx_2dx$$

An error occurs when

1. $|x-x_1| < |x-x_2|$, if $x \in \omega_2$
2. $|x-x_1| > |x-x_2|$, if $x \in \omega_1$

This can be further broken up into 4 cases below:

1. $x < \frac{x_1+x_2}{2}$, if $x \in \omega_2$ and $x_2 > x_1$
2. $x > \frac{x_1+x_2}{2}$, if $x \in \omega_2$ and $x_2 < x_1$
3. $x > \frac{x_1+x_2}{2}$, if $x \in \omega_1$ and $x_2 > x_1$
4. $x < \frac{x_1+x_2}{2}$, if $x \in \omega_1$ and $x_2 < x_1$

In probability, the above idea is expressed as below. Denote x_{ti} as the test sample from ω_i:

$$\begin{aligned}
P(e) &= \int_{x_1}\int_{x_2}\int_{x_{t2}} p(e|x_1,x_2,x)p(x_1,x_2,x)dx_1dx_2dx \\
&= \int_{x_1}\int_{x_2}\int_{x_{t1}} p(|x_{t2}-x_1| < |x_{t2}-x_2||x_1,x_2,x_{t2})p(x_1,x_2,x_{t2})p(\omega_2)dx_1dx_2dx_{t2} \\
&+ \int_{x_1}\int_{x_2}\int_{x} p(|x_{t1}-x_1| > |x_{t1}-x_2||x_1,x_2,x_{t1})p(x_1,x_2,x_{t1})p(\omega_1)dx_1dx_2dx_{t1} \\
&= p(\omega_2)\int_{x_1}\int_{x_2}\int_{x_{t2}} p(|x_{t2}-x_1| < |x_{t2}-x_2||x_1,x_2,x_{t2})p(x_1,x_2,x_{t2})dx_1dx_2dx_{t2} \\
&+ p(\omega_1)\int_{x_1}\int_{x_2}\int_{x_{t1}} p(|x_{t1}-x_1| > |x_{t1}-x_2||x_1,x_2,x_{t1})p(x_1,x_2,x)dx_1dx_2dx_{t1} \\
&= p(\omega_2)I_1 + p(\omega_1)I_2
\end{aligned}$$

By symmetry, the first integral is equal to the second integral, i.e., $I_1 = I_2$, therefore we only need to evaluate the first one.

$$
\begin{aligned}
I_1 &= \int_{x_1}\int_{x_2}\int_{x_{t2}} p(|x_{t2}-x_1| < |x_{t2}-x_2||x_1,x_2,x_{t2})p(x_1,x_2,x)dx_1dx_2dx_{t2} \\
&= \int_{x_1}\int_{x_2}\int_{x_{t2}} p(x_{t2} < \frac{x_1+x_2}{2}|x_1,x_2,x_{t2},x_2>x_1)p(x_2>x_1|x_1,x_2)p(x_1,x_2,x_{t2}) \\
&+ p(x_{t2} > \frac{x_1+x_2}{2}|x_1,x_2,x_{t2},x_2<x_1)p(x_2<x_1|x_1,x_2)p(x_1,x_2,x_{t2})dx_1dx_2dx_{t2}
\end{aligned}
$$

Observe that,

$p(x < \frac{x_1+x_2}{2}|x_1,x_2,x,x\in\omega_2,x_2>x_1) = 1$ if $x < \frac{x_1+x_2}{2}$, and it is equal to 0 otherwise.

Similarly,

$p(x_2 > x_1|x_1,x_2) = 1$ if $x_2 > x_1$, and it is equal to 0 otherwise.

Furthermore, as x_1, x_2, x_{ti} are independent, so $p(x_1,x_2,x_{t2}) = p(x_1|\omega_1)p(x_2|\omega_2)p(x_{ti}|\omega_i)$.

Therefore,

$$
\begin{aligned}
I_1 &= \int_{x_1=0}^{1} dx_1 p(x_1|\omega_1)\int_{x_2=x1}^{1} dx_2 p(x_2|\omega_2)\int_{x_{t2}=0}^{\frac{x_1+x_2}{2}} dx_{t2}p(x_{t2}|\omega_2) \\
&+ \int_{x_2=0}^{1} dx_2 p(x_2|\omega_2)\int_{x_1=x2}^{1} dx_1 p(x_1|\omega_1)\int_{x_{t2}=\frac{x_1+x_2}{2}}^{1} dx_{t2}p(x_{t2}|\omega_2) \\
&= 0.35 = I_2
\end{aligned}
$$

Since $P(\omega_1) = P(\omega_2) = 0.5$,

$$P(e) = p(\omega_2)I_1 + p(\omega_1)I_2 = 0.35$$

P.2 Computing distances in a high-dimensional feature space sometimes could be costly prohibitive. One popular trick is to compute a certain distance in a lower dimension space as a pre-filtering step.

Assume $\vec{x} = \{x_1, x_2, \ldots, x_d\}$ and $\vec{y} = \{y_1, y_2, \ldots, y_d\}$ are two feature vectors in a d-dimensional space. Prove that

$$\left\{\frac{1}{\sqrt{d}}\sum_{i=1}^{d} x_i - \frac{1}{\sqrt{d}}\sum_{i=1}^{d} y_i\right\}^2 \leq \sum_{i=1}^{d}(x_i - y_i)^2$$

Namely the distance between the scaled means of two vectors is less than their L_2 distance. Discuss how we may use this property to reduce the computational complexity of the process of finding the nearest neighbor point.

Answer:

Let $z_i = x_i - y_i$,

$$\left(\frac{1}{\sqrt{d}}\sum_{i=1}^{d} x_i - \frac{1}{\sqrt{d}}\sum_{i=1}^{d} y_i\right)^2$$
$$= d\left(\frac{1}{d}\sum_{i=1}^{d} z_i\right)^2$$

Let z be a random variable with $z \in \{z_i | i = 1, \ldots, d\}$ and $p(z_i) = \frac{1}{d}, \sum_{i=1}^{d} p(z_i) = 1$. By Jensen's inequality, we have $f(E[z]) \leq E[f(z)]$ for a convex function f. As $f(x) = x^2$ is a convex function. Therefore,

$$d\left(\frac{1}{d}\sum_{i=1}^{d} z_i\right)^2 = d\left(\sum_{i=1}^{d}\frac{1}{d} z_i\right)^2$$
$$\leq d\sum_{i=1}^{d}\frac{1}{d} z_i^2 = \sum_{i=1}^{d} z_i^2$$

To reduce computational complexity for the nearest neighbor classifiers, we can pre-compute the scaled mean of the training data. Then, when given a test data, we first compute its scaled mean and then compute its distance to the pre-computed training data. As the distance function is 1d instead of the original dimension, there is a reduced computational complexity.