Problem 10, Page 67

Consider the following decision rule for a two-category one-dimensional probelm: Decide ω_1 if $x > \theta$; otherwise decide ω_2.

1. Show that the probobility of error for this rule is given by

   $P(error) = P(\omega_1) \int_{-\infty}^{\theta} p(x \mid \omega_1) dx + P(\omega_2) \int_{\theta}^{\infty} p(x \mid \omega_2) dx$

2. By differentiating, show that a necessary condition to minimize $P(error)$ is that θ satisfies

$$p(\theta \mid \omega_1) P(\omega_1) = p(\theta \mid \omega_2) P(\omega_2)$$

3. Does this equation define θ uniquely?
4. Give an example where a value of θ satisfying the equation actually maximizes the probability of error.

Problem 12, Page 68

Let $\omega_{max}(x)$ be the state of nature for which

$$P(\omega_{max} \mid x) \geq P(\omega_i \mid x) \ \text{ for all } \ i, \ \ i = 1, \ldots, c.$$

1. Show that $P(\omega_{max} \mid x) \geq 1/c$.
2. Show that for the minimum-error-rate decision rule the average probability of error is given by
$$P(error) = 1 - \int P(\omega_{max} \mid x) p(x) dx.$$
3. Use these two results to show that
$$P(error) \leq (c-1)/c.$$
4. Describe a situation for which $P(error) = (c-1)/c$.

We will see –

1. How to apply Baye's rule in practice?

2. How to atleast approximately find class conditionals $p(x|\omega_i)$?

3. Finally, this leads to a very simple classifier called k nearest neighbor classifier.

How to find class conditionals $p(X|\omega_i)$?

Let $\Omega = \{\omega_1, \ldots, \omega_c\}$ be the set of classes.

Let $\mathcal{D} = \{(X_1, y_1), \ldots, (X_n, y_n)\}$ be the training set.

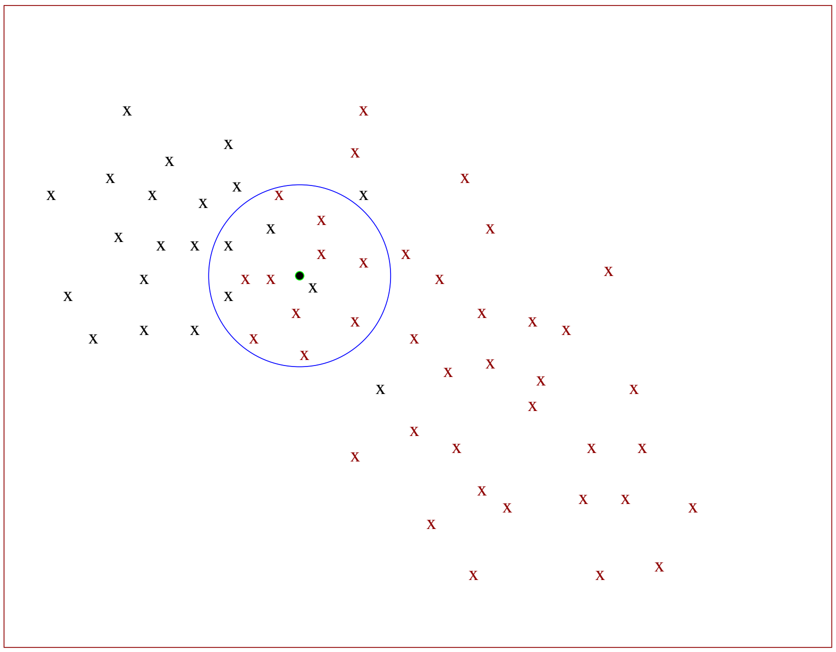
Here $y_i \in \Omega$ is the class label for pattern X_i.

To find $p(X|\omega_i)$, do the following:

1. Draw a a small hyper-sphere at X. Let its volume be V.
2. Count the number of training patterns which belongs to class ω_i that are falling in the sphere. Let this be k_i.
3. Let n_i be the total number of training patterns that belongs to class ω_i.
4. Then an estimate of $p(X|\omega_i)$ which is $\hat{p}(X|\omega_i)$ is:

$$\hat{p}(X|\omega_i) = \frac{k_i/n_i}{V}$$

Consider a two class problem. Classes are *black* and *red*



Let X be the centre of the sphere whose volume is 2 units.

$\hat{p}(X|$ class is *black* $) = 4/(20 * 2) = 1/10.$

$\hat{p}(X|$ class is *red* $) = 10/(40 * 2) = 1/8.$

How to find $\hat{p}(X)$? In this example what is its value?

An estimate of priors would be:

$$\hat{P}(\omega_i) = \frac{n_i}{n}$$

Posterior

$$\hat{P}(\omega_i|X) = \frac{\frac{k_i}{n_i} \times \frac{n_i}{n}}{p(X)} = \frac{k_i/n}{p(X)}$$

Predicted class is according to

$$\text{max. in } \{\hat{P}(\omega_1|X), \ldots, \hat{P}(\omega_c|X)\}$$

$$= \text{max. in } \{\tfrac{k_1/n}{p(X)}, \ldots, \tfrac{k_c/n}{p(X)}\}$$

$$= \text{max. in } \{k_1, \ldots, k_c\}$$

That is, count the number of patterns for each class that are falling in the sphere, and decide the class for which the count is maximum.

Still one problem is: how to decide the radius for the sphere?

Instead of this, for a given value k, we can draw a sphere which is just big enough to hold k nearest neighbors of X.

This classifier is called the *k-nearest neighbor classifier (k-NNC)*.

For a given test pattern X which is to be classified, do:

1. Find k nearest neighbors of X.
2. Among these k nearest neighbors, according to majority voting decide the predicted class label.

k-NNC is an approximation of Bayes classifier!

Mathematically it can be shown that when $n \to \infty$, $k \to \infty$ and $k/n \to 0$, k-NNC is exactly the Bayes classifier.

When $k = 1$, k-NNC is simply called the *nearest neighbor classifier (NNC).*

In our example with *black* and *red* classes with $k = 14$ what is the prediction?

Similarly for NNC what is the prediction?

Still there is one problem with k-NNC

That is, how to decide the value of k ?

For this, there is a practical method called *cross validation.*

One more issue is : *How to find the distances?*

There are many distance functions like Euclidean distance.

Cross Validation