MODULE 10
Bayes Classifier

LESSON 18
Bayes Classifier

Keywords: Bayes, Minimum Error-Rate, Classifier

Classification using Bayes Decision Theory

- In this approach classification is carried out using probabilities of classes.
- It is assumed that we know the *a priori* or the prior probability of each class. If we have two classes C_1 and C_2, then the prior probability for class C_1 is P_{C1} and the prior probability for class C_2 is P_{C2}.
- If the prior probability is not known, the classes are taken to be equally likely.
- If prior probability is not known and there are two classes C_1 and V_2, then it is assumed $P_{C1} = P_{C2} = 0.5$.
- If P_{C1} and P_{C2} are known, then when a new pattern x comes along, we need to calculate $P(C_1|x)$ and $P(C_2|x)$.
- The bayes theorem is used to compute $P(C_1|x)$ and $P(C_2|x)$.
- Then if $P(C_1|x) \geq P(C_2|x)$, the pattern is assigned to Class 1 and if $P(C_1|x) < P(C_2|x)$, it is assigned to Class 2. This is called the Bayes decision rule.

Bayes Rule

- If $P(C_i)$ is the prior probability of Class i, and $P(X|C_i)$ is the conditional density of X given class C_i, then the *a posteriori* or posterior probability of C_i is given by

  $P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$

- Bayes theorem provides a way of calculating the posterior probability $P(C_i \mid X)$. In other words, after observing X, the posterior probability that the class is c_i can be calculated from Bayes theorem. It is useful to convert prior probabilities to posterior probabilities.
- $P(X)$ is given by

  $P(X) = \sum_i P(X \mid C_i)P(C_i)$

- Let the probability that an elephant is black be 80% and that an elephant is white be 20%. This means P(elephant is black) = 0.8 and P(elephant is white) =0.2 . With only this information, any elephant will be classified as being black. This is because the probability of error in this case is only 0.2 as opposed to classifying the elephant as white which results in a probability of error of 0.8 . When additional information is available, it can be used along with the information above.

  If we have probability that elephant belongs to region X is 0.2 . Now if the elephant belongs to region X, we need to calculate the posterior probability that the elephant is white. i.e. P(elephant is white | elephant belongs to region) or $P(W \mid X)$. This can be calculated by using Bayes theorem. If 95% of the time when the elephant is white, it is because it belongs to the region X. Then

  $P(W \mid X) = \frac{P(X|W)*P(W)}{P(X)} = \frac{0.95*0.2}{0.2} = 0.95$

  The probability of error is 0.05 which is the probability that the elephant is not white given that it belongs to region X.

Minimum Error Rate Classifier

- If it is required to classify a pattern X, then the minimum error rate classifier classifies the pattern X to the class C which has the maximum posterior probability $P(C \mid X)$.
- If the test pattern X is classified as belonging to Class C, then the error in the classification will be (1 - $P(C \mid X)$).
- It is evident to reduce the error, X has to be classified as belonging to the class for which $P(C \mid X)$ is maximum.
- The expected probability of error is given by

$$\int_X (1 - P(C|X))P(X)dX$$

This is minimum when $P(C \mid X)$ is maximum (for a specified value of $P(X)$.

- Let us consider an example of how to use minimum error rate classifier for a classification problem. Let us consider an example with three classes *small*, *medium* and *large* with prior probability

  $P(small) = \frac{1}{3}$
  $P(medium) = \frac{1}{2}$
  $P(large) = \frac{1}{6}$

  We have a set of nails, bolts and rivets in a box and the three classes correspond to the size of these objects in the box.

  Now let us consider the class-conditional probabilities of these objects :

  For Class *small* we get

  $P(nail \mid small) = \frac{1}{4}$
  $P(bolt \mid small) = \frac{1}{2}$
  $P(rivet \mid small) = \frac{1}{4}$

  For Class *medium* we get

  $P(nail \mid medium) = \frac{1}{2}$
  $P(bolt \mid medium) = \frac{1}{6}$
  $P(rivet \mid medium) = \frac{1}{3}$

  For Class *large* we get

  $P(nail \mid large) = \frac{1}{3}$
  $P(bolt \mid large) = \frac{1}{3}$

$P(rivet \mid large) = \frac{1}{3}$

Now we can find the probability of the class labels given that it is a nail, bolt or rivet. For doing this we need to use Bayes Classifier. Once we get these probabilities, we can find the corresponding class labels of the objects.

$$P(small \mid nail) = \frac{P(nail)|small)P(small)}{P(nail|small).P(small)+P(nail|medium).P(medium)+P(nail|large).P(large)}$$

This will give

$$P(small \mid nail) = \frac{\frac{1}{4}\cdot\frac{1}{3}}{\frac{1}{4}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{1}{6}} = 0.2143$$

Similarly, we calculate $P(medium \mid nail)$ and we get

$$P(medium \mid nail) = \frac{\frac{1}{2}.frac12}{\frac{1}{4}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{1}{6}} = 0.6429$$

and also $P(large \mid nail)$

$$P(large \mid nail) = \frac{\frac{1}{3}\cdot\frac{1}{6}}{\frac{1}{4}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{1}{6}} = 0.1429$$

Since $P(medium \mid nail) > P(small \mid nail)$ and
$P(medium \mid nail) > P(large \mid nail)$

we classify nail as belonging to the class *medium*. The probability of error $P(error \mid nail) = 1 - 0.6429 = 0.3571$

In a similar way, we can find the posterior probability for bolt

$$P(small \mid bolt) = \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{6}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{1}{6}} = 0.5455$$

$$P(medium \mid bolt) = \frac{\frac{1}{6}\cdot\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{6}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{1}{6}} = 0.2727$$

$P(large \mid bolt) = \frac{\frac{1}{3} \cdot \frac{1}{6}}{\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{6} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{6}} = 0.1818$

Since $P(small \mid bolt) > P(medium \mid bolt)$ and
$P(small \mid bolt) > P(large \mid bolt)$

we classify bolt as belonging to the class *small* and the probability of error $P(error \mid bolt) = 1 - 0.5455 = 0.4545$

In a similar way, we can find the posterior probability for rivet

$P(small \mid rivet) = \frac{\frac{1}{4} \cdot \frac{1}{3}}{\frac{1}{4} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{6}} = 0.2727$

$P(medium \mid rivet) = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{4} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{6}} = 0.5455$

$P(large \mid rivet) = \frac{\frac{1}{3} \cdot \frac{1}{6}}{\frac{1}{4} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{6}} = 0.1818$

Since $P(medium \mid rivet) > P(small \mid rivet)$ and
$P(medium \mid rivet) > P(large \mid rivet)$

we classify bolt as belonging to the class *medium* and the probability of error $P(error \mid rivet) = 1 - 0.5455 = 0.4545$