Recap

- We have gone through a general overview of pattern classification.

Recap

- We have gone through a general overview of pattern classification.
- The Bayes classifier minimizes *risk*.
- If we have complete knowledge of the underlying probability distributions, then Bayes classifier is optimal (for minimizing risk).

Recap

- We have gone through a general overview of pattern classification.
- The Bayes classifier minimizes *risk*.
- If we have complete knowledge of the underlying probability distributions, then Bayes classifier is optimal (for minimizing risk).
- There are other classifiers possible (e.g., nearest neighbour classifier)

Recap

- Another important form for a classifier is

$$h(\mathbf{X}) = 0 \quad \text{if} \quad g(\mathbf{W}, \mathbf{X}) > 0$$

g is called Discriminant function.

Recap

- Another important form for a classifier is

$$h(\mathbf{X}) = 0 \quad \text{if} \quad g(\mathbf{W}, \mathbf{X}) > 0$$

  g is called Discriminant function.
- Important special case is linear discriminant functions $h(\mathbf{X}) = \text{sgn}(\mathbf{W}^T\mathbf{X})$

Recap

- Another important form for a classifier is

$$h(\mathbf{X}) = 0 \quad \text{if} \quad g(\mathbf{W}, \mathbf{X}) > 0$$

  g is called Discriminant function.
- Important special case is linear discriminant functions $h(\mathbf{X}) = \text{sgn}(\mathbf{W}^T\mathbf{X})$
- There are different approaches to learn *nonlinear* classifiers.

Bayes Classifier

- In this class we will derive the Bayes classifier for M classes under a general loss function.
- This can actually be looked at as a special case of a more general problem of decision making under uncertainity.

Bayesian Decision Making

- The task: Decision making under uncertainity

Bayesian Decision Making

- The task: Decision making under uncertainity
- We want to decide on one of finitely many ‘actions’ based on some observation.
- Our ‘payoff’ or ‘cost’ depends on the *unknown* ‘state of nature’ and the observation gives some (stochastic) information on the ‘state of nature’.

Bayesian Decision Making

- The task: Decision making under uncertainity
- We want to decide on one of finitely many ‘actions’ based on some observation.
- Our ‘payoff’ or ‘cost’ depends on the *unknown* ‘state of nature’ and the observation gives some (stochastic) information on the ‘state of nature’.
- A Loss function gives ‘costs’ for each decision for every ‘true’ state of nature.
- We want a strategy of decision making that minimizes, e.g., expected loss.

In the context of classifier design

- Observation is the feature vector.
- The ‘state of nature’ is the ‘true’ class label of the feature vector.
- We need to decide on a class label based on the observation.

Bayes Classifier

- We now consider the Bayes classifier with M classes and arbitrary loss function.

Bayes Classifier

- We now consider the Bayes classifier with M classes and arbitrary loss function.
- $C_0, C_1, \cdots, C_{M-1}$ – the class labels.
  $y(\mathbf{X}) \in \{C_0, \cdots, C_{M-1}\}$.

Bayes Classifier

- We now consider the Bayes classifier with M classes and arbitrary loss function.
- $C_0, C_1, \cdots, C_{M-1}$ – the class labels.
  $y(\mathbf{X}) \in \{C_0, \cdots, C_{M-1}\}$.
- Let $h(\mathbf{X}) \in \{\alpha_0, \alpha_1, \cdots, \alpha_{K-1}\}$.
  The output of classifier would be α_j's.

Bayes Classifier

- We now consider the Bayes classifier with M classes and arbitrary loss function.
- $C_0, C_1, \cdots, C_{M-1}$ – the class labels.
  $y(\mathbf{X}) \in \{C_0, \cdots, C_{M-1}\}$.
  (States of Nature)
- Let $h(\mathbf{X}) \in \{\alpha_0, \alpha_1, \cdots, \alpha_{K-1}\}$.
  The output of classifier would be α_j's.
  (Actions of decision maker)

- In general, we may have $M \neq K$.
  The classifier output need not always be a class label.
- For example, we can have $K = M + 1$ and α_M may denote the decision of 'rejection'.
- We take $K = M$ unless specified otherwise.

- $L(\alpha_j, C_k)$ – loss when classifier says α_j and ‘true class’ is C_k. We assume that loss function is non-negative.
- Notation makes it easy to understand arguments of loss function.

- $L(\alpha_j, C_k)$ – loss when classifier says α_j and ‘true class’ is C_k. We assume that loss function is non-negative.
- Notation makes it easy to understand arguments of loss function.
- As earlier, the risk of a classifier h is

$$R(h) = EL(h(\mathbf{X}), y(\mathbf{X}))$$

- We want the classifier that has the least risk value.

- Given a $\mathbf{X}$, let $R(\alpha_i \mid \mathbf{X})$ denote the expected loss when classifier says α_i and conditioned on $\mathbf{X}$.

- Given a $\mathbf{X}$, let $R(\alpha_i \mid \mathbf{X})$ denote the expected loss when classifier says α_i and conditioned on $\mathbf{X}$.

$$R(\alpha_i \mid \mathbf{X}) \;=\; E\left[L(h(\mathbf{X}), y(\mathbf{X})) \mid h(\mathbf{X}) = \alpha_i, \mathbf{X}\right]$$

- Given a $\mathbf{X}$, let $R(\alpha_i \mid \mathbf{X})$ denote the expected loss when classifier says α_i and conditioned on $\mathbf{X}$.

$$\begin{aligned} R(\alpha_i \mid \mathbf{X}) &= E\left[L(h(\mathbf{X}), y(\mathbf{X})) \mid h(\mathbf{X}) = \alpha_i, \mathbf{X}\right] \\ &= E\left[L(\alpha_i, y(\mathbf{X})) \mid \mathbf{X}\right] \end{aligned}$$

- Given a $\mathbf{X}$, let $R(\alpha_i \mid \mathbf{X})$ denote the expected loss when classifier says α_i and conditioned on $\mathbf{X}$.

$$
\begin{aligned}
R(\alpha_i \mid \mathbf{X}) &= E\left[L(h(\mathbf{X}), y(\mathbf{X})) \mid h(\mathbf{X}) = \alpha_i, \mathbf{X}\right] \\
&= E\left[L(\alpha_i, y(\mathbf{X})) \mid \mathbf{X}\right] \\
&= \sum_{j=0}^{M-1} L(\alpha_i, C_j)\text{Prob}(y(\mathbf{X}) = C_j \mid \mathbf{X})
\end{aligned}
$$

- Given a $\mathbf{X}$, let $R(\alpha_i \mid \mathbf{X})$ denote the expected loss when classifier says α_i and conditioned on $\mathbf{X}$.

$$
\begin{aligned}
R(\alpha_i \mid \mathbf{X}) &= E\left[L(h(\mathbf{X}), y(\mathbf{X})) \mid h(\mathbf{X}) = \alpha_i, \mathbf{X}\right] \\
&= E\left[L(\alpha_i, y(\mathbf{X})) \mid \mathbf{X}\right] \\
&= \sum_{j=0}^{M-1} L(\alpha_i, C_j) \text{Prob}(y(\mathbf{X}) = C_j \mid \mathbf{X}) \\
&= \sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X})
\end{aligned}
$$

- We saw

$$\begin{aligned} R(\alpha_i \mid \mathbf{X}) &= E\left[L(h(\mathbf{X}), y(\mathbf{X})) \mid h(\mathbf{X}) = \alpha_i, \mathbf{X}\right] \\ &= \sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X}) \end{aligned}$$

- In general, we have

$$R(h(\mathbf{X}) \mid \mathbf{X}) = \sum_{j=0}^{M-1} L(h(\mathbf{X}), C_j) q_j(\mathbf{X})$$

- Let f denote the density of $\mathbf{X}$. Now risk of any classifier is

$$R(h) \;=\; E\,[\;E[L(h(\mathbf{X}), y(\mathbf{X}))\mid \mathbf{X}]\;]$$

- Let f denote the density of $\mathbf{X}$. Now risk of any classifier is

$$\begin{aligned} R(h) &= E\left[\, E[L(h(\mathbf{X}), y(\mathbf{X})) \mid \mathbf{X}]\,\right] \\ &= \int R(h(\mathbf{X}) \mid \mathbf{X}) f(\mathbf{X}) \; d\mathbf{X} \end{aligned}$$

- Let f denote the density of $\mathbf{X}$. Now risk of any classifier is

$$\begin{aligned} R(h) &= E\left[\, E[L(h(\mathbf{X}), y(\mathbf{X})) \mid \mathbf{X}]\,\right] \\ &= \int R(h(\mathbf{X}) \mid \mathbf{X}) f(\mathbf{X})\; d\mathbf{X} \end{aligned}$$

- The optimal classifier:

  for each $\mathbf{X}$, $h(\mathbf{X})$ should minimize $R(h(\mathbf{X}) \mid \mathbf{X})$.

The Bayes Classifier

- Recall $R(h(\mathbf{X}) \mid \mathbf{X}) = \sum_{j=0}^{M-1} L(h(\mathbf{X}), C_j) q_j(\mathbf{X})$

The Bayes Classifier

- Recall $R(h(\mathbf{X}) \mid \mathbf{X}) = \sum_{j=0}^{M-1} L(h(\mathbf{X}), C_j) q_j(\mathbf{X})$
- The Bayes classifier, h_B for the M-class case is:

  $h_B(\mathbf{X}) = \alpha_i$ if

$$\sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X}) \leq \sum_{j=0}^{M-1} L(\alpha_k, C_j) q_j(\mathbf{X}), \ \forall k$$

  (Break ties arbitrarily)

The Bayes Classifier

- Recall $R(h(\mathbf{X}) \mid \mathbf{X}) = \sum_{j=0}^{M-1} L(h(\mathbf{X}), C_j) q_j(\mathbf{X})$
- The Bayes classifier, h_B for the M-class case is:

  $h_B(\mathbf{X}) = \alpha_i$ if

  $$\sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X}) \leq \sum_{j=0}^{M-1} L(\alpha_k, C_j) q_j(\mathbf{X}), \ \forall k$$

  (Break ties arbitrarily)
- Thus $R(h_B(\mathbf{X}) \mid \mathbf{X}) \leq R(h(\mathbf{X}) \mid \mathbf{X}), \ \forall h$ and thus Bayes classifier is optimal.

- Take $M = 2$. Now the Bayes classifier is:

  $h_B(\mathbf{X}) = \alpha_0$ if

$$L(\alpha_0, C_0)q_0(\mathbf{X}) + L(\alpha_0, C_1)q_1(\mathbf{X}) \leq$$

$$L(\alpha_1, C_0)q_0(\mathbf{X}) + L(\alpha_1, C_1)q_1(\mathbf{X})$$

- Take $M = 2$. Now the Bayes classifier is:

  $h_B(\mathbf{X}) = \alpha_0$ if

$$L(\alpha_0, C_0)q_0(\mathbf{X}) + L(\alpha_0, C_1)q_1(\mathbf{X}) \leq$$

$$L(\alpha_1, C_0)q_0(\mathbf{X}) + L(\alpha_1, C_1)q_1(\mathbf{X})$$

- Same as

$$\frac{q_0(\mathbf{X})}{q_1(\mathbf{X})} \geq \frac{L(\alpha_0, C_1)}{L(\alpha_1, C_0)} \quad \text{if} \quad L(\alpha_0, C_0) = L(\alpha_1, C_1) = 0.$$

- Same as the Bayes classifier we saw earlier.

- Take M-class case and consider 0–1 loss function. Then

$$R(\alpha_i \mid \mathbf{X}) = \sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X})$$

- Take M-class case and consider 0–1 loss function. Then

$$
\begin{aligned}
R(\alpha_i \mid \mathbf{X}) &= \sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X}) \\
&= \sum_{j \neq i} q_j(\mathbf{X}) = 1 - q_i(\mathbf{X})
\end{aligned}
$$

- Thus, for M-class case and 0–1 loss function

$$R(\alpha_i \mid \mathbf{X}) \;=\; \sum_{j \neq i} q_j(\mathbf{X}) \;=\; 1 - q_i(\mathbf{X})$$

- Thus the Bayes classifier is: $h_B(\mathbf{X}) = \alpha_i$ if

$$(1 - q_i(\mathbf{X})) \leq (1 - q_j(\mathbf{X})) \;\text{ or }\; q_i(\mathbf{X}) \geq q_j(\mathbf{X}), \quad \forall j$$

- Thus, for M-class case and 0–1 loss function

$$R(\alpha_i \mid \mathbf{X}) \;=\; \sum_{j \neq i} q_j(\mathbf{X}) \;=\; 1 - q_i(\mathbf{X})$$

- Thus the Bayes classifier is: $h_B(\mathbf{X}) = \alpha_i$ if

$$(1 - q_i(\mathbf{X})) \leq (1 - q_j(\mathbf{X})) \;\text{ or }\; q_i(\mathbf{X}) \geq q_j(\mathbf{X}), \quad \forall j$$

- This is the M-class classifier for 0–1 loss function. Minimizes probability of misclassification.

Bayes Classifier – General Case

- The Bayes classifier that minimizes risk is:

  $h_B(\mathbf{X}) = \alpha_i$ if (Break ties arbitrarily)

$$R(\alpha_i \mid \mathbf{X}) \leq R(\alpha_j \mid \mathbf{X}), \ \forall j.$$

  where

$$R(\alpha_i \mid \mathbf{X}) = \sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X})$$

Bayes Classifier – General Case

- The Bayes classifier that minimizes risk is:

  $h_B(\mathbf{X}) = \alpha_i$ if (Break ties arbitrarily)

$$R(\alpha_i \mid \mathbf{X}) \leq R(\alpha_j \mid \mathbf{X}), \ \forall j.$$

  where

$$R(\alpha_i \mid \mathbf{X}) = \sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X})$$

- Note that this is the most general case. (Even when $L(\alpha_i, C_i) \neq 0$). This is optimal for minimizing risk.

Some special cases

- Let us consider the two class case and explicitly write down Bayes classifier for some specific class conditional densities.

Some special cases

- Let us consider the two class case and explicitly write down Bayes classifier for some specific class conditional densities.
- For simplicity we write $L(\alpha_i, C_j) = L(i, j)$. We also assume $L(0, 0) = L(1, 1) = 0$.

Some special cases

- Let us consider the two class case and explicitly write down Bayes classifier for some specific class conditional densities.
- For simplicity we write $L(\alpha_i, C_j) = L(i, j)$. We also assume $L(0, 0) = L(1, 1) = 0$.
- For the 2-class case, we decide on C_0 if

$$\frac{q_0(\mathbf{X})}{q_1(\mathbf{X})} = \frac{f_0(\mathbf{X})p_0}{f_1(\mathbf{X})p_1} \geq \frac{L(0,1)}{L(1,0)}$$

Normal class conditional densities

- We start with the simple case of $\mathbf{X} \in \Re$ (hence use X for $\mathbf{X}$) and both class conditional densities normal.

$$f_i(X) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(\frac{-(X-\mu_i)^2}{2\sigma_i^2}\right), \quad i = 0, 1$$

Normal class conditional densities

- We start with the simple case of $\mathbf{X} \in \Re$ (hence use X for $\mathbf{X}$) and both class conditional densities normal.

$$f_i(X) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(\frac{-(X-\mu_i)^2}{2\sigma_i^2}\right), \quad i = 0, 1$$

- $h_B(X) = 0$ if

$$p_0 f_0(X) L(1,0) > p_1 f_1(X) L(0,1)$$

Normal class conditional densities

- We start with the simple case of $\mathbf{X} \in \Re$ (hence use X for $\mathbf{X}$) and both class conditional densities normal.

$$f_i(X) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(\frac{-(X-\mu_i)^2}{2\sigma_i^2}\right), \quad i = 0, 1$$

- $h_B(X) = 0$ if

$$p_0 f_0(X) L(1,0) > p_1 f_1(X) L(0,1)$$

- Same as

$$\ln(p_0 L(1,0)) + \ln(f_0(X)) > \ln(p_1 L(0,1)) + \ln(f_1(X))$$

- That is, $h_B(X) = 0$ if

$$\ln(p_0 L(1,0)) - \ln(\sigma_0) - \frac{1}{2}\ln(2\pi) - \frac{(X-\mu_0)^2}{2\sigma_0^2} >$$
$$\ln(p_1 L(0,1)) - \ln(\sigma_1) - \frac{1}{2}\ln(2\pi) - \frac{(X-\mu_1)^2}{2\sigma_1^2}$$

- That is, $h_B(X) = 0$ if
$$\ln(p_0 L(1,0)) - \ln(\sigma_0) - \frac{1}{2}\ln(2\pi) - \frac{(X-\mu_0)^2}{2\sigma_0^2} >$$
$$\ln(p_1 L(0,1)) - \ln(\sigma_1) - \frac{1}{2}\ln(2\pi) - \frac{(X-\mu_1)^2}{2\sigma_1^2}$$
- That is, $h_B(X) = 0$ if
$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$
$$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$

- That is, $h_B(X) = 0$ if

$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$
$$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$

- That is, $h_B(X) = 0$ if
$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$
$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$
- This is of the form

$$h_B(X) = 0 \quad \text{if} \quad aX^2 + bX + c > 0$$

where a, b, c are some constants.

- That is, $h_B(X) = 0$ if
$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$
$$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$
- This is of the form
$$h_B(X) = 0 \quad \text{if} \quad aX^2 + bX + c > 0$$
where a, b, c are some constants.
- Thus the Bayes classifier in this case is a quadratic discriminant function.

some special cases

- The Bayes classifier is: $h_B(X) = 0$ if
  $$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$
  $$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$

some special cases

- The Bayes classifier is: $h_B(X) = 0$ if
$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right)+X\left(\frac{\mu_0}{\sigma_0^2}-\frac{\mu_1}{\sigma_1^2}\right)$$
$$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2}-\frac{\mu_0^2}{\sigma_0^2}\right)+\ln\left(\frac{\sigma_1^2}{\sigma_0^2}\right)+\ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right)>0$$
- Take $\sigma_0 = \sigma_1 = \sigma$, $p_0 = p_1$ and $L(1,0) = L(0,1)$.

some special cases

- The Bayes classifier is: $h_B(X) = 0$ if
$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$
$$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1^2}{\sigma_0^2}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$
- Take $\sigma_0 = \sigma_1 = \sigma$, $p_0 = p_1$ and $L(1,0) = L(0,1)$. Then
$$\frac{X}{\sigma^2}(\mu_0 - \mu_1) - \frac{1}{2\sigma^2}(\mu_0^2 - \mu_1^2) > 0$$

some special cases

- The Bayes classifier is: $h_B(X) = 0$ if
$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$
$$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$
- Take $\sigma_0 = \sigma_1 = \sigma$, $p_0 = p_1$ and $L(1,0) = L(0,1)$. Then
$$\frac{X}{\sigma^2}(\mu_0 - \mu_1) - \frac{1}{2\sigma^2}(\mu_0^2 - \mu_1^2) > 0$$
- That is, $X > \frac{\mu_0+\mu_1}{2}$, assuming $\mu_0 > \mu_1$.

some special cases

- The Bayes classifier is: $h_B(X) = 0$ if
$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$
$$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$
- Take $\sigma_0 = \sigma_1 = \sigma$, $p_0 = p_1$ and $L(1,0) = L(0,1)$. Then
$$\frac{X}{\sigma^2}(\mu_0 - \mu_1) - \frac{1}{2\sigma^2}(\mu_0^2 - \mu_1^2) > 0$$
- That is, $X > \frac{\mu_0+\mu_1}{2}$, assuming $\mu_0 > \mu_1$.
- Intuitively the classifier is very clear.

some special cases

- The Bayes classifier is: $h_B(X) = 0$ if
  $$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$
  $$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$

some special cases

- The Bayes classifier is: $h_B(X) = 0$ if
$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$
$$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$
- Take $\mu_0 = \mu_1 = 0$, $p_0 = p_1$ and $L(1,0) = L(0,1)$. Then
$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) - \ln\left(\frac{\sigma_0}{\sigma_1}\right) > 0$$

some special cases

- The Bayes classifier is: $h_B(X) = 0$ if

$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$

$$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) > 0$$

- Take $\mu_0 = \mu_1 = 0$, $p_0 = p_1$ and $L(1,0) = L(0,1)$. Then

$$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) - \ln\left(\frac{\sigma_0}{\sigma_1}\right) > 0$$

- Assuming $\sigma_0 > \sigma_1$, this is same as

$$X^2 > \frac{\sigma_1^2\sigma_0^2 \ln(\sigma_0/\sigma_1)}{(\sigma_0^2 - \sigma_1^2)}$$

some special cases

- The Bayes classifier is: $h_B(X) = 0$ if
$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + X\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$
$+\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left(\frac{p_0 L(1{,}0)}{p_1 L(0{,}1)}\right) > 0$
- Take $\mu_0 = \mu_1 = 0$, $p_0 = p_1$ and $L(1,0) = L(0,1)$. Then
$\frac{1}{2}X^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) - \ln\left(\frac{\sigma_0}{\sigma_1}\right) > 0$
- Assuming $\sigma_0 > \sigma_1$, this is same as
$X^2 > \frac{\sigma_1^2\sigma_0^2\ln(\sigma_0/\sigma_1)}{(\sigma_0^2 - \sigma_1^2)}$ (again, intuitively clear).

- Now let us consider the case of $\mathbf{X} \in \Re^n$ and normal class conditional densities.

$$f_i(\mathbf{X}) =$$
$$((2\pi)^n|\Sigma_i|)^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_i)^T\Sigma_i^{-1}(\mathbf{X}-\boldsymbol{\mu}_i)), \;\; i = 0, 1$$

- Now let us consider the case of $\mathbf{X} \in \Re^n$ and normal class conditional densities.

$$f_i(\mathbf{X}) = ((2\pi)^n|\Sigma_i|)^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_i)^T\Sigma_i^{-1}(\mathbf{X}-\boldsymbol{\mu}_i)), \quad i = 0, 1$$

- The Bayes classifier is: $h_B(\mathbf{X}) = 0$ if

$$\ln(p_0 L(1,0)) + \ln(f_0(\mathbf{X})) > \ln(p_1 L(0,1)) + \ln(f_1(\mathbf{X})).$$

- Now let us consider the case of $\mathbf{X} \in \Re^n$ and normal class conditional densities.

$$f_i(\mathbf{X}) = \left((2\pi)^n |\Sigma_i|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{X}-\boldsymbol{\mu}_i)\right), \quad i = 0, 1$$

- The Bayes classifier is: $h_B(\mathbf{X}) = 0$ if

$$\ln(p_0 L(1,0)) + \ln(f_0(\mathbf{X})) > \ln(p_1 L(0,1)) + \ln(f_1(\mathbf{X})).$$

- That is,

$$\ln(p_0 L(1,0)) - \frac{1}{2}\ln(|\Sigma_0|) - \frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\mathbf{X}-\boldsymbol{\mu}_0) > \ln(p_1 L(0,1)) - \frac{1}{2}\ln(|\Sigma_1|) - \frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\mathbf{X}-\boldsymbol{\mu}_1)$$

- We have $h_B(\mathbf{X}) = 0$ if

$$\ln(p_0 L(1,0)) - \frac{1}{2}\ln(|\Sigma_0|) - \frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_0)^T \Sigma_0^{-1}(\mathbf{X}-\boldsymbol{\mu}_0) > \ln(p_1 L(0,1)) - \frac{1}{2}\ln(|\Sigma_1|) - \frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_1)^T \Sigma_1^{-1}(\mathbf{X}-\boldsymbol{\mu}_1)$$

- We have $h_B(\mathbf{X}) = 0$ if

$$\ln(p_0 L(1,0)) - \frac{1}{2}\ln(|\Sigma_0|) - \frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_0)^T \Sigma_0^{-1}(\mathbf{X}-\boldsymbol{\mu}_0) > \ln(p_1 L(0,1)) - \frac{1}{2}\ln(|\Sigma_1|) - \frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_1)^T \Sigma_1^{-1}(\mathbf{X}-\boldsymbol{\mu}_1)$$

- That is

$$\frac{1}{2}\mathbf{X}^T(\Sigma_1^{-1} - \Sigma_0^{-1})\mathbf{X} + \mathbf{X}^T(\Sigma_0^{-1}\boldsymbol{\mu}_o - \Sigma_1^{-1}\boldsymbol{\mu}_1) + \frac{1}{2}(\boldsymbol{\mu}_1^T \Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T \Sigma_0^{-1}\boldsymbol{\mu}_0) + \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) + \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_0|}\right) > 0$$

- We have $h_B(\mathbf{X}) = 0$ if

$$\begin{aligned}\ln(p_0 L(1,0)) - \tfrac{1}{2}\ln(|\Sigma_0|) - \tfrac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\mathbf{X} - \boldsymbol{\mu}_0) > \\ \ln(p_1 L(0,1)) - \tfrac{1}{2}\ln(|\Sigma_1|) - \tfrac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\mathbf{X} - \boldsymbol{\mu}_1)\end{aligned}$$

- That is

$$\begin{aligned}&\tfrac{1}{2}\mathbf{X}^T(\Sigma_1^{-1} - \Sigma_0^{-1})\mathbf{X} + \mathbf{X}^T(\Sigma_0^{-1}\boldsymbol{\mu}_o - \Sigma_1^{-1}\boldsymbol{\mu}_1) \\ &\quad + \tfrac{1}{2}(\boldsymbol{\mu}_1^T \Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T \Sigma_0^{-1}\boldsymbol{\mu}_0) \\ &+ \ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) + \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_0|}\right) > 0\end{aligned}$$

- Once again, the Bayes classifier is a quadratic discriminant function.

- The Bayes classifier is a discriminant function given by

$$\frac{1}{2}\mathbf{X}^T(\Sigma_1^{-1} - \Sigma_0^{-1})\mathbf{X} + \mathbf{X}^T(\Sigma_0^{-1}\boldsymbol{\mu}_o - \Sigma_1^{-1}\boldsymbol{\mu}_1)$$
$$+\frac{1}{2}(\boldsymbol{\mu}_1^T\Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\Sigma_0^{-1}\boldsymbol{\mu}_0)$$
$$+\ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) + \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_0|}\right) > 0$$

- The Bayes classifier is a discriminant function given by

$$\frac{1}{2}\mathbf{X}^T(\Sigma_1^{-1} - \Sigma_0^{-1})\mathbf{X} + \mathbf{X}^T(\Sigma_0^{-1}\boldsymbol{\mu}_o - \Sigma_1^{-1}\boldsymbol{\mu}_1)$$
$$+\frac{1}{2}(\boldsymbol{\mu}_1^T\Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\Sigma_0^{-1}\boldsymbol{\mu}_0)$$
$$+\ln\left(\frac{p_0 L(1{,}0)}{p_1 L(0{,}1)}\right) + \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_0|}\right) > 0$$

- Consider the specisl case $\Sigma_i = \Sigma$.

- The Bayes classifier is based on the discriminant function

$$\frac{1}{2}\mathbf{X}^T(\Sigma_1^{-1} - \Sigma_0^{-1})\mathbf{X} + \mathbf{X}^T(\Sigma_0^{-1}\boldsymbol{\mu}_o - \Sigma_1^{-1}\boldsymbol{\mu}_1)$$
$$+\frac{1}{2}(\boldsymbol{\mu}_1^T\Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\Sigma_0^{-1}\boldsymbol{\mu}_0)$$
$$+\ln\left(\frac{p_0 L(1,0)}{p_1 L(0,1)}\right) + \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_0|}\right) > 0$$

- Consider the specisl case $\Sigma_i = \Sigma$.
- Then the quadratic term Vanishes.
- The Bayes classifier now becomes a linear discriminant function.

- In the specisl case $\Sigma_i = \Sigma$, the Bayes classifier is:
- $h_B(\mathbf{X}) = 0$ if $g(\mathbf{X}) > 0$, where

  $g(\mathbf{X}) = \mathbf{W}^T\mathbf{X} + w_0$, with

$$\mathbf{W} = \Sigma^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)$$

$$w_0 = \frac{1}{2}(\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\Sigma^{-1}\boldsymbol{\mu}_0) + \ln\left(\frac{p_0 L(1{,}0)}{p_1 L(0{,}1)}\right)$$

- This is a linear discriminant function

- Bayes classifier can be similarly derived for any other class conditional densities.

- Bayes classifier can be similarly derived for any other class conditional densities.
- Depending on the nature of the densities the final expressions can be complicated. However, given all the class conditional densities and prior probabilities, we can easily decide on the class of any given feature vector.

- Bayes classifier can be similarly derived for any other class conditional densities.
- Depending on the nature of the densities the final expressions can be complicated. However, given all the class conditional densities and prior probabilities, we can easily decide on the class of any given feature vector.
- For example, in a 2-class case with 0-1 loss function, given a $\mathbf{X}$, we decide on the class based on whether or not the inequality $p_0 f_0(\mathbf{X}) > p_1 f_1(\mathbf{X})$ is satisfied.

- Bayes classifier can be similarly derived for any other class conditional densities.
- Depending on the nature of the densities the final expressions can be complicated. However, given all the class conditional densities and prior probabilities, we can easily decide on the class of any given feature vector.
- For example, in a 2-class case with 0-1 loss function, given a $\mathbf{X}$, we decide on the class based on whether or not the inequality $p_0 f_0(\mathbf{X}) > p_1 f_1(\mathbf{X})$ is satisfied.
- Given full statistical information, this is the optimal decision.

- We discussed discriminant functions only for two class case.

- We discussed discriminant functions only for two class case.
- $h(\mathbf{X}) = 0$ if $g(\mathbf{X}) \geq 0$; $\quad h(\mathbf{X}) = 1$ otherwise

- We discussed discriminant functions only for two class case.
- $h(\mathbf{X}) = 0$ if $g(\mathbf{X}) \geq 0; \quad h(\mathbf{X}) = 1$ otherwise
- It is not immediately obvious how this is to be extended to the multi-class case.

- We discussed discriminant functions only for two class case.
- $h(\mathbf{X}) = 0$ if $g(\mathbf{X}) \geq 0$; $h(\mathbf{X}) = 1$ otherwise
- It is not immediately obvious how this is to be extended to the multi-class case.
- The Bayes classifier for the multi-class case is one such generalization.

- Bayes classifier for M-class case is: $h_B(\mathbf{X}) = \alpha_i$ if

$$\sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X}) \leq \sum_{j=0}^{M-1} L(\alpha_k, C_j) q_j(\mathbf{X}), \ \forall k$$

- Bayes classifier for M-class case is: $h_B(\mathbf{X}) = \alpha_i$ if

$$\sum_{j=0}^{M-1} L(\alpha_i, C_j) q_j(\mathbf{X}) \leq \sum_{j=0}^{M-1} L(\alpha_k, C_j) q_j(\mathbf{X}), \ \forall k$$

- Consider the case of 0-1 loss function. Then the above is same as $p_i f_i(\mathbf{X}) \geq p_k f_k(\mathbf{X}), \ \forall k$

  or $\ln(p_i f_i(\mathbf{X})) \geq \ln(p_k f_k(\mathbf{X})), \ \forall k$

- Define $g_i(\mathbf{X}) = \ln(f_i(\mathbf{X})) + \ln(p_i)$, $i = 0, 1, \cdots, M-1$.
- Now, the Bayes classifier is:

  Decide on class-i if $g_i(\mathbf{X}) \geq g_j(\mathbf{X}) \ \forall j$

- Define $g_i(\mathbf{X}) = \ln(f_i(\mathbf{X})) + \ln(p_i)$, $i = 0, 1, \cdots, M-1$.
- Now, the Bayes classifier is:

  Decide on class-i if $g_i(\mathbf{X}) \geq g_j(\mathbf{X}) \; \forall j$
- This is the form of discriminant function based classifier for the M-class case.