Displays for Statistics 5401

Lecture 35

November 30, 2005

Christopher Bingham, Instructor
612-625-1024

Class Web Page
http://www.stat.umn.edu/~kb/classes/5401

The minimum ECM (expected cost of misclassification) and minimum TPM (total probability of misclassification) rules are based on $\text{ECM}_i(\mathbf{x})$, where

- $\text{ECM}_i(\mathbf{x})$ = conditional expected cost, given $\mathbf{x}$ (but not knowing the population $\mathbf{x}$ comes from), of classifying $\mathbf{x}$ as from π_i.

$\text{ECM}_i(\mathbf{x})$ weights the costs $C(i|j)$, $j \neq i$, by posterior probabilities $P(\pi_j|\mathbf{x})$.

Since $C(i|i) = 0$,

$$\text{ECM}_i(\mathbf{x}) = \sum_{1\leq j\leq g} P(\pi_j|\mathbf{x})C(i|j)$$

The posterior probabilities are

$$P(\pi_j|x) = p_j f_j(\mathbf{x})/\{\sum_{1\leq k\leq g} p_k f_k(\mathbf{x})\},\ 1\leq j\leq g$$

so

$$\text{ECM}_i(\mathbf{x}) = \frac{\sum_{1\leq j\leq g} p_j f_j(\mathbf{x})C(i|j)}{\sum_{1\leq k\leq g} p_k f_k(\mathbf{x})}$$

Statement of minimum ECM rule

> Select the π_i for which $ECM_i(\mathbf{x})$ is smallest.

More precisely,

$\hat{\pi}_{\text{min ECM}}(\mathbf{x}) = \pi_j$, where

$$ECM_j(\mathbf{x}) = \min_{1\le i\le g} ECM_i(\mathbf{x})$$

In words, the minimum ECM rule is:

"Select the population with the least *posterior* expected misclassification cost."

The denominator $\sum_{1\le k\le g} p_k f_k(\mathbf{x})$ is the same for all $ECM_i(\mathbf{x})$ $i = 1, \ldots, g$.

This means that you can restate $\hat{\pi}_{\text{min ECM}}$ as:

- Select π_i so as to minimize

$$\sum_{1\le j\le g} p_j f_j(\mathbf{x}) C(i\,|\,j) = \sum_{j\ne i} p_j f_j(\mathbf{x}) C(i\,|\,j)$$

When costs are equal ($C(i\,|\,j) = c$, $i \ne j$), $\hat{\pi}_{\text{min TPM}}(\mathbf{x}) = \hat{\pi}_{\text{min ECM}}(\mathbf{x})$ and

$$\begin{aligned} ECM_i(\mathbf{x}) &= c\sum_{j\ne i} p_j f_j(\mathbf{x}) / \sum_{1\le k\le g} p_k f_k(\mathbf{x}) \\ &= c(1 - p_i f_i(\mathbf{x}) / \sum_{1\le k\le g} p_k f_k(\mathbf{x})) \\ &= c(1 - P(\pi_i\,|\,\mathbf{x})) \\ &= c(1 - \text{posterior probability of } \pi_i \text{ given } \mathbf{x}) \end{aligned}$$

This means you can state $\hat{\pi}_{\text{min TPM}}(\mathbf{x})$ as

Select π_i to *maximize* $P(\pi_i\,|\,\mathbf{x})$

In words this is

"Select the population with the largest *posterior* probability."

Since all denominators are the same, the rule simplifies to

"Select π_i with largest $p_i f_i(\mathbf{x})$"

or

"Select π_i with largest $\log(p_i) + \log(f_i(x))$"

Two group case ($g = 2$)

When selecting one of *two* groups, only ratios of posterior probabilities or expected costs are important.

- For minimum TPM, the relevant ratio is (since $p_2 = 1 - p_1$):

  $R(\mathbf{x}) \equiv p_1 f_1(\mathbf{x})/((1-p_1)f_2(\mathbf{x})) = OR\times\lambda(\mathbf{x})$

  where

  $\lambda(\mathbf{x}) \equiv f_1(\mathbf{x})/f_2(\mathbf{x})$, the likelihood ratio

  $OR = p_1/(1-p_1)$ = (prior) odds ratio

- For minimum ECM the ratio is:

  $$R(\mathbf{x}) \equiv OR\times CR\times\lambda(\mathbf{x})$$

  $CR = C(2|1)/C(1|2)$ = cost ratio

In both cases, the rule is:

Classify as π_1 when $R(\mathbf{x}) \geq 1$

Classify as π_2 when $R(\mathbf{x}) < 1$

These classification rules (minimum ECM or minimum TPM) are fully specified *only* when you

- can provide prior probabilities p_i (needed for OR)
- can specify costs (needed for CR)
- can compute the likelihood ratio $\lambda(\mathbf{x})$ for which you need $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$,

When you can't specify costs, it is usual to treat them as constant.

With certain types of data, you may be able to estimate p_i. Otherwise, if you don't know p_i, you might assume $p_1 = p_2 = \ldots = p_g = 1/g$.

In practice, you seldom if ever know $f_i(\mathbf{x})$ so you can't compute $\lambda(\mathbf{x})$. Somehow you must *estimate* $f_i(\mathbf{x})$, $i = 1, \ldots, g$.

Typically you have a *training sample* - a a body of data with

n_1 observations $\mathbf{x}_{11}$, $\mathbf{x}_{21}$, ..., $\mathbf{x}_{n_1,1}$ *known* to come from π_1

n_2 observations $\mathbf{x}_{12}$, $\mathbf{x}_{22}$, ..., $\mathbf{x}_{n_2,2}$ *known* to come from π_2

. . . .

n_g observations $\mathbf{x}_{1g}$, $\mathbf{x}_{2g}$, ..., $\mathbf{x}_{n_g,g}$ *known* to come from π_g

You use these data to find estimates of densities $\hat{f}_i(\mathbf{x})$, computable for *any* $\mathbf{x}$.

Then, in the two group case, you estimate the likelihood ratio by

$$\hat{\lambda}(\mathbf{x}) = \hat{f}_1(\mathbf{x})/\hat{f}_2(\mathbf{x}).$$

Finally you use the rule obtained by "plugging" $\hat{\lambda}(\mathbf{x})$ into the minimum TPM or minimum ECM rule.

There are at least two types of estimates for densities, non-parametric and parametric.

Non-parametric density estimates

Histogram estimate

$\hat{f}_i(\mathbf{x})$ = height of the bar of a (multivariate) *histogram* (computed from the training sample from π_i) which contains $\mathbf{x}$.

This amounts to "binning" the observations from each π_i in rectangular cells or "boxes" and estimating the density at $\mathbf{x}$ by

$$\hat{f}_i(\mathbf{x}) = \frac{\text{relative frequency in cell}(\mathbf{x})}{\text{area or volume of cell}(\mathbf{x})}$$

where cell($\mathbf{x}$) $\equiv$ cell containing $\mathbf{x}$.

This is generally feasible only when p is small, unless the samples sizes are huge.

Kernel estimate

$$\hat{f}_i(\mathbf{x}) = n_i^{-1}\sum_{1\le k\le n_i} W(\mathbf{x} - \mathbf{x}_{ki})$$

where $W(\mathbf{x}) \geq 0$ is a multivariate density function with a mode at $\mathbf{0}$.

Examples

- $W(\mathbf{x})$ is $N_p(\mathbf{0},\mathbf{\Sigma})$ density
- $W(\mathbf{x})$ = uniform density over a square or cube centered at $\mathbf{0}$.
- $W(\mathbf{x})$ = uniform density over a circle or sphere centered at $\mathbf{0}$.

You can check that $\hat{f}_i(\mathbf{x})$ is a density (non-negative, integrates to 1).

Usually $W(\mathbf{x})$ is from a family of distributions, which vary in concentration, say $W(\mathbf{x}) = h^p V(h\mathbf{x})$, p = dimension of $\mathbf{x}$, where $V(\mathbf{u})$ is a multivariate density such as $N_p(\mathbf{0}, \mathbf{I}_p)$ or uniform over $\{\mathbf{u} \mid |u_i| < .5\}$ or $\{\mathbf{u} \mid \|\mathbf{u}\| \leq 1\}$.

When $V(\mathbf{u}) = e^{-\|\mathbf{u}\|^2/2}/\{2\pi\}^{p/2}$ is the $N_p(\mathbf{0},\mathbf{I}_p)$ density, $W(\mathbf{x})$ is the $N_p(\mathbf{0},h^{-1}\mathbf{I}_p)$ density

The <u>larger</u> h is,

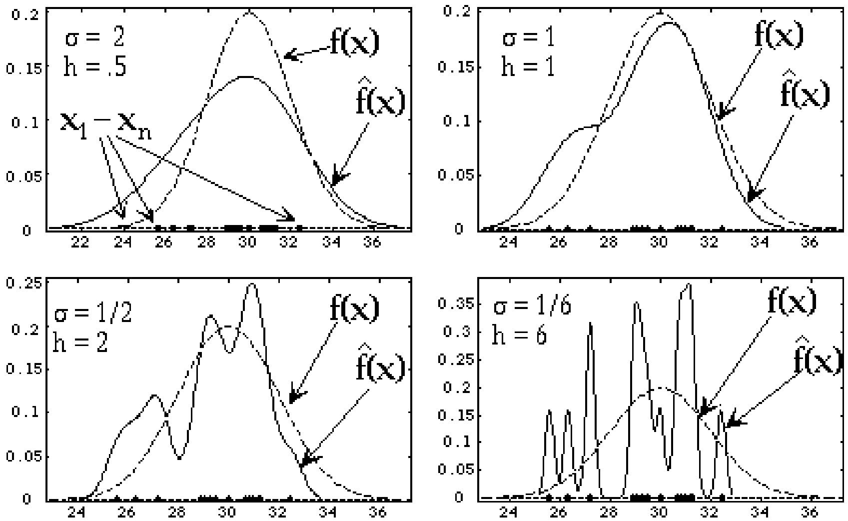
- the more concentrated around the sample point $\mathbf{x}_{ki}$ is $W(\mathbf{x} - \mathbf{x}_{ki})$
- the "bumpier" is $\hat{f}_i(\mathbf{x})$.

The <u>smaller</u> h is,

- the more spread out is $W(\mathbf{x} - \mathbf{x}_{ki})$
- This can result in a featureless estimate with excessive bias.

The key to successful kernel density estimation is determining the degree of concentration (choice of h). h is what is sometimes called a <u>tuning constant</u>. The optimal value of h is usually determined by <u>cross validation</u>.

Univariate (p = 1) example, with W(x) = hV(hx), V(z) standard normal density, with h = 1/σ, σ = 1/h = 2, 1, 1/2, 1/6.



The dashed line is the true $N(30,2^2)$ density and artifical $N(30,2^2)$ data are marked on the x-axis.

The narrower the density W(x) is (smaller σ here), the less smoothing is done and the rougher is the estimated density.

As $\sigma \to 0$, $\hat{f}(\mathbf{x})$ has sharp spikes at the training sample data values.

Parametric density estimates

Suppose you know (or can assume) that $f_i(\mathbf{x}) = g(\mathbf{x},\boldsymbol{\theta}_i)$, $g(\mathbf{x},\boldsymbol{\theta})$ a known density (say $N_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$) with vector of parameters $\boldsymbol{\theta}$.

When $\hat{\boldsymbol{\theta}}_i$ is an estimate of $\boldsymbol{\theta}_i$ computed from training sample data from π_i, you estimate $f_i(\mathbf{x})$ and $\lambda(\mathbf{x}) = f_1(\mathbf{x})/f_2(\mathbf{x})$ by

$$\hat{f}_i(\mathbf{x}) = g(\mathbf{x},\hat{\boldsymbol{\theta}}_i) \text{ and } \hat{\lambda}(\mathbf{x}) = g(\mathbf{x},\hat{\boldsymbol{\theta}}_1)/g(\mathbf{x},\hat{\boldsymbol{\theta}}_2)$$

$g(\mathbf{x},\hat{\boldsymbol{\theta}}_i)$ is often called a "plug-in" density estimate.

This is the approach we focus on, with $f_i(\mathbf{x})$ a $N_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$ density.

- When the $\boldsymbol{\Sigma}_i$'s are equal, you classify using linear functions of $\mathbf{x}$
- With $\boldsymbol{\Sigma}_i$'s that differ, you classify using quadratic functions of $\mathbf{x}$.

Parameter estimates for multivariate normal

Suppose $\mathbf{x}$ in π_i is $N_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$, so

$\boldsymbol{\theta} = [\mu_1,\mu_2,\ldots,\mu_p, \sigma_{11},\sigma_{12},\sigma_{22},\ldots,\sigma_{p-1,p},\sigma_{pp}]'$, $p(p+3)/2$ parameters.

Estimates of the $\boldsymbol{\mu}_i$'s are

- $\hat{\boldsymbol{\mu}}_i = \overline{\mathbf{x}}_i$, $i = 1,\ldots,g$

When you can assume $\boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$, you estimate of $\boldsymbol{\Sigma}$ by

$$\hat{\boldsymbol{\Sigma}} = \mathbf{S}_{pooled} = (N - g)^{-1}\sum_{1\leq i\leq g}(n_i-1)\mathbf{S}_i = f_e^{-1}\mathbf{E},$$

$\mathbf{E}$ the MANOVA error matrix, $f_e = N - g$.

With unrestricted $\boldsymbol{\Sigma}_i$'s, you estimate $\boldsymbol{\Sigma}_i$ by

$$\hat{\boldsymbol{\Sigma}}_i = \mathbf{S}_i,\ i = 1,\ldots,g.$$

There are other possibilites, such as $\boldsymbol{\Sigma}_i = k_i\boldsymbol{\Sigma}$, k_i unknown, but I will not explore them.

Classifying data from Multivariate Normal Populations

The $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ density for π_i is

$$f_i(\mathbf{x}) = \frac{\exp\{-(\mathbf{x} - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)/2\}}{(2\pi)^{p/2}\{\det(\boldsymbol{\Sigma}_i)\}^{1/2}}$$

Note: $\exp\{...\}$ means $e^{(\ldots)}$.

Things are neater using log densities:

$$\log f_i(\mathbf{x}) = \text{const}_1 - \log(\det(\boldsymbol{\Sigma}_i))/2 - (\mathbf{x} - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)/2,$$

a quadratic function of $\mathbf{x}$.

You can ignore $\text{const}_1 = -(p/2)\log(2\pi)$ because it the same for all $f_i(\mathbf{x})$ and doesn't affect any comparisons of densities.

<u>Equal variance case</u>: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \ldots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$.

Then $\log f_i(\mathbf{x})$

$$= \text{const}_2 - (\mathbf{x} - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)/2$$
$$= \text{const}_2 - q(\mathbf{x}) - \boldsymbol{\mu}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i/2 + \boldsymbol{\mu}_i'\boldsymbol{\Sigma}^{-1}\mathbf{x}$$
$$= \text{const}_2 - q(\mathbf{x}) - \quad \mathbf{c}_i \quad + \quad \boldsymbol{\ell}_i'\mathbf{x}$$

- $\text{const}_2 = \text{const}_1 - \log(\det(\boldsymbol{\Sigma}))/2$
  $= -(p/2)\log(2\pi) - \log(\det(\boldsymbol{\Sigma}))/2$
- $q(\mathbf{x}) = \mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}/2$, the same for all π_i
- $\boldsymbol{\ell}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i$, $\mathbf{c}_i = \boldsymbol{\mu}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i/2 = \boldsymbol{\ell}_i'\boldsymbol{\mu}_i/2$

You can ignore const_2 and $q(\mathbf{x})$ because they are the same for all π_i.

The part that *does* depend on π_i is

$$-c_i + \boldsymbol{\ell}_i'\mathbf{x} = \boldsymbol{\ell}_i'(\mathbf{x} - \boldsymbol{\mu}_i/2).$$

You classify by comparing g *linear functions* of $\mathbf{x}$,

$$-c_i + \boldsymbol{\ell}_i'\mathbf{x},\ i = 1, \ldots, g.$$

Two groups with $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

When $g = 2$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$

$$\log \lambda(\mathbf{x}) = \log f_1(\mathbf{x}) - \log f_2(\mathbf{x})$$
$$= (\boldsymbol{\ell}_1'\mathbf{x} - \boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1/2) - (\boldsymbol{\ell}_2'\mathbf{x} - \boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2/2)$$
$$= (\boldsymbol{\ell}_1'\mathbf{x} - \quad c_1 \quad) - (\boldsymbol{\ell}_2'\mathbf{x} - \quad c_2 \quad)$$

because $\text{const}_2 - q(\mathbf{x})$ cancel out.

Here

- $\boldsymbol{\ell}_1 = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1$ and $\boldsymbol{\ell}_2 = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2$
- $c_1 = \boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1/2 = \boldsymbol{\ell}_1'\boldsymbol{\mu}_1/2$
  $c_2 = \boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2/2 = \boldsymbol{\ell}_2'\boldsymbol{\mu}_2/2$

Define $\boldsymbol{\ell} \equiv \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \boldsymbol{\ell}_1 - \boldsymbol{\ell}_2$. Then

$$\log \lambda(\mathbf{x}) = \boldsymbol{\ell}'(\mathbf{x} - (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2)$$
$$= \textstyle\sum_{1 \le j \le p} \ell_j\{x_j - (\mu_{j1} + \mu_{j2})/2\}$$

a single *linear* function of $\mathbf{x}$.

$$\lambda(\mathbf{x}) > 1 \Leftrightarrow \boldsymbol{\ell}'(\mathbf{x} - (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2) > 0$$
$$\lambda(\mathbf{x}) < 1 \Leftrightarrow \boldsymbol{\ell}'(\mathbf{x} - (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2) < 0$$

Good rules are based on $\lambda(\mathbf{x}) = f_1(\mathbf{x})/f_2(\mathbf{x})$

You can specify a rule by choosing a suitable constant "cutpoint" k_0:

- Classify as π_1 when $\log \lambda(\mathbf{x}) = \boldsymbol{\ell}'(\mathbf{x} - (\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)/2) \geq k_0$
- Classify as π_2 when $\log \lambda(\mathbf{x}) < k_0$

k_0 is a *cutpoint* or *threshold*.

k_0 depends on prior probabilities and costs, but not parameters.

Define $m = \boldsymbol{\ell}'(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2$. Then you can restate the rule as

- Classify as π_1 when $\boldsymbol{\ell}'\mathbf{x} \geq k_1 \equiv m + k_0$
- Classify as π_2 when $\boldsymbol{\ell}'\mathbf{x} < k_1$

Recall that the minimum ECM rule is

Classify as π_1 when $OR\times CR\times\lambda(\mathbf{x}) \geq 1$

Classify as π_2 when $OR\times CR\times\lambda(\mathbf{x}) < 1$

where

$OR = p_1/(1-p_1) = p_1/p_2$

$= $ (prior) odds ratio

$CR = C(2|1)/C(1|2) =$ cost ratio

That is

Classify as π_1 when $\lambda(\mathbf{x}) \geq 1/(OR\times CR)$

Classify as π_2 when $\lambda(\mathbf{x}) < 1/(OR\times CR)$

Therefore minimum ECM rule uses

- $k_0 = \log(1/(OR\times CR)) = -\log(OR) - \log(CR)$
  $= \log(p_2/p_1) + \log\{C(1|2)/C(2|1)\}$
- $k_1 = \boldsymbol{\ell}'(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2 +$
  $\log(p_2/p_1) + \log\{C(1|2)/C(2|1)\}$

Cutpoints k_0 (for $\log \lambda(\mathbf{x})$) and k_1 (for $\boldsymbol{\ell}'\mathbf{x}$) combine log prior odds and log *mis-classification cost* ratios.

$\boldsymbol{\ell}'\mathbf{x} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\mathbf{x}$, is Fisher's *linear discriminant function*. It was derived under the assumption that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$

The constant

$$k_1 = m + \log(p_2/p_1) + \log\{C(1|2)/C(2|1)\}$$

is a *threshold* or cut-off value separating values of $\boldsymbol{\ell}'\mathbf{x}$ favoring π_1 ($\boldsymbol{\ell}'\mathbf{x} \geq k_1$) from values of $\boldsymbol{\ell}'\mathbf{x}$ favoring π_2 ($\boldsymbol{\ell}'\mathbf{x} < k_1$).

- The more the prior *odds ratio* OR = p_1/p_2 *favors* π_2 (is small)

or

- the more the error *cost ratio* $C(1|2)/C(2|1)$ *disadvantages* π_1

the higher is the threshold $\boldsymbol{\ell}'\mathbf{x}$ must reach in order to select π_1.

Simple case with equal priors and costs:

$$p_1 = p_2 \text{ and } C(1|2) = C(2|1) \Rightarrow k_0 = 0$$

The threshold for $\boldsymbol{\ell}'\mathbf{x}$ is

$$k_1 = m \equiv \boldsymbol{\ell}'(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2,$$

halfway between $\boldsymbol{\ell}'\boldsymbol{\mu}_1$ and $\boldsymbol{\ell}'\boldsymbol{\mu}_2$. That is, classify in π_1 if and only if

$$\boldsymbol{\ell}'\mathbf{x} > \boldsymbol{\ell}'(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2$$

Univariate (p = 1) case

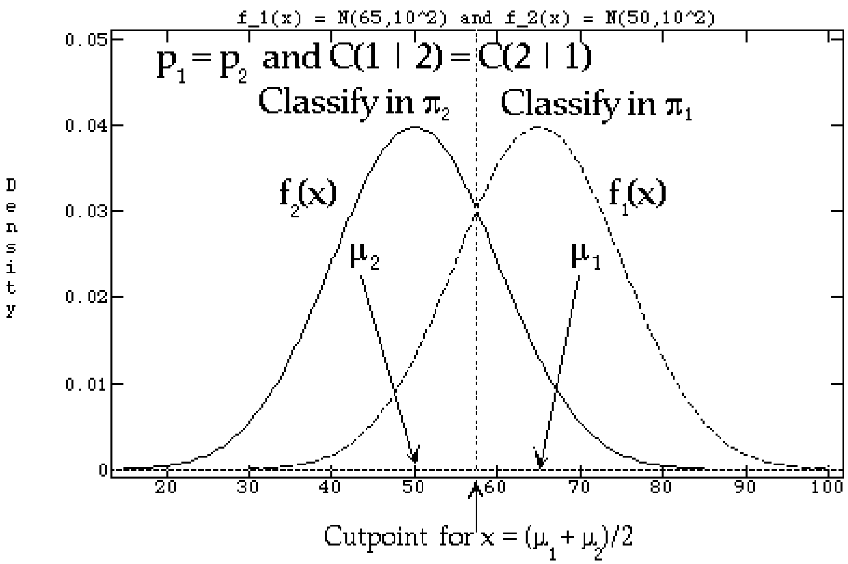
$\boldsymbol{\ell} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = (\mu_1 - \mu_2)/\sigma^2$, a scalar

$m = \boldsymbol{\ell}'(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2 = (\mu_1 - \mu_2)(\mu_1 + \mu_2)/(2\sigma^2)$

$\boldsymbol{\ell}'\mathbf{x} - m =$

$$\{(\mu_1 - \mu_2)/\sigma^2\}x - (\mu_1 - \mu_2)(\mu_1 + \mu_2)/(2\sigma^2)$$
$$= \{(\mu_1 - \mu_2)/(2\sigma^2)\}\{x - (\mu_1 + \mu_2)/2\}$$

When $\mu_1 > \mu_2$, $\boldsymbol{\ell}'\mathbf{x} \geq m \Leftrightarrow$ if $x \geq (\mu_1 + \mu_2)/2$



The graph shows

- $\lambda(x) > 1$ to the left of $(\mu_1 + \mu_2)/2$
- $\lambda(x) < 1$ to the right of $(\mu_1 + \mu_2)/2$

Unequal costs and prior probabilities

- Classify in π_1 when

$$\boldsymbol{\ell}'\mathbf{x} \equiv [(\mu_1 - \mu_2)/\sigma^2]x > (\mu_1 - \mu_2)(\mu_1 + \mu_2)/(2\sigma^2) + \log(p_2/p_1) + \log\{C(1|2)/C(2|1)\}$$

When $\mu_1 < \mu_2$, this is

- Classify in π_1 when

$$x < (\mu_1 + \mu_2)/2 + (\sigma^2/(\mu_1 - \mu_2))\times \{\log(p_2/p_1) + \log\{C(1|2)/C(2|1)\}\}$$

Cut points when $C(1|2) = C(2|1) = 1$ and $p_1 = 0.5$, 0.1 and 0.01.