Maximin Decision Rule

Robert M. Haralick

Computer Science, Graduate Center
City University of New York
Convex Sets and Linear Functions

Proposition

*Images of linear functions of convex sets are convex.*

Proof.

Let $C$ be a convex set and $f : C \rightarrow \mathbb{R}^N$ be a linear function. Define $D = \{ y \in \mathbb{R}^N | y = f(x), x \in C \}$ Let $y_1, y_2 \in D$ and let $0 \leq \lambda \leq 1$. Then there exists $x_1, x_2 \in C$ such that $y_1 = f(x_1)$ and $y_2 = f(x_2)$.

\[
\lambda y_1 + (1 - \lambda y_2) = \lambda f(x_1) + (1 - \lambda) f(x_2) \\
= f(\lambda x_1 + (1 - \lambda) x_2)
\]

But $x_1, x_2 \in C$ and $C$ is convex. Therefore $\lambda x_1 + (1 - \lambda) x_2 \in C$. Hence, $f(\lambda x_1 + (1 - \lambda) x_2) \in D$. And this makes $\lambda y_1 + (1 - \lambda y_2) \in D$. 

\[\square\]
Dependence on Prior Class Probabilities

**Proposition**

Expected economic gain for a decision rule is an affine function of the expected economic conditional gains with coefficients $P(c^1), \ldots, P(c^{K-1})$.

**Proof.**

\[
E[e; f] = \sum_{j=1}^{K} E[e | c^j; f]P(c^j)
\]

\[
= \sum_{j=1}^{K-1} E[e | c^j; f]P(c^j) + E[e | c^K; f](1 - \sum_{j=1}^{K-1} P(c^j))
\]

\[
= \sum_{j=1}^{K-1} \{E[e | c^j; f] - E[e | c^K; f]\}P(c^j) + E[e | c^K; f]
\]
Example

<table>
<thead>
<tr>
<th>e</th>
<th>Assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>$c^1$</td>
</tr>
<tr>
<td>$c^1$</td>
<td>2</td>
</tr>
<tr>
<td>$c^2$</td>
<td>-1</td>
</tr>
</tbody>
</table>

\[
P(d \mid c)
\]

<table>
<thead>
<tr>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Class</td>
</tr>
<tr>
<td>$c^1$</td>
</tr>
<tr>
<td>$c^2$</td>
</tr>
</tbody>
</table>

\[
E[e \mid c^j; f] = \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k)P(d \mid c^j)f_d(c^k)
\]

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Conditional Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>$d^1$</td>
</tr>
<tr>
<td>$f^1$</td>
<td>$c^1$</td>
</tr>
<tr>
<td>$f^2$</td>
<td>$c^1$</td>
</tr>
<tr>
<td>$f^3$</td>
<td>$c^1$</td>
</tr>
<tr>
<td>$f^4$</td>
<td>$c^1$</td>
</tr>
<tr>
<td>$f^5$</td>
<td>$c^2$</td>
</tr>
<tr>
<td>$f^6$</td>
<td>$c^2$</td>
</tr>
<tr>
<td>$f^7$</td>
<td>$c^2$</td>
</tr>
<tr>
<td>$f^8$</td>
<td>$c^2$</td>
</tr>
</tbody>
</table>
Expected Conditional Gain and Expected Gain

\[ E[e \mid c^i; f] = \sum_{d \in D} \sum_{k=1}^{K} e(c^i, c^k) P(d \mid c^i) f_d(c^k) \]

\[ E[e; f] = \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j) \]

\[ = \left[ \sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) \right] + \left[ E[e \mid c^K; f](1 - \sum_{j=1}^{K-1} P(c^j)) \right] \]

\[ = \left[ \sum_{j=1}^{K-1} \{E[e \mid c^j; f] - E[e \mid c^K; f]\} P(c^j) \right] + E[e \mid c^K; f] \]

\[
\begin{align*}
E[e; f^1] &= [2 - (-1)]P(c^1) + (-1) = 3.0P(c^1) - 1 \\
E[e; f^2] &= [.5 - (-.7)]P(c^1) + (.7) = 1.2P(c^1) - .7 \\
E[e; f^3] &= [1.1 - .2]P(c^1) + .2 = 0.9P(c^1) + .2 \\
E[e; f^4] &= [-.4 - .5]P(c^1) + .5 = -0.9P(c^1) + .5 \\
E[e; f^5] &= [1.4 - .5]P(c^1) + .5 = 0.9P(c^1) + .5 \\
E[e; f^6] &= [-.1 - .8]P(c^1) - .8 = -0.9P(c^1) + .8 \\
E[e; f^7] &= [.5 - 1.7]P(c^1)1.7 = -1.2P(c^1) + 1.7 \\
E[e; f^8] &= [-1.0 - 2.0]P(c^1) + 2.0 = -3.0P(c^1) + 2.0
\end{align*}
\]
# Expected Conditional Gain and Expected Gain

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Expected Conditional Gain</th>
<th>Expected Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f )</td>
<td>( d^1 )</td>
<td>( d^2 )</td>
</tr>
<tr>
<td>( f^1 )</td>
<td>( c^1 )</td>
<td>( c^1 )</td>
</tr>
<tr>
<td>( f^2 )</td>
<td>( c^1 )</td>
<td>( c^1 )</td>
</tr>
<tr>
<td>( f^3 )</td>
<td>( c^1 )</td>
<td>( c^2 )</td>
</tr>
<tr>
<td>( f^4 )</td>
<td>( c^1 )</td>
<td>( c^2 )</td>
</tr>
<tr>
<td>( f^5 )</td>
<td>( c^2 )</td>
<td>( c^1 )</td>
</tr>
<tr>
<td>( f^6 )</td>
<td>( c^2 )</td>
<td>( c^1 )</td>
</tr>
<tr>
<td>( f^7 )</td>
<td>( c^2 )</td>
<td>( c^2 )</td>
</tr>
<tr>
<td>( f^8 )</td>
<td>( c^2 )</td>
<td>( c^2 )</td>
</tr>
</tbody>
</table>
\[ E[e; f] = \sum_{j=1}^{K-1} \{ E[e \mid c^j; f] - E[e \mid c^K; f] \} P(c^j) + E[e \mid c^K; f] \]

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Conditional Gain</th>
<th>Expected Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f )</td>
<td>( d^1 )</td>
<td>( d^2 )</td>
</tr>
<tr>
<td>( f^1 )</td>
<td>( c^1 )</td>
<td>( c^1 )</td>
</tr>
<tr>
<td>( f^2 )</td>
<td>( c^1 )</td>
<td>( c^1 )</td>
</tr>
<tr>
<td>( f^3 )</td>
<td>( c^1 )</td>
<td>( c^2 )</td>
</tr>
<tr>
<td>( f^4 )</td>
<td>( c^1 )</td>
<td>( c^2 )</td>
</tr>
<tr>
<td>( f^5 )</td>
<td>( c^2 )</td>
<td>( c^1 )</td>
</tr>
<tr>
<td>( f^6 )</td>
<td>( c^2 )</td>
<td>( c^1 )</td>
</tr>
<tr>
<td>( f^7 )</td>
<td>( c^2 )</td>
<td>( c^2 )</td>
</tr>
<tr>
<td>( f^8 )</td>
<td>( c^2 )</td>
<td>( c^2 )</td>
</tr>
</tbody>
</table>
Dependence on Class Prior Probabilities

The Bayes Gain is the upper envelope
**Convex Functions**

**Definition**

A function $h : \mathbb{R}^N \rightarrow \mathbb{R}$, is a convex function if and only if for every $\lambda$, $0 \leq \lambda \leq 1$,

$$h(\lambda(x_1, \ldots, x_N) + (1 - \lambda)(y_1, \ldots, y_N)) \leq \lambda h(x_1, \ldots, x_N) + (1 - \lambda) h(y_1, \ldots, y_N)$$

**Diagram**

$$w = \lambda u + (1 - \lambda)v$$

$(u, h(u))$ and $(v, h(v))$ are points on the graph of $h$ with $w = \lambda u + (1 - \lambda)v$ lying between them, illustrating the convex property.
Bayes Gain is Convex

Bayes Gain is a convex function of class prior probabilities.
Bayes Gain Is Convex

\[
E[e; f] = \sum_{j=1}^{K} E[e | c^j; f] P(c^j)
\]

\[
G_B = \max_f E[e; f] \text{ Bayes Gain}
\]

Let \(f^n, n = 1, \ldots, N\) be the \(N = |C|^{|D|}\) deterministic decision rules.
Define for \(j = 1, \ldots, K\)

\[
a_{jn} = E[e | c^j; f^n]
\]

\[
p_j = P(c^j)
\]

\[
G_B(P(c^1), \ldots, P(c^K)) = \max_n \sum_{j=1}^{K} E[e | c^j; f^n] P(c^j)
\]

\[
G_B(p_1, \ldots, p_K) = \max_n \sum_{j=1}^{K} a_{jn} p_j
\]
Let \( p = (p_1, \ldots, p_K) \) and \( q = (q_1, \ldots, q_K) \). Let \( 0 \leq \lambda \leq 1 \).

\[
G_B(\lambda p + (1 - \lambda)q) \leq \lambda G_B(p) + (1 - \lambda)G_B(q)
\]

**Proof.**

\[
G_B(\lambda p + (1 - \lambda)q) = \max_n \sum_{j=1}^{K} a_{jn}(\lambda p_j + (1 - \lambda)q_j)
\]

\[
= \max_n \left\{ \lambda \sum_{j=1}^{K} a_{jn}p_j + (1 - \lambda) \sum_{j=1}^{K} a_{jn}q_j \right\}
\]

\[
\leq \left[ \max_n \lambda \sum_{j=1}^{K} a_{jn}p_j \right] + \left[ \max_n (1 - \lambda) \sum_{j=1}^{K} a_{jn}q_j \right]
\]

\[
\leq \lambda G_B(p) + (1 - \lambda)G_B(q)
\]
Definition

Let $f : \mathbb{R}^N \to \mathbb{R}$. The epigraph of $f$, denoted $\text{Epi}(f)$ is the set of points lying on or above the graph of $f$.

$$\text{Epi}(f) = \{(x, u) \in \mathbb{R}^N \times \mathbb{R} \mid u \geq f(x)\}$$
Proposition

If a function is convex then its epigraph is a convex set.

Proof.

Suppose \( f \) is convex. Let \((x, u), (y, v) \in \text{Epi}(f)\) and \(0 \leq \lambda \leq 1\). Then by definition of \(\text{Epi}(f)\), \(f(x) \leq u, f(y) \leq v\) and, therefore, \(\lambda f(x) + (1 - \lambda)f(v) \leq \lambda u + (1 - \lambda)v\). Since \( f \) is convex, \(f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)\). But \(\lambda f(x) + (1 - \lambda)f(v) \leq \lambda u + (1 - \lambda)v\). Now by definition of \(\text{Epi}(f)\), \((\lambda x + (1 - \lambda)y, \lambda u + (1 - \lambda)v) \in \text{Epi}(f)\) making \(\text{Epi}(f)\) convex.
Proposition

If the epigraph of a function is a convex set, then the function is convex.

Proof.

Suppose $\text{Epi}(f)$ is a convex set. Then by definition of $\text{Epi}(f)$, $(x, f(x)) \in \text{Epi}(f)$ and $(y, f(y)) \in \text{Epi}(f)$. Since $\text{Epi}(f)$ is convex, $\lambda(x, f(x)) + (1 - \lambda)(y, f(y)) \in \text{Epi}(f)$. Hence $(\lambda x + (1 - \lambda)y, \lambda f(x) + (1 - \lambda)f(y)) \in \text{Epi}(f)$. By definition of $\text{Epi}(f)$, $f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$. And by definition of a convex function, this implies that $f$ is convex.
Epigraph and Convexity

**Theorem**

*A function is convex if and only if its epigraph is a convex set.*
Basin sets of Convex Functions

**Definition**

Let \( f : \mathbb{R}^N \rightarrow \mathbb{R} \) and \( c \in \mathbb{R} \). A basin set of \( f \) is any set of the form

\[
L = \{ x \in \mathbb{R}^N | f(x) \leq c \}
\]

**Theorem**

Let \( C \) be a convex set, \( h \) be a convex function on \( C \) and \( L = \{ c \in C | h(c) \leq b \} \). Then \( L \) is a convex set.

**Proof.**

Let \( x, y \in L \) so that \( h(x) \leq b \) and \( h(y) \leq b \) and let \( 0 \leq \lambda \leq 1 \). Since \( x, y \in L \subseteq C \) and since \( C \) is a convex set, \( \lambda x + (1 - \lambda)y \in C \). Then since \( h \) is a convex function,

\[
h(\lambda x + (1 - \lambda)y) \leq \lambda h(x) + (1 - \lambda)h(y) \leq \lambda b + (1 - \lambda)b = b
\]

This implies by definition of \( L \) that \( \lambda x + (1 - \lambda)y \in L \).
Corollary

Let $C \subset \mathbb{R}^N$ be a closed and bounded convex set. Let $h : C \to \mathbb{R}$ be a convex function. Suppose $b = \min_{c \in C} h(c)$. Then $M = \{x \in C \mid h(x) = b\}$ is a convex set.

Proof.

Note that since $b = \min_{c \in C} h(c)$, $M = \{x \in C \mid h(x) \leq b\}$. $C$ being closed and bounded is needed because the minima of $h$ may be on the boundary.
Theorem

Let $C$ be a convex set and $h$ be a convex function on $C$. Suppose $h$ has a local minima at $x_0 \in C$. Then for any $x \in C$, $h(x_0) \leq h(x)$.

Proof.

Let $x \in C$ and $1 \geq \alpha > 0$ be sufficiently small so that $(1 - \alpha)x_0 + \alpha x \in C$. Then,

\[
\begin{align*}
    h(x_0) & \leq h((1 - \alpha)x_0 + \alpha x) \\
    0 & \leq \alpha(h(x) - h(x_0)) \\
    h(x_0) & \leq h(x)
\end{align*}
\]
$E[e; f] = \sum_{j=1}^{K-1} \{E[e \mid c^j; f] - E[e \mid c^K; f]\} P(c^j) + E[e \mid c^K; f]$
Dependence on Class Prior Probabilities

Where are the probabilistic decision rules?

\[ \lambda f + (1 - \lambda)g \]
Pick a prior probability $P(c^1)$
For decision rule $f$ there is an Expected Gain $E[e; f]$
For decision rule $g$ there is a Expected Gain $E[e; g]$
For decision rule $\lambda f + (1 - \lambda)g$, the Expected Gain is

$$\lambda E[e; f] + (1 - \lambda)E[e; g]$$

In between the Expected Gain for $f$ and the Expected Gain for $g$
Dependence on Class Prior Probabilities

Decision Rules $f^7$ and $f^1$

Examine the blue line

Where are the probabilistic decision rules?

Lines contained in the area between the lower and the upper envelopes

$P(c_1) = .2$

$E[e; f]$

$P(c^1)$
Expected gain of a mixed decision rule is the mixture of the expected gains of the component decision rules.

\[ E[e; \lambda f + (1 - \lambda)g, P(c^1)] = \lambda E[e; f, P(c^1)] + (1 - \lambda)E[e; g; P(c^1)] \]
Two Class Case

Proposition

Let $0 \leq \lambda \leq 1$. Let $f_1$ and $f_2$ be two decision rules and suppose there are two classes, then $E[e; \lambda f_1 + (1 - \lambda)f_2, P(c^1)]$ is an affine function of $P(c^1)$.

Proof.

\[
E[e; f_1, P(c^1)] = \alpha_1 P(c^1) + \beta_1 \\
E[e; f_2, P(c^1)] = \alpha_2 P(c^1) + \beta_2 \\
E[e; \lambda f_1 + (1 - \lambda)f_2, P(c^1)] = \lambda(\alpha_1 P(c^1) + \beta_1) + (1 - \lambda)(\alpha_2 P(c^1) + \beta_2) \\
= (\lambda \alpha_1 + (1 - \lambda)\alpha_2)P(c^1) + \lambda \beta_1 + (1 - \lambda)\beta_2
\]
Proposition

Fix $P(c^1)$. Let $0 \leq \lambda \leq 1$.

If $E[e; f, P(c^1)] \leq E[e; g, P(c^1)] = \text{then}$

\[
E[e; f, P(c^1)] \leq E[e; \lambda f + (1 - \lambda)g] \leq E[e; g, P(c^1)]
\]

Proof.

\[
\begin{align*}
E[e; f, P(c^1)] &= \lambda E[e; f, P(c^1)] + (1 - \lambda)E[e; f, P(c^1)] \\
E[e; f, P(c^1)] &\leq \lambda E[e; f, P(c^1)] + (1 - \lambda)E[e; g, P(c^1)] \leq E[e; g, P(c^1)] \\
E[e; f, P(c^1)] &\leq E[e; \lambda f + (1 - \lambda)g] \leq E[e; g, P(c^1)]
\end{align*}
\]
Dependence on Class Prior Probabilities

Where are the probabilistic decision rules?

Lines in the area between the lower and the upper envelopes
Dependence on Class Prior Probabilities

Worst Class Priors

\[ P(c^1) = \frac{4}{7} \approx 0.5714 \]
\[ P(c^2) = \frac{3}{7} \approx 0.4286 \]
Finding Worst Class Priors

Two Class Case
Decision Rules of Mixture are Known

\[
E[e; f_5; P(c^1)] = 0.9P(c^1) + 0.5 \\
E[e; f_7; P(c^1)] = -1.2P(c^1) + 1.7 \\
\text{Set } E[e; f_5; P(c^1)] = E[e; f_7; P(c^1)] \\
0.9P(c^1) + 0.5 = -1.2P(c^1) + 1.7; \\
2.1P(c^1) = 1.2 \\
P(c^1) = \frac{1.2}{2.1} = \frac{4}{7} \\
P(c^2) = 1 - P(c^1) = \frac{3}{7}
\]
Suppose we know the deterministic decision rules to make up the mixture: $f_5$ and $f_7$

Since $E[e; f] = E[e; c^1, f]P(c^1) + E[e; c^2, f]P(c^2)$

$$E[e; \lambda f^5 + (1 - \lambda)f^7] = E[e|c^1; \lambda f^5 + (1 - \lambda)f^7]P(c^1) + E[e|c^2; \lambda f^5 + (1 - \lambda)f^7]P(c^2)$$

Since Expectation is a linear operator $E[e|c; \alpha f + \beta g] = \alpha E[e|c; f] + \beta E[e|c; g]$

$$E[e; \lambda f^5 + (1 - \lambda)f^7] = \left(\lambda E[e|c^1; f^5] + (1 - \lambda)E[e|c^1; f^7]\right) P(c^1) +$$

$$\left(\lambda E[e|c^2; f^5] + (1 - \lambda)E[e|c^2; f^7]\right) (1 - P(c^1))$$

$$= \left\{\left(\lambda E[e|c^1; f^5] + (1 - \lambda)E[e|c^1; f^7]\right) - \right.$$

$$\left(\lambda E[e|c^2; f^5] + (1 - \lambda)E[e|c^2; f^7]\right)\} P(c^1) +$$

$$E[e|c^2; \lambda f^5 + (1 - \lambda)f^7]$$

When there is no dependence on priors, the coefficient of $P(c^1)$ must be zero

$$\left(\lambda E[e|c^1; f^5] + (1 - \lambda)E[e|c^1; f^7]\right) - \left(\lambda E[e|c^2; f^5] + (1 - \lambda)E[e|c^2; f^7]\right) = 0$$

The class conditional expected gains must be equal

$$\lambda E[e|c^1; f^5] + (1 - \lambda)E[e|c^1; f^7] = \lambda E[e|c^2; f^5] + (1 - \lambda)E[e|c^2; f^7]$$

$$E[e|c^1; \lambda f_5 + (1 - \lambda)f_7] = E[e|c^3; \lambda f_5 + (1 - \lambda)f_7]$$
Maximin Decision Rule

\[ E[e|c^2; f] \]

\[ E[e|c^1; f] \]
Dependence of a Probabilistic Decision Rule on Priors

\[ \lambda E[e|c^1; f^5] + (1 - \lambda)E[e|c^1; f^7] - \left( \lambda E[e|c^2; f^5] + (1 - \lambda)E[e|c^2; f^7] \right) = 0 \]

\[ \lambda \left( E[e|c^1; f^5] - E[e|c^1; f^7] - E[e|c^2; f^5] + E[e|c^2; f^7] \right) = E[e|c^2; f^7] - E[e|c^1; f^7] \]

\[ \lambda = \frac{E[e|c^2; f^7] - E[e|c^1; f^7]}{E[c^1; f^5] - E[c^1; f^7] - E[e|c^2; f^5] + E[e|c^2; f^7]} \]

\[ = \frac{1.7 - .5}{1.4 - .5 - .5 + 1.7} = \frac{1.2}{2.1} = \frac{4}{7} \]
Require $0 \leq \lambda \leq 1$

$$\lambda = \frac{E[e|c^2; f^7] - E[e|c^1; f^7]}{E[c^1; f^5] - E[c^1; f^7] - E[e|c^2; f^5] + E[e|c^2; f^7]}$$

$\lambda \geq 0$ implies

$$\text{Sign} \left( E[e|c^2; f^7] - E[e|c^1; f^7] \right) = \text{Sign} \left( E[c^1; f^5] - E[c^1; f^7] - E[e|c^2; f^5] + E[e|c^2; f^7] \right)$$

$\lambda \leq 1$ implies

$$|E[e|c^2; f^7] - E[e|c^1; f^7]| \leq |E[c^1; f^5] - E[c^1; f^7] - E[e|c^2; f^5] + E[e|c^2; f^7]|$$

If either of these inequality cannot be satisfied, it implies that the mixture of $f_5$ and $f_7$ is wrong.
Expected Gain As A Function of Priors

The Expected economic gain can be related to the class conditional expected economic gain and prior probabilities.

\[
E[e; f] = \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j)
\]

\[
= \left\{ \sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) \right\} + E[e \mid c^K; f] P(c^K)
\]

\[
= \left\{ \sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) \right\} + E[e \mid c^K; f] \left( 1 - \sum_{j=1}^{K-1} P(c^j) \right)
\]

\[
= \left\{ \sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) \right\} + E[e \mid c^K; f] - \sum_{j=1}^{K-1} E[e \mid c^K; f] P(c^j)
\]

\[
E[e; f, P(c^1), \ldots, P(c^{K-1})] = \left\{ \sum_{j=1}^{K-1} \left( E[e \mid c^j; f] - E[e \mid c^K; f] \right) P(c^j) \right\} + E[e \mid c^K; f]
\]
Expected Gain As A Function Of Priors

\[ E[e; f, P(c^1), \ldots, P(c^{K-1})] = \left\{ \sum_{j=1}^{K-1} \left( E[e | c^j; f] - E[e | c^K; f] \right) P(c^j) \right\} + E[e | c^K; f] \]

Two Class Case

\[ E[e; f, P(c^1)] = \left( E[e | c^1; f] - E[e | c^2; f] \right) P(c^1) + E[e | c^2; f] \]
\[ = \alpha P(c^1) + \gamma \]
\[ E[e; f_1, P(c^1)] = \alpha_{11} P(c^1) + \gamma_1 \]
\[ E[e; f_2, P(c^1)] = \alpha_{21} P(c^1) + \gamma_2 \]

When the expected gains of \( f_1 \) and \( f_2 \) are the same

\[ E[e; f_1, P(c^1)] = E[e; f_2, P(c^1)] \]
\[ \alpha_{11} P(c^1) + \gamma_1 = \alpha_{21} P(c^1) + \gamma_2 \]
\[ (\alpha_{11} - \alpha_{21}) P(c^1) = \gamma_2 - \gamma_1 \]
\[ P(c^1) = \frac{\alpha_{11} - \alpha_{21}}{\gamma_2 - \gamma_1} \]
Three Class Case

\[ E[e; f_i; P(c^1), P(c^2)] = \alpha_{i1} P(c^1) + \alpha_{i2} P(c^2) + \gamma_i, \quad i = 1, 2, 3 \]

\[ E[e; f_i; P(c^1), P(c^2)] = E[e; f_3; P(c^1), P(c^2)], \quad i = 1, 2 \]

\[ \alpha_{11} P(c^1) + \alpha_{12} P(c^2) + \gamma_1 = \alpha_{31} P(c^1) + \alpha_{32} P(c^2) + \gamma_3 \]

\[ \alpha_{21} P(c^1) + \alpha_{22} P(c^2) + \gamma_2 = \alpha_{31} P(c^1) + \alpha_{32} P(c^2) + \gamma_3 \]

\[
\begin{pmatrix}
\alpha_{11} - \alpha_{31} & \alpha_{12} - \alpha_{32} \\
\alpha_{21} - \alpha_{31} & \alpha_{22} - \alpha_{32}
\end{pmatrix}
\begin{pmatrix}
P(c^1) \\
P(c^2)
\end{pmatrix}
= 
\begin{pmatrix}
\gamma_1 - \gamma_3 \\
\gamma_2 - \gamma_3
\end{pmatrix}
\]
$E[e; f_k; P(c^1), \ldots, P(c^{K-1})] = \sum_{i=1}^{K-1} \alpha_{ki} P(c^i) + \gamma_k, \ k = 1, \ldots, K$ 

$E[e; f_k; P(c^1), \ldots, P(c^{K-1})] = E[e; f_K; P(c^1), \ldots, P(c^{K-1})], \ k = 1, \ldots, K - 1$

\[
\begin{pmatrix}
\alpha_{11} - \alpha_{K1} & \alpha_{12} - \alpha_{K2} & \cdots & \alpha_{1,K-1} - \alpha_{K,K-1} \\
\alpha_{K-1,1} - \alpha_{K1} & \alpha_{K-1,2} - \alpha_{K2} & \cdots & \alpha_{K-1,K-1} - \alpha_{K,K-1} \\
\end{pmatrix}
\begin{pmatrix}
P(c^1) \\
P(c^{K-1})
\end{pmatrix} =
\begin{pmatrix}
\gamma_1 - \gamma_K \\
\gamma_{K-1} - \gamma_K
\end{pmatrix}
\]

$0 \leq P(x^k) \leq 1, \ k = 1, \ldots, K - 1$

$\sum_{k=1}^{K} P(c^k) = 1$
Finding The Convex Combination

Two Class Case

\[ \lambda E[e|c^1; f^5] + (1 - \lambda)E[e|c^1; f^7] - (\lambda E[e|c^2; f^5] + (1 - \lambda)E[e|c^2; f^7]) = 0 \]

\[ E[e; \lambda f^5 + (1 - \lambda)f^7] = E[e|c^1; \lambda f^5 + (1 - \lambda)f^7] P(c^1) + E[e|c^2; \lambda f^5 + (1 - \lambda)f^7] P(c^2) \]

Find \( P(c^1) \) that solves \( \alpha_{11} P(c^1) + \gamma_1 = \alpha_{21} P(c^1) + \gamma_2 \). Call the solution \( P_0(c^1) \).

Consider the expected gain of a mixed decision rule that has expected gain \( \alpha_{21} P_0(c^1) + \gamma_2 \) for any prior \( P(c^1) \).

\[ \lambda(\alpha_{11} P(c^1) + \gamma_1) + (1 - \lambda)(\alpha_{21} P(c^1) + \gamma_2) = \alpha_{21} P_0(c^1) + \gamma_2 \]

\[ (\lambda \alpha_{11} + (1 - \lambda) \alpha_{21}) P(c^1) = \alpha_{21} P_0(c^1) + \gamma_2 - \lambda \gamma_1 - (1 - \lambda) \gamma_2 \]

\[ = \alpha_{21} P_0(c^1) - \lambda(\gamma_1 + \gamma_2) \]

Therefore, \( \lambda \alpha_{11} + (1 - \lambda) \alpha_{21} = 0 \) and \( \lambda = \frac{-\alpha_{21}}{\alpha_{11} - \alpha_{21}} = \frac{\alpha_{21} P_0(c^1)}{\gamma_1 + \gamma_2} \)
Two Class Case
Identity in $P(c^1)$ meaning For all $P(c^1)$

\[ 0 \leq \lambda_1, \lambda_2 \leq 1 \]
\[ \lambda_1 + \lambda_2 = 1 \]
\[ \lambda_1(\alpha_{11} P(c^1) + \gamma_1) + \lambda_2(\alpha_{21} P(c^1) + \gamma_2) = \alpha_{21} P_0(c^1) + \gamma_2 \]

\[ (\lambda_1 \alpha_{11} + \lambda_2 \alpha_{21}) P(c^1) = \alpha_{21} P_0(c^1) + \gamma_2 - \lambda_1 \gamma_1 - \lambda_2 \gamma_2 \]

This implies

\[ \lambda_1 \alpha_{11} + \lambda_2 \alpha_{21} = 0 \]
\[ \lambda_1 \gamma_1 + \lambda_2 \gamma_2 = \alpha_{21} P_0(c^1) + \gamma_2 \]
\[ \lambda_1 + \lambda_2 = 1 \]
Finding the Convex Combination

\[
\begin{align*}
\lambda_1 \alpha_{11} + \lambda_2 \alpha_{21} &= 0 \\
\lambda_1 \gamma_1 + \lambda_2 \gamma_2 &= \alpha_{21} P_0(c^1) + \gamma_2 \\
\lambda_1 + \lambda_2 &= 1
\end{align*}
\]

\[
\begin{align*}
\lambda_2 &= -\lambda_1 \frac{\alpha_{11}}{\alpha_{21}} \\
\lambda_1 + \lambda_2 &= \lambda_1 (1 - \frac{\alpha_{11}}{\alpha_{21}}) = 1 \\
\lambda_1 &= \frac{\alpha_{21}}{\alpha_{21} - \alpha_{11}}
\end{align*}
\]
Finding the Convex Combination: Consistency Check

\[ 0 \leq \lambda_1, \lambda_2 \leq 1 \]
\[ \lambda_1 = \frac{\alpha_{21}}{\alpha_{21} - \alpha_{11}} \]

Either \( \alpha_{21} - \alpha_{11} > 0 \) or \( < 0 \).

If \( \alpha_{21} - \alpha_{11} > 0 \) then

\[ \alpha_{21} > \alpha_{11} \]
\[ \alpha_{21} > 0 \]

If \( \alpha_{21} - \alpha_{11} < 0 \) then,

\[ \alpha_{21} < \alpha_{11} \]
\[ \alpha_{21} < 0 \]
Once $\lambda_1$ and $\lambda_2$ are known, the exact value for $P_0(c^1)$ can be determined.

\[
\begin{align*}
\lambda_1 \gamma_1 + (1 - \lambda_1) \gamma_2 &= \alpha_{21} P_0(c^1) + \gamma_2 \\
\lambda_1 (\gamma_1 - \gamma_2) &= \alpha_{21} P_0(c^1) \\
\alpha_{21} P_0(c^1) &= \frac{\lambda_1 (\gamma_1 - \gamma_2)}{\alpha_{21}} \\
&= \frac{\alpha_{21}}{\alpha_{21} - \alpha_{11}} \frac{\gamma_1 - \gamma_2}{\alpha_{21}} \\
&= \frac{\gamma_1 - \gamma_2}{\alpha_{21} - \alpha_{11}}
\end{align*}
\]
Finding The Convex Combination

*K Class Case*

Identity in $P(c^1) \ldots, P(c^{K-1})$

$$
\sum_{k=1}^{K} \lambda_k \left( \sum_{i=1}^{K-1} \alpha_{ik} P(c^i) + \gamma_k \right) = \sum_{i=1}^{K} \alpha_{Ki} P_0(c^i) + \gamma_K
$$

$$
\sum_{i=1}^{K-1} \left( \sum_{k=1}^{K} \lambda_k \alpha_{ik} \right) P(c^i) = \sum_{i=1}^{K} \alpha_{Ki} P_0(c^i) + \gamma_K - \sum_{k=1}^{K} \lambda_k \gamma_k
$$

Implies

$$
\sum_{k=1}^{K} \lambda_k \alpha_{ik} = 0, \ i = 1, \ldots, K - 1
$$

$$
\sum_{k=1}^{K} \lambda_k = 1
$$
Each component decision rule of the mixture has an expected gain that is a hyperplane in the axes $P(c^1), \ldots, P(c^{K-1})$.

The first $K - 1$ rows of the $i$th column consists of the coefficients of $P(c^1), \ldots, P(c^{K-1})$ for the $i$th hyperplane.

\[
\begin{pmatrix}
\alpha_{11} & \alpha_{21} & \ldots & \alpha_{K1} \\
\alpha_{21} & \alpha_{22} & \ldots & \alpha_{K2} \\
\vdots & & & \\
\alpha_{K-1,1} & \alpha_{K-1,2} & \ldots & \alpha_{K-1,K} \\
1 & 1 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_{K-1} \\
\lambda_K
\end{pmatrix}
= 
\begin{pmatrix}
0 \\
0 \\
\vdots \\
0 \\
1
\end{pmatrix}
\]
Dependence on Class Prior Probabilities

\[ f_M = \frac{4}{7} f^5 + \frac{3}{7} f^7 \]
Conditional Expected Gains: All Decision Rules

\[ E[e|c^2; f] \]

\[ E[e|c^1; f] \]
The game is played for a large number of trials.

- Nature chooses class $c$ in accordance with class priors $P(c^1), \ldots, P(c^K)$
- A measurement $d$ is sampled in accordance with $P(d | c)$
- Bayes chooses decision rule to maximize expected gain under given class priors

Suppose nature chooses class priors so that the Bayes gain is minimized. Bayes chooses to maximize expected gain under worst priors. But suppose nature does not choose $c$ in accordance with worst priors.
There is a mixed decision rule that guarantees that regardless of what class priors nature chooses, the expected gain is equal to the Bayes gain under the worst class priors. This is the maximin decision rule.
Dependence on Class Prior Probabilities

\[ f_M = \frac{4}{7} f^5 + \frac{3}{7} f^7 \]
A decision rule $f$ is a **Maximin Decision Rule** if and only if

$$\min_{P(c^1), \ldots, P(c^K)} \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j) \geq \min_{P(c^1), \ldots, P(c^K)} \sum_{j=1}^{K} E[e \mid c^j; g] P(c^j)$$

for any decision rule $g$ where

$$E[e \mid c^j; f] = \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P(d \mid c^j) f_d(c^k)$$
Determining the Maximin Decision Rule

\[ E[e; f] = E[e|c^1; f]P(c^1) + E[e|c^2; f]P(c^2) \]
\[ = E[e|c^1; f]P(c^1) + E[e|c^2; f](1 - P(c^1)) \]
\[ = (E[e|c^1; f] - E[e|c^2; f])P(c^1) + E[e|c^2; f] \]

Since a maximin decision rule has no dependence on the prior probability, we must have

\[ E[e|c^1; f] - E[e|c^2; f] = 0 \]
\[ E[e|c^1; f] = E[e|c^2; f] \]

In this case,

\[ E[e; f] = E[e|c^1; f] \]
\[ = E[e|c^2; f] \]
A decision rule $f$ is a maximin decision rule if and only if

$$\min_{j=1,...,K} E[e | c^j; f] \geq \min_{j=1,...,K} E[e | c^j, g]$$

for any decision rule $g$. 
Maximin Decision Rule

**Theorem**

A decision rule $f$ is a maximin decision rule if and only if

$$\min_{P(c^1), \ldots, P(c^K)} E[e; f, P(c^1), \ldots, P(c^K)] \geq \min_{P(c^1), \ldots, P(c^K)} E[e; g, P(c^1), \ldots, P(c^K)]$$

for any decision rule $g$.

**Proof.**

Recall

$$E[e; f, P(c^1), \ldots, P(c^K)] = E[e; f] = \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j)$$
A decision rule $f$ is a maximin decision rule if and only if the expected gain of $f$ is the same as the expected gain of the Bayes rule under the worst possible prior class probabilities.

**Theorem**

Let $G$ be the Bayes Economic Gain under the worst prior class probabilities. Then $f$ is a maximin decision rule if and only if

$$E[e \mid c^j; f] = G, \; j = 1, \ldots, K$$
Maximin Decision Rule

\[ E[e|c^2; f] \]

\[ E[e|c^1; f] \]

Points: \( f^1, f^5, f^7, f^8 \)

Decision boundary: \( f_M \)

Value: 1.014

55 / 60
Maximin Decision Rule

Let $P(c^1), \ldots, P(c^K)$ be given class prior probabilities. Let $f^m$, $m = 1, \ldots, M$ be $M$ deterministic decision rules satisfying

$$G = \sum_{j=1}^{K} E[e \mid c^j; f^m] P(c^j), \ m = 1, \ldots, M$$

Then there exists $\lambda_m$, $\lambda_m \geq 0$, $m = 1, \ldots, M$, and $\sum_{m=1}^{M} \lambda_m = 1$ satisfying

$$G = E[e \mid c^j; \sum_{m=1}^{M} \lambda_m f^m], \ j = 1, \ldots, K$$

Note:

$$E[e \mid c^j; \sum_{m=1}^{M} \lambda_m f^m] = \sum_{m=1}^{M} \lambda_m E[e \mid c^j; f^m]$$
Maximin Decision Rule

- Let $P(c^1), \ldots, P(c^K)$ be the worst priors
- Let $G_w$ be the worst Bayes gain
- Let $f^m$ be deterministic decision rules, $m = 1, \ldots, M$
  - $G_w = \sum_{j=1}^{K} E[e | c^j; f^m] P(c^j)$
- Find convex combination $\lambda_1, \ldots, \lambda_M$
  - $G_w = E[e | c^k; \sum_{m=1}^{M} \lambda_m f^m] = \sum_{m=1}^{M} \lambda_m E[e | c^j; f^m], j = 1, \ldots, K$
- Let $a_{jm} = E[e | c^j; f^m]$
- Find convex combination $\lambda_1, \ldots, \lambda_M$ satisfying
  - $G_w = \sum_{m=1}^{M} \lambda_m a_{jm}, j = 1, \ldots, K$
Existence of Mixed Decision Rule Strategy

**Theorem**

Let $a_{jm}$ be a real number, $j = 1, \ldots, K; m = 1, \ldots, M$. Let $p_j \geq 0$ and $\sum_{j=1}^{K} p_j = 1$. Suppose

$$G = \sum_{j=1}^{K} p_j a_{jm}, \ m = 1, \ldots, M$$

Then there exists $\lambda_m, \ m = 1, \ldots, M, \ \lambda_m \geq 0$ and $\sum_{m=1}^{M} \lambda_m = 1$ satisfying

$$G = \sum_{m=1}^{M} a_{jm} \lambda_m, \ j = 1, \ldots, K$$
Dependence on Class Prior Probabilities

Worst Class Priors

\[
P(c^1) = \frac{4}{7} \approx 0.5714
\]

\[
P(c^2) = \frac{3}{7} \approx 0.4286
\]
Maximin Decision Rule

\[ E[e|c^1; f_M] = E[e|c^2; f_M] \]

\[ f_M = \frac{4}{7} f^5 + \frac{3}{7} f^7 \]

\[ E[e|c^1; f] \]

\[ E[e|c^2; f] \]