Maximin Decision Rule

Robert M. Haralick

Computer Science, Graduate Center
City University of New York

Outline

Conditional Expectation

Definition

Let X and Y be a discrete random variables that take values from the set $A \times B$. The Conditional Expectation of X given Y is defined by

$$E[Y \mid X = a] = \sum_{b \in B} b P_{XY}(a, b)$$

$E[Y \mid X]$ is a function of the various values that X can take.

The Event (c^j, c^k, d)

$$
\begin{aligned}
P_{TA}(c^j, c^k, d) &= P_{TA}(c^j, c^k | d) P(d) \\
&= P_T(c^j | d) P_A(c^k | d) P(d) \\
&= \frac{P_T(d | c^j) P_T(c^j)}{P(d)} P_A(c^k | d) P(d) \\
&= P_T(d | c^j) P_A(c^k | d) P_T(c^j) \\
P_{AT}(c^k, d | c^j) &= \frac{P_{TA}(c^j, c^k, d)}{P_T(c^j)} \\
&= P_T(d | c^j) P_A(c^k | d) = P_T(d | c^j) f_d(c_k)
\end{aligned}
$$

Expected Conditional Economic Gain Given Class

Definition

The conditional expectation of the economic gain given class c^j for decision rule f is defined by

$$\begin{aligned}
E[e \mid c^j; f] &= \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P_{TA}(c^j, c^k, d) \\
&= \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P(d \mid c^j) f_d(c^k) \\
&= \sum_{k=1}^{K} e(c^j, c^k) \sum_{d \in D} P(d \mid c^j) f_d(c^k)
\end{aligned}$$

where $f_d(c)$ is the conditional probability that the decision rule assigns class c given measurement d.

Class Conditional Probability and Prior Probability

- $P(d|c)$
  - Conditional probability of measurement d given class c
  - Class conditional probability
- $P(c)$
  - Prior probability of class c
  - Prior probability

Economic Gain

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

$$
\begin{aligned}
E[e;f] &= \sum_{d \in D} \sum_{k=1}^{K} \sum_{j=1}^{K} e(c^j, c^k) P(c^j, d) f_d(c^k) \\
&= \sum_{j=1}^{K} \sum_{k=1}^{K} \sum_{d \in D} e(c^j, c^k) P(d \mid c^j) P(c^j) f_d(c^k) \\
&= \sum_{j=1}^{K} \left[\sum_{k=1}^{K} \sum_{d \in D} e(c^j, c^k) P(d \mid c^j) f_d(c^k) \right] P(c^j) \\
&= \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j)
\end{aligned}
$$

Economic Gain

When the economic gain is represented in terms of the prior class probabilities, we write

$$E[e; f, P(c^1), \ldots, P(c^K)]$$

When f is a Bayes decision rule,

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

for any other decision rule g.

Definition

When f is a Bayes decision rule, $E[e; f, P(c^1), \ldots, P(c^K)]$ is called the Bayes gain.

Convex Combinations

Definition

Let $x, y \in \mathbb{R}^N$ and $0 \leq \lambda \leq 1$. Then $\lambda x + (1-\lambda)y$ is called a convex combination of x and y.

Proposition

If $0 \leq x, y, \lambda \leq 1$, *then* $0 \leq \lambda x + (1-\lambda)y \leq 1$

Proof.

$0 \leq x, y, \lambda$ *implies* $\lambda x + (1-\lambda)y \leq \lambda + (1-\lambda) = 1$.
$\lambda \leq 1$ *implies* $0 \leq 1 - \lambda$.
$x, y, \lambda, 1-\lambda \geq 0$ *implies* $\lambda x + (1-\lambda)y \geq 0$.
Therefore, $0 \leq \lambda x + (1-\lambda)y \leq 1$. $\square$

Structure of Decision Rules

Consider the structure of a decision rule $f_d(c)$.
Suppose $D = \{d^1, \ldots, d^Q\}$ and $C = \{c^1, \ldots, c^K\}$.
Then this decision rule f can be thought of as a vector in $\mathbb{R}^{KQ}$

$$f' = (f_{d^1}(c^1), \ldots, f_{d^1}(c^K), \ldots, f_{d^Q}(c^1), \ldots, f_{d^Q}(c^K))$$

There are some constraints:

- For $q \in \{1, \ldots, Q\}$ and $k \in \{1, \ldots, K\},\ 0 \leq f_{d^q}(c^k) \leq 1$
- For $q \in \{1, \ldots, Q\},\ \sum_{k=1}^{K} f_{d^q}(c^k) = 1$

Therefore, a decision rule must lie in the unit hypercube of $\mathbb{R}^{QK}$ and it must lie in the manifold defined by the Q linear constraints

$$\sum_{k=1}^{K} f_{d^q}(c^k) = 1$$

Convex Combinations of Decision Rules

Proposition

Convex combinations of decision rules are decision rules

Proof.

Let f and g be two decision rules. Let $0 \leq \lambda \leq 1$. Consider $\lambda f_d(c) + (1-\lambda) g_d(c)$. We have already proven that $0 \leq \lambda f_d(c) + (1-\lambda) g_d(c) \leq 1$. Consider the convex combination:

$$\begin{aligned}
\sum_{c \in C} [\lambda f_d(c) + (1-\lambda) g_d(c)] &= \lambda \sum_{c \in C} f_d(c) + (1-\lambda) \sum_{c \in C} g_d(c) \\
&= \lambda + (1-\lambda) \\
&= 1
\end{aligned}$$

□

Convex Sets

Definition

A set $C \subseteq \mathbb{R}^N$ is a convex set if and only if $x, y \in C$ imply $\lambda x + (1 - \lambda)y \in C$ for every $0 \leq \lambda \leq 1$.

Proposition

The set F of all convex combinations of decision rules is a convex set.

Intersection of Convex Sets are Convex

Proposition

Let C and D be convex sets. Then $C \cap D$ is a convex set.

Proof.

Let $x, y \in C \cap D$ and $0 \leq \lambda \leq 1$. Consider $\lambda x + (1 - \lambda)y$.
Since $x, y \in C \cap D$, $x, y \in C$ and $x, y \in D$.
Since C is convex and $0 \leq \lambda \leq 1$, $\lambda x + (1 - \lambda)y \in C$.
Since D is convex and $0 \leq \lambda \leq 1$, $\lambda x + (1 - \lambda)y \in D$.
$\lambda x + (1 - \lambda)y \in C$ and $\lambda x + (1 - \lambda)y \in D$ imply
$\lambda x + (1 - \lambda)y \in C \cap D$. □

Mixed Decision Rules

Definition

Let f and g be decision rules and $0 \leq \lambda \leq 1$.
Then
$$h_d(c) = \lambda f_d(c) + (1 - \lambda) g_d(c)$$
is called a mixed decision rule of f and g.

- With probability λ apply decision rule f and probability $1 - \lambda$ apply decision rule g.
- If we apply decision rule f, the we assign class c with probability $f(c|d)$
- If we apply decision rule g, then we assign class c with probability $g(c|d)$

Extreme Points

Definition

Let $A \subseteq \mathbb{R}^N$. A point $e \in A$ is called an Extreme Point of A if and only if $b, c \in A$ with $e = \frac{b+c}{2}$ implies $e = b = c$.

Deterministic Decision Rules are Extreme Points

Proposition

Let F be the set of all convex combinations of decision rules. Let f be a deterministic decision rule. Then f is an extreme point of F.

Proof.

Let $g, h \in F$ satisfy $f = \frac{g+h}{2}$. Hence for every $d \in D$ and $c \in C$,

$$f_d(c) = \frac{g_d(c) + h_d(c)}{2}$$

Since f is a deterministic decision rule, for some $c^* \in C$, $f_d(c^*) = 1$ and for all $c \in C - \{c^*\}$, $f_d(c) = 0$. Consider $c \in C$ for which $f_d(c) = 0$.

$$f_d(c) = 0 = \frac{g_d(c) + h_d(c)}{2}$$

Since $g_d(c), h_d(c) \geq 0$ and since $g_d(c) + h_d(c) = 0$, it follows that $g_d(c) = h_d(c) = 0$. □

Proof Continued

Proof.

Now consider c^*.

$$f_d(c^*) = 1 = \frac{g_d(c^*) + h_d(c^*)}{2}$$

Hence, $g_d(c^*) + h_d(c^*) = 2$. But $g_d(c^*), h_d(c^*) \leq 1$. Therefore, $g_d(c^*) = 1$ and $h_d(c^*) = 1$.

Now, by definition of extreme point, a deterministic decision rule $f \in F$ is an extreme point of F, the set of all convex combinations of decision rules. □

Convex Polyhedrons

Definition

A Closed Convex Polyhedron is a non-empty set P formed as the solutions to a matrix equation $Ax \leq b$.

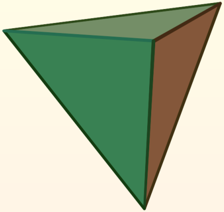
$$P = \{x \mid Ax \leq b\}$$

Each row of the matrix equation specifies a hyperplane half space and P is the intersection of these hyperplane half spaces.

Definition

A bounded polyhedron is a polytope.

Closed Convex Polytope Example Tetrahedron



$$P = \{x \mid Ax \leq b\}$$

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

$$b = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The Set of Decision Rules is a Closed Convex Polytope

Proposition

Let F be the set of all decision rules formed from the finite set C of classes and the finite set D of measurements.
The set F is a closed convex polytope lying in a linear manifold of dimension $|C|\ |D| - |D|$.

Proof.

Let $f \in F$. We already know that $f \in \mathbb{R}^{|C|\ |D|}$. The $|D|$ linear constraints are formed from the requirement that $\sum_{c \in C} f_d(c) = 1$. The remaining constraints are of the form

- $f_d(c) \geq 0$ *which is equivalent to* $-f_d(c) \leq 0$
- $f_d(c) \leq 1$

□

Minkowski's Theorem

Definition

Let $X = \{x_1, \ldots, x_M\} \subset \mathbb{R}^N$. The Convex Hull of X is defined by

$$\mathcal{CH}(X) = \{y \in \mathbb{R}^N \mid y = \sum_{m=1}^{M} \lambda_m x_m, \text{where } \lambda_m \geq 0, \sum_{m=1}^{M} \lambda_m = 1\}$$

Theorem

Any closed convex polytope is the convex hull of its extreme points.

Probabilistic Decision Rules

Any Probabilistic Decision Rule can be represented as a convex combination of the deterministic decision rules.

Theorem

Let f be a probabilistic decision rule and let $f^1, \ldots, f^M$ be the set of all possible deterministic decision rules. Then there exists a convex combination $\lambda_1, \ldots, \lambda_M$ such that

$$f_d(c) = \sum_{m=1}^{M} \lambda_m f_d^m(c)$$

Extreme Points Convex Sets

Proposition

Let $C \subseteq \mathbb{R}^N$ be a convex set. Let e be an extreme point of C. Let D be a convex subset of C. If $e \in D$, then e is an extreme point of D.

Proof.

Let e be an extreme point of C. Suppose $e \in D$. Let $a, b \in D$ satisfy $e = \frac{a+b}{2}$. Since $D \subseteq C$, $a, b \in C$. Now, $a, b \in D \subseteq C$, with $e = \frac{a+b}{2}$. Since e is an extreme point of C, $e = a = b$. But now we have $e \in D$ and $a, b \in D$ satisfying $e = \frac{a+b}{2}$. And we have just proved that $e = a = b$. Therefore, e is an extreme point of D. □

Expected Conditional Gain: Mixed Decision Rules

Proposition

$$E[e \mid c^j; \lambda f + (1-\lambda) g] = \lambda E[e \mid c^j; f] + (1-\lambda) E[e \mid c^j; g]$$

Proof.

$$\begin{aligned}
E[e \mid c^j; \lambda f + (1-\lambda) g] &= \sum_{k=1}^{K} \sum_{d \in D} e(c^j, c^k) P(d \mid c^j) \{\lambda f(c^k \mid d) + (1-\lambda) g(c^k \mid d)\} \\
&= \lambda \sum_{k=1}^{K} \sum_{d \in D} e(c^j, c^k) P(d \mid c^j) f(c^k \mid d) + \\
&\quad (1-\lambda) \sum_{k=1}^{K} \sum_{d \in D} e(c^j, c^k) P(d \mid c^j) g(c^k \mid d) \\
&= \lambda E[e \mid c^j; f] + (1-\lambda) E[e \mid c^j; g]
\end{aligned}$$

□

Example

e	Assigned	
True	c^1	c^2
c^1	2	-1
c^2	-1	2

$P(d \mid c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	.2	.3	.5
c^2	.5	.4	.1

$f_d(c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	1	0	0
c^2	0	1	1

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

$$\begin{aligned}
E[e \mid c^1; f] &= e(c^1, c^1)P(d^1 \mid c^1)f_{d^1}(c^1) + e(c^1, c^2)P(d^1 \mid c^1)f_{d^1}(c^2) + \\
&\quad e(c^1, c^1)P(d^2 \mid c^1)f_{d^2}(c^1) + e(c^1, c^2)P(d^2 \mid c^1)f_{d^2}(c^2 + \\
&\quad e(c^1, c^1)P(d^3 \mid c^1)f_{d^3}(c^1) + e(c^1, c^2)P(d^3 \mid c^1)f_{d^3}(c^2) \\
&= 2 * .2 * 1 + (-1) * .2 * 0 + \\
&\quad 2 * .3 * 0 + (-1) * .3 * 1 + \\
&\quad 2 * .5 * 0 + (-1) * .5 * 1 \\
&= .4 - .3 - .5 = -.4
\end{aligned}$$

Example

e	Assigned	
True	c^1	c^2
c^1	2	-1
c^2	-1	2

$P(d \mid c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	.2	.3	.5
c^2	.5	.4	.1

$f_d(c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	1	0	0
c^2	0	1	1

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

$$\begin{aligned}
E[e \mid c^2; f] &= e(c^2, c^1)P(d^1 \mid c^2)f_{d^1}(c^1) + e(c^2, c^2)P(d^1 \mid c^2)f_{d^1}(c^2) + \\
&\quad e(c^2, c^1)P(d^2 \mid c^2)f_{d^2}(c^1) + e(c^2, c^2)P(d^2 \mid c^2)f_{d^2}(c^2) + \\
&\quad e(c^2, c^1)P(d^3 \mid c^2)f_{d^3}(c^1) + e(c^2, c^2)P(d^3 \mid c^2)f_{d^3}(c^2) \\
&= (-1) * .5 * 1 + 2 * .5 * 0 + \\
&\quad (-1) * .4 * 0 + 2 * .4 * 1 + \\
&\quad (-1) * .1 * 0 + 2 * .1 * 1 \\
&= -.5 + .8 + .2 = .5
\end{aligned}$$

Example

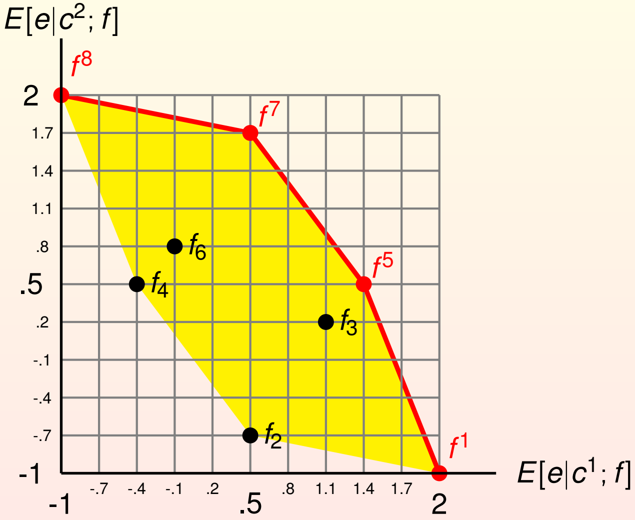
e	Assigned	
True	c^1	c^2
c^1	2	-1
c^2	-1	2

$P(d \mid c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	.2	.3	.5
c^2	.5	.4	.1

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

	Measurements			Conditional Gain	
f	d^1	d^2	d^3	$E[e \mid c^1; f]$	$E[e \mid c^2; f]$
f^1	c^1	c^1	c^1	2.0	-1.0
f^2	c^1	c^1	c^2	.5	-.7
f^3	c^1	c^2	c^1	1.1	.2
f^4	c^1	c^2	c^2	-.4	.5
f^5	c^2	c^1	c^1	1.4	.5
f^6	c^2	c^1	c^2	-.1	.8
f^7	c^2	c^2	c^1	.5	1.7
f^8	c^2	c^2	c^2	-1.	2.0

Conditional Expected Gains: All Decision Rules



Convex Sets and Linear Functions

Proposition

Images of linear functions of convex sets are convex.

Proof.

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* $y_2 = f(x_2)$.

$$\begin{aligned} \lambda y_1 + (1 - \lambda y_2 &= \lambda f(x_1) + (1 - \lambda) f(x_2) \\ &= f(\lambda x_1 + (1 - \lambda) x_2) \end{aligned}$$

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* $\lambda y_1 + (1 - \lambda y_2 \in D$ □

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.

$$
\begin{aligned}
E[e; f] &= \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j) \\
&= \sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) + E[e \mid c^K; f](1 - \sum_{j=1}^{K-1} P(c^j)) \\
&= \sum_{j=1}^{K-1} \{E[e \mid c^j; f] - E[e \mid c^K; f]\} P(c^j) + E[e \mid c^K; f]
\end{aligned}
$$

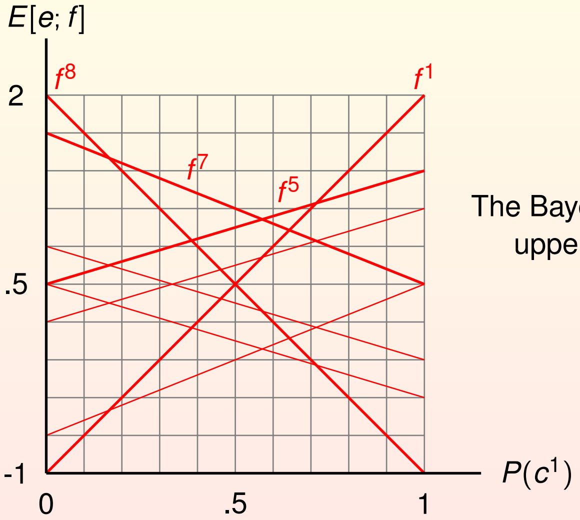
□

Dependence on Prior Class Probabilities

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

	Measurements			Conditional Gain		Expected Gain
f	d^1	d^2	d^3	$E[e\|c^1;f]$	$E[e\|c^2;f]$	$E[e;f,P(c^1)]$
f^1	c^1	c^1	c^1	2.0	-1.0	$3P(c^1) - 1$
f^2	c^1	c^1	c^2	.5	-.7	$1.2P(c^1) - .7$
f^3	c^1	c^2	c^1	1.1	.2	$.9P(c^1) + .2$
f^4	c^1	c^2	c^2	-.4	.5	$-.9P(c^1) + .5$
f^5	c^2	c^1	c^1	1.4	.5	$.9P(c^1) + .5$
f^6	c^2	c^1	c^2	-.1	.8	$-.9P(c^1) + .8$
f^7	c^2	c^2	c^1	.5	1.7	$-1.2P(c^1) + 1.7$
f^8	c^2	c^2	c^2	-1.	2.0	$-3P(c^1) + 2.0$

Dependence on Class Prior Probabilities



The Bayes Gain is the upper envelope

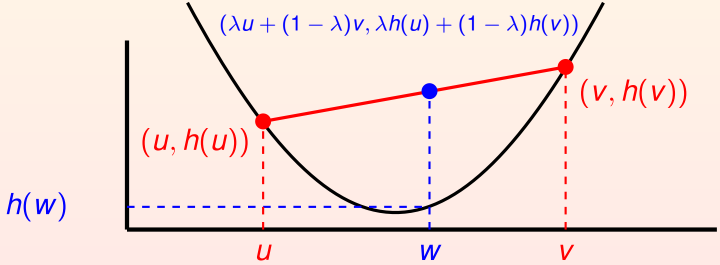
Convex Functions

Definition

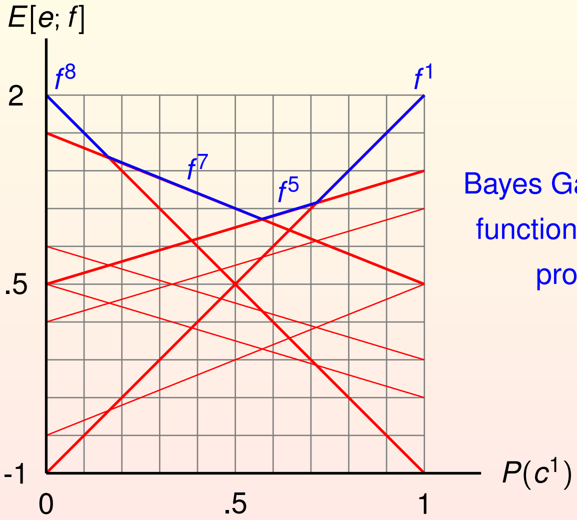
A function h, $h : \mathbb{R}^N \to \mathbb{R}$, is a convex function if and only if for every λ, $0 \leq \lambda \leq 1$,

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

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



Bayes Gain is Convex



Bayes Gain is a convex function of class prior probabilities

Bayes Gain Is Convex

$$
\begin{aligned}
E[e;f] &= \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j)] \\
G_B &= \max_f E[e;f] \text{ Bayes Gain}
\end{aligned}
$$

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

$$
\begin{aligned}
a_{jn} &= E[e \mid c^j; f^n] \\
p_j &= P(c^j) \\
G_B(P(c^1), \ldots, P(c^K)) &= \max_n \sum_{j=1}^{K} E[e \mid c^j; f^n] P(c^j) \\
G_B(p_1, \ldots, p_K) &= \max_n \sum_{j=1}^{K} a_{jn} p_j
\end{aligned}
$$

Bayes Gain Is Convex

Theorem

Let $p = (p_1, \ldots, p_K)$ *and* $q = (q_1, \ldots q_K)$. *Let* $0 \le \lambda \le 1$.

$$G_B(\lambda p + (1-\lambda)q) \quad \le \quad \lambda G_B(p) + (1-\lambda)G_B(q)$$

Proof.

$$\begin{aligned}
G_B(\lambda p + (1-\lambda)q) &= \max_n \sum_{j=1}^{K} a_{jn}(\lambda p_j + (1-\lambda)q_j) \\
&= \max_n \left\{ \lambda \sum_{j=1}^{K} a_{jn}p_j + (1-\lambda) \sum_{j==1}^{K} a_{jn}q_j \right\} \\
&\le \left[\max_n \lambda \sum_{j=1}^{K} a_{jn}p_j \right] + \left[\max_n (1-\lambda) \sum_{j=1}^{K} a_{jn}q_j \right] \\
&\le \lambda G_B(p) + (1-\lambda)G_B(q)
\end{aligned}$$

□

Epigraph and Convex Sets

Definition

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

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

Epigraph and Convex Sets

Proposition

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

Proof.

Suppose f *is convex. Let* $(x, u), (y, v) \in Epi(f)$ *and* $0 \leq \lambda \leq 1$. *Then by definition of* $Epi(f)$, $f(x) \leq u$, $f(y) \leq v$ *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* $Epi(f)$, $(\lambda x + (1-\lambda)y, \lambda u + (1-\lambda)v) \in Epi(f)$ *making* $Epi(f)$ *convex.* □

Epigraph and Convex Sets

Proposition

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

Proof.

Suppose $Epi(f)$ *is a convex set. Then by definition of* $Epi(f)$, $(x, f(x)) \in Epi(f)$ *and* $(y, f(y)) \in Epi(f)$. *Since* $Epi(f)$ *is convex,* $\lambda(x, f(x)) + (1-\lambda)(y, f(y)) \in Epi(f)$. *Hence* $(\lambda x + (1-\lambda y), \lambda f(x) + (1-\lambda)f(y)) \in Epi(f)$. *By definition of* $Epi(f)$, $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$. *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 $f : \mathbb{R}^N \to \mathbb{R}$ and $c \in \mathbb{R}$. A basin set of f is any set of the form

$$L = \{x \in \mathbb{R}^N \mid f(x) \leq c\}$$

Theorem

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

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

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

Minima Set of A Convex Function is Convex

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 $M = \{x \in C \mid h(x) \leq b\}$. C being closed and bounded is needed because the minima of h may be on the boundary. □

For Convex Functions Local Minima are Global Minima

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

$$\begin{aligned} h(x_0) &\leq h((1-\alpha)x_0 + \alpha x) \leq (1-\alpha)h(x_0) + \alpha h(x) \\ 0 &\leq \alpha(h(x) - h(x_0)) \\ h(x_0) &\leq h(x) \end{aligned}$$

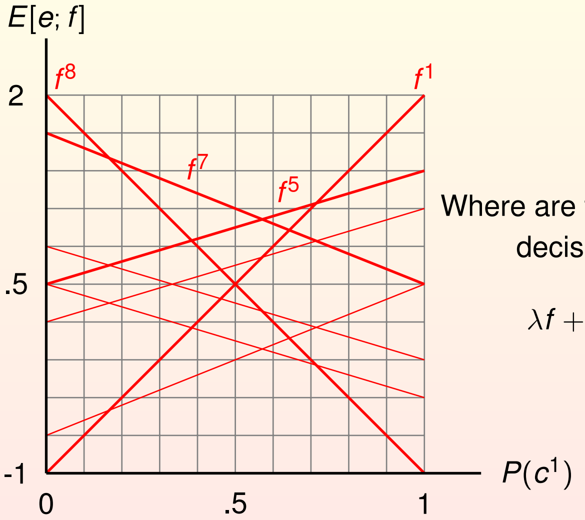
□

Dependence on Prior Class Probabilities

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

	Measurements			Conditional Gain		Expected Gain
f	d^1	d^2	d^3	$E[e\|c^1;f]$	$E[e\|c^2;f]$	$E[e;f,P(c^1)]$
f^1	c^1	c^1	c^1	2.0	-1.0	$3P(c^1) - 1$
f^2	c^1	c^1	c^2	.5	-.7	$1.2P(c^1) - .7$
f^3	c^1	c^2	c^1	1.1	.2	$.9P(c^1) + .2$
f^4	c^1	c^2	c^2	-.4	.5	$-.9P(c^1) + .5$
f^5	c^2	c^1	c^1	1.4	.5	$.9P(c^1) + .5$
f^6	c^2	c^1	c^2	-.1	.8	$-.9P(c^1) + .8$
f^7	c^2	c^2	c^1	.5	1.7	$-1.2P(c^1) + 1.7$
f^8	c^2	c^2	c^2	-1.	2.0	$-3P(c^1) + 2.0$

Dependence on Class Prior Probabilities



Where are the probabilistic decision rules?

$$\lambda f + (1 - \lambda) g$$

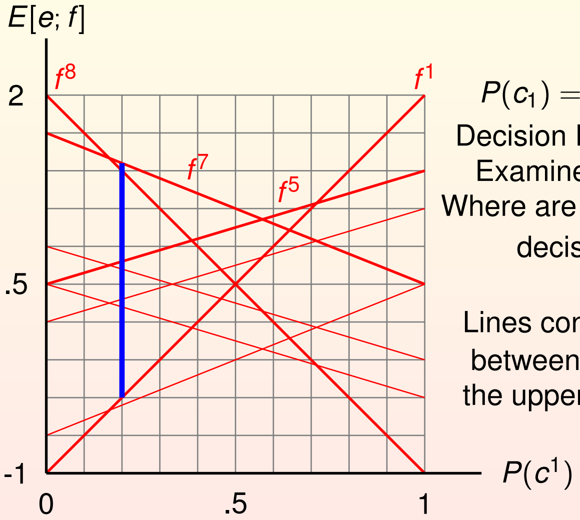
Probabilistic Decision Rules

- 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



$P(c_1) = .2$

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

Probabilistic Decision Rules Are In Between

Expected gain of a mixed decision rule is the mixture of the expected gains of the component decision rules.

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

Probabilistic Decision Rules Are In Between

Two Class Case

Proposition

Let $0 \le \lambda \le 1$*. If* $E[e; f_1, P(c^1)]$ *and* $E[e; f_2, P(c^1)]$ *are affine functions of* $P(c^1)$*, then* $E[e; \lambda f_1 + (1-\lambda) f_2, P(c^1)]$ *is an affine function of* $P(c^1)$*.*

Proof.

$$
\begin{aligned}
E[e; f_1, P(c^1)] &= \alpha_1 P(c^1) + \beta_1 \\
E[e; f_2, P(c^1)] &= \alpha_2 P(c^1) + \beta_2 \\
E[e; \lambda f_1 + (1-\lambda) f_2, P(c^1)] &= \lambda(\alpha_1 P(c^1) + \beta_1) + (1-\lambda)(\alpha_2 P(c^1) + \beta_2) \\
&= (\lambda\alpha_1 + (1-\lambda)\alpha_2) P(c^1) + \lambda\beta_1 + (1-\lambda)\beta_2
\end{aligned}
$$

□

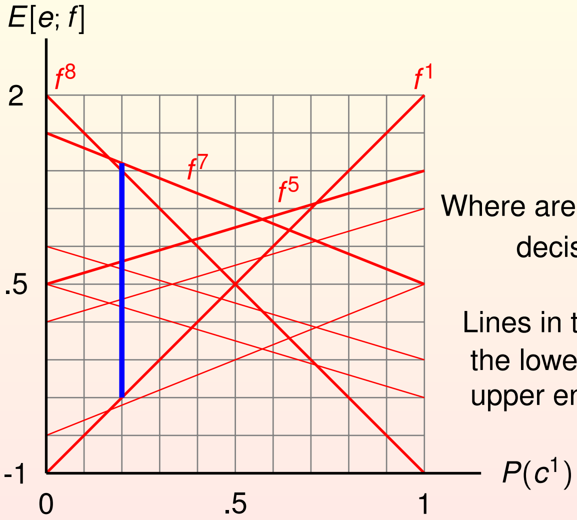
Probabilistic Decision Rules Are In Between

Corollary

Fix $P(c^1)$. *Let* $0 \leq \lambda \leq 1$.
If $\alpha_1 P(c^1) + \beta_1 < \alpha_2 P(c^1) + \beta_2$ *then*

$$\alpha_1 P(c^1) + \beta_1 \leq (\lambda\alpha_1 + (1-\lambda)\alpha_2)P(c^1) + \lambda\beta_1 + (1-\lambda)\beta_2 \leq \alpha_2 P(c^1) + \beta_2$$

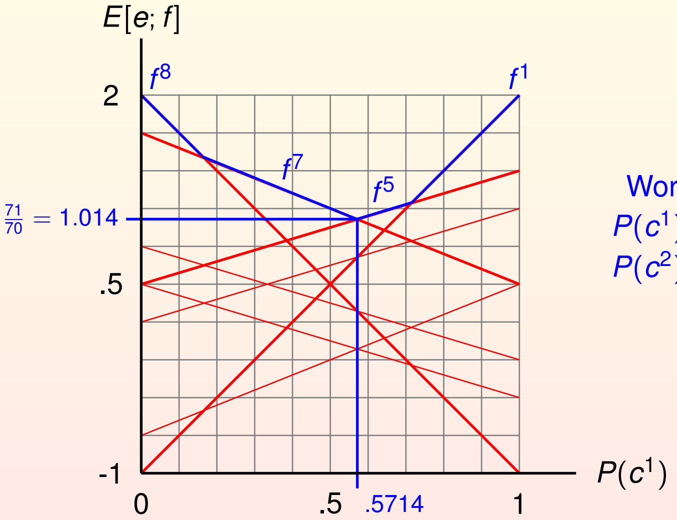
Dependence on Class Prior Probabilities



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

$P(c^1) = 4/7 \approx .5714$

$P(c^2) = 3/7 \approx .4286$

Finding Worst Class Priors

Two Class Case
Decision Rules of Mixture are Known

$$
\begin{aligned}
E[e; f_5; P(c^1)] &= .9P(c^1) + .5 \\
E[e; f_7; P(c^1)] &= -1.2P(c^1) + 1.7 \\
\text{Set } E[e; f_5; P(c^1)] &= E[e; f_7; P(c^1)] \\
.9P(c^1) + .5 &= -1.2P(c^1) + 1.7; \\
2.1P(c^1) &= 1.2 \\
P(c^1) &= \frac{1.2}{2.1} = \frac{4}{7} \\
P(c^2) &= 1 - P(c^1) = \frac{3}{7}
\end{aligned}
$$

Determining the Maximin Decision Rule

$$
\begin{aligned}
E[e;f] &= E[e|c^1;f]P(c^1) + E[e|c^2;f]P(c^2) \\
&= E[e|c^1;f]P(c^1) + E[e|c^2;f](1 - P(c^1)) \\
&= (E[e|c^1;f] - E[e|c^2;f])P(c^1) + E[e|c^2;f]
\end{aligned}
$$

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

$$
\begin{aligned}
E[e|c^1;f] - E[e|c^2;f] &= 0 \\
E[e|c^1;f] &= E[e|c^2;f]
\end{aligned}
$$

In this case,

$$
\begin{aligned}
E[e;f] &= E[e|c^1;f] \\
&= E[e|c^2;f]
\end{aligned}
$$

Dependence of a Probabilistic Decision Rule on Priors

$$
\begin{aligned}
E[e; \lambda f^5 + (1-\lambda) f^7] &= E[e|c^1; \lambda f^5 + (1-\lambda) f^7] P(c^1) + E[e|c^2; \lambda f^5 + (1-\lambda) f^7] P(c^2) \\
&= (\lambda E[e|c^1; f^5] + (1-\lambda) E[e|c^1; f^7]) P(c^1) + \\
&\qquad (\lambda E[e|c^2; f^5] + (1-\lambda) E[e|c^2; f^7])(1 - P(c^1)) \\
&= ((\lambda E[e|c^1; f^5] + (1-\lambda) E[e|c^1; f^7]) - (\lambda E[e|c^2; f^5] + (1-\lambda) E[e|c^2; f^7]) P(\\
&\qquad E[e|c^2; \lambda f^5 + (1-\lambda) f^7]
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
\lambda &= \frac{E[e|c^2; f^7] - E[e|c^1; f^7]}{E[c^1; f^5] - E[c^1; f^7] - E[e|c^2; f^5] + E[e|c^2; f^7]} \\
&= \frac{1.7 - .5}{1.4 - .5 - .5 + 1.7} = \frac{1.2}{2.1} = \frac{4}{7}
\end{aligned}
$$

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

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

$\lambda \leq 1$ implies

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

Finding Worst Class Priors

Three Class Case

$$
\begin{aligned}
E[e; f_i; P(c^1), P(c^2)] &= \alpha_{i1} P(c^1) + \alpha_{i2} P(c^2) + \gamma_i, \;\; i = 1, 2, 3 \\
E[e; f_i; P(c^1), P(c^2)] &= E[e; f_3; P(c^1), P(c^2)], \;\; i = 1, 2
\end{aligned}
$$

$$
\begin{aligned}
\alpha_{11} P(c^1) + \alpha_{12} P(c^2) + \gamma_1 &= \alpha_{31} P(c^1) + \alpha 32 P(c^2) + \gamma_3 \\
\alpha_{21} P(c^1) + \alpha_{22} P(c^2) + \gamma_2 &= \alpha_{31} P(c^1) + \alpha 32 P(c^2) + \gamma_3
\end{aligned}
$$

$$
\begin{pmatrix} \alpha_{11} - \alpha_{31} & \alpha_{12} - \alpha_{32} \\ \alpha_{21} - \alpha_{31} & \alpha_{22} - \alpha_{32} \end{pmatrix} \begin{pmatrix} P(c^1) \\ P(c^2) \end{pmatrix} = \begin{pmatrix} \gamma_1 - \gamma_3 \\ \gamma_2 - \gamma_3 \end{pmatrix}
$$

Finding Worst Class Priors

N Class Case

$$E[e; f_n; P(c^1), \ldots, P(c^{N-1})] = \sum_{i=1}^{N-1} \alpha_{ni} P(c^i) + \gamma_n, \; n = 1, \ldots, N$$

$$E[e; f_n; P(c^1), \ldots, P(c^{N-1})] = E[e; f_N; P(c^1), \ldots, P(c^{N-1})], \; i = 1, \ldots, N-1$$

$$\begin{pmatrix} \alpha_{11} - \alpha_{N1} & \alpha_{12} - \alpha_{N2} & \cdots & \alpha_{1N-1} - \alpha_{NN-1} \\ & & \vdots & \\ \alpha_{N-11} - \alpha_{N1} & \alpha_{N-12} - \alpha_{N2} & \cdots & \alpha_{N-1N-1} - \alpha_{NN-1} \end{pmatrix} \begin{pmatrix} P(c^1) \\ \vdots \\ P(c^{N-1}) \end{pmatrix} = \begin{pmatrix} \gamma_1 - \gamma_N \\ \vdots \\ \gamma_{N-1} - \gamma_N \end{pmatrix}$$

Finding The Convex Combination

Two Class Case

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 $P(c^1)$.

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

$$\begin{aligned}(\lambda\alpha_{11} + (1-\lambda)\alpha_{21}) P(c^1) &= \alpha_{21} P_0(c^1) + \gamma_2 - \lambda\gamma_1 - (1-\lambda)\gamma_2 \\ &= \alpha_{21} P_0(c^1) - \lambda(\gamma_1 + \gamma_2)\end{aligned}$$

Therefore, $\lambda\alpha_{11} + (1-\lambda)\alpha_{21} = 0$ and $\lambda = \frac{-\alpha_{21}}{\alpha_{11} - \alpha_{21}}$.

Finding The Convex Combination

Two Class Case
Identity in $P(c^1)$ meaning For all $P(c^1)$

$$
\begin{aligned}
0 \leq \lambda_1, \lambda_2 \leq 1 \\
\lambda_1 + \lambda_2 = 1 \\
\lambda_1(\alpha_{11} P(c^1) + \gamma_1) + \lambda_2(\alpha_{21} P(c^1) + \gamma_2) &= \alpha_{21} P_0(c^1) + \gamma_2
\end{aligned}
$$

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

This implies

$$
\begin{aligned}
\lambda_1 \alpha_{11} + \lambda_2 \alpha_{21} &= 0 \\
\lambda_1 \gamma_1 + \lambda_2 \gamma_2 &= \alpha_{21} P_0(c^1) + \gamma_2 \\
\lambda_1 + \lambda_2 &= 1
\end{aligned}
$$

Finding the Convex Combination

$$
\begin{aligned}
\lambda_1 \alpha_{11} + \lambda_2 \alpha_{21} &= 0 \\
\lambda_1 \gamma_1 + \lambda_2 \gamma_2 &= \alpha_{21} P_0(c^1) + \gamma_2 \\
\lambda_1 + \lambda_2 &= 1
\end{aligned}
$$

$$
\begin{aligned}
\lambda_2 &= -\lambda_1 \frac{\alpha_{11}}{\alpha_{21}} \\
\lambda_1 + \lambda_2 &= \lambda_1 \left(1 - \frac{\alpha_{11}}{\alpha_{21}}\right) = 1 \\
\lambda_1 &= \frac{\alpha_{21}}{\alpha_{21} - \alpha_{11}}
\end{aligned}
$$

Finding the Convex Combination: Consistency Check

$$\begin{aligned} 0 \leq \quad \lambda_1, \lambda_2 \quad &\leq 1 \\ \lambda_1 \quad &= \quad \frac{\alpha_{21}}{\alpha_{21} - \alpha_{11}} \end{aligned}$$

Either $\alpha_{21} - \alpha_{11} > 0$ *or* < 0.
If $\alpha_{21} - \alpha_{11} > 0$ then

$$\begin{aligned} \alpha_{21} &> \alpha_{11} \\ \alpha_{21} &> 0 \end{aligned}$$

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

$$\begin{aligned} \alpha_{21} &< \alpha_{11} \\ \alpha_{21} &< 0 \end{aligned}$$

Finding the Convex Combination

Once λ_1 and λ_2 are known, the exact value for $P_0(c^1)$ can be determined.

$$
\begin{aligned}
\lambda_1\gamma_1 + (1-\lambda_1)\gamma_2 &= \alpha_{21}P_0(c^1) + \gamma_2 \\
\lambda_1(\gamma_1 - \gamma_2) &= \alpha_{21}P_0(c^1) \\
P_0(c^1) &= \frac{\lambda_1(\gamma_1 - \gamma_2)}{\alpha_{21}} \\
&= \frac{\alpha_{21}}{\alpha_{21} - \alpha_{11}} \frac{\gamma_1 - \gamma_2}{\alpha_{21}} \\
&= \frac{\gamma_1 - \gamma_2}{\alpha_{21} - \alpha_{11}}
\end{aligned}
$$

Finding The Convex Combination

N Class Case
Identity in $P(c^1)\ldots,P(c^{N-1})$

$$\begin{aligned}
\sum_{n=1}^{N}\lambda_n\left(\sum_{i=1}^{N-1}\alpha_{in}P(c^i)+\gamma_n\right) &= \sum_{i=1}^{N}\alpha_{Ni}P_0(c^i)+\gamma_N \\
\sum_{i=1}^{N-1}\left(\sum_{n=1}^{N}\lambda_n\alpha_{in}\right)P(c^i) &= \sum_{i=1}^{N}\alpha_{Ni}P_0(c^i)+\gamma_N-\sum_{n=1}^{N}\lambda_n\gamma_n
\end{aligned}$$

Implies

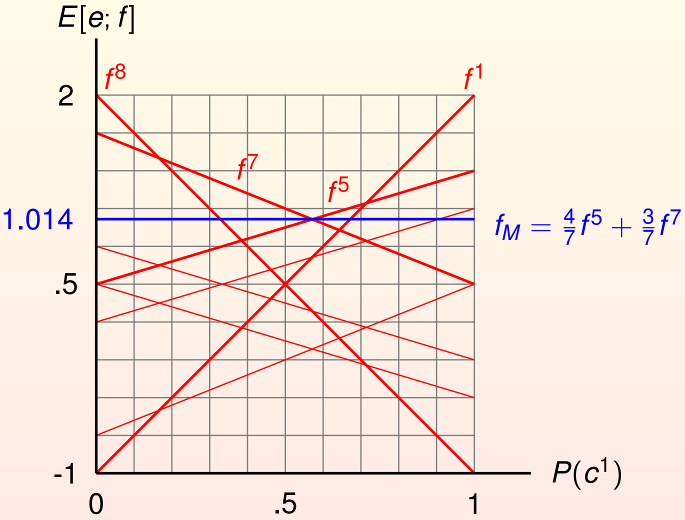
$$\begin{aligned}
\sum_{n=1}^{N}\lambda_n\alpha_{in} &= 0,\ i=1,\ldots,N-1 \\
\sum_{n=1}^{N}\lambda_n &= 1
\end{aligned}$$

Finding The Convex Combination

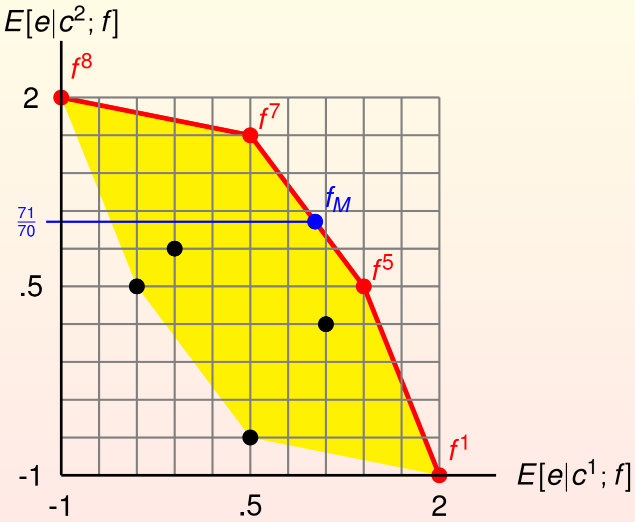
- Each component decision rule of the mixture has an expected gain that is a hyperplane in the axes $P(c^1)\ldots, P(c^{N-1})$
- The first $N-1$ rows of the $i^t h$ column consists of the coefficients of $P(c^1)\ldots, P(c^{N-1})$ for the i^{th} hyperplane

$$\begin{pmatrix} \alpha_{11} & \alpha_{21} & \ldots & \alpha_{N1} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{N2} \\ & & \vdots & \\ \alpha_{N-11} & \alpha_{N-12} & \ldots & \alpha_{N-1N} \\ 1 & 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Dependence on Class Prior Probabilities



Conditional Expected Gains: All Decision Rules



Two Entity Game

The game is played for a large number of trials.

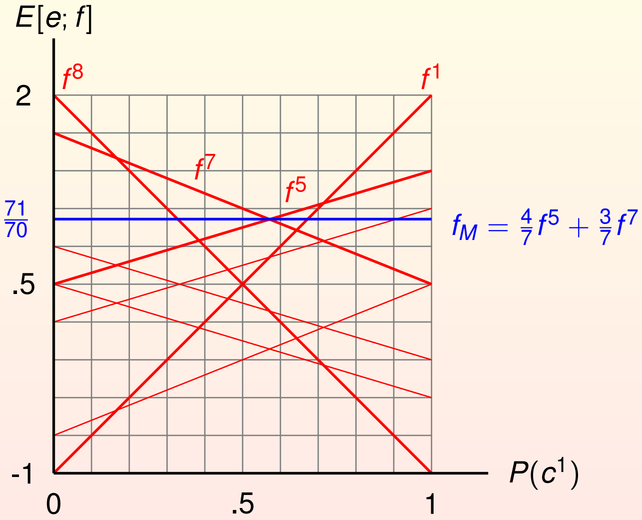
- Nature chooses class c in accordance with class priors $P(c^1) \ldots, P(c^K)$
- A measurement d is sampled in accordance with $P(d \mid c)$
- Bayes chooses decision rule to maximize expected gain under given class priors

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.

Maximin Decision Rule

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



Maximin Decision Rule

Definition

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

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

Maximin Decision Rule

Theorem

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

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

for any decision rule g.

Theorem

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

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

for any decision rule g.

Proof.

Recall

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

□

Maximin Decision Rule

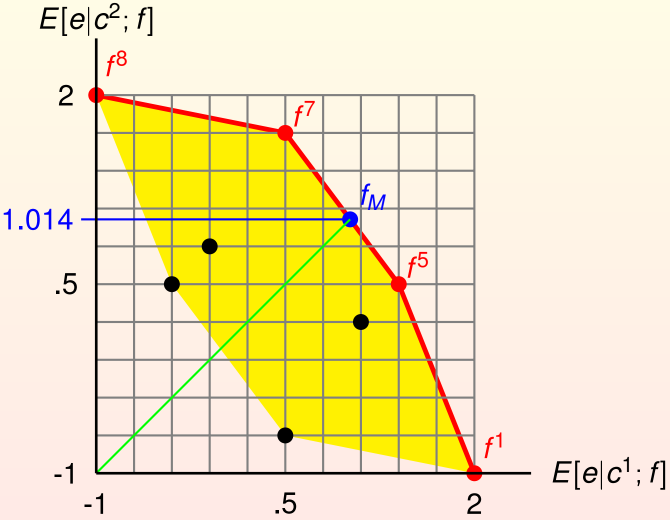
A decision rule f is a maximin decision rule if and only if the expected gain of f is the same as the expected gain of the Bayes rule under the worst possible prior class probabilities.

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



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, $m = 1, \ldots, M$
  - $G_w = \sum_{j=1}^{K} E[e \mid c^j; f^m] P(c^j)$
- Find convex combination $\lambda_1, \ldots, \lambda_M$
  - $G_w = E[e \mid c^k; \sum_{m=1}^{M} \lambda_m f^m] = \sum_{m=1}^{M} \lambda_m E[e \mid c^j; f^m], \; j = 1, \ldots, K$
- Let $a_{jm} = E[e \mid c^j; f^m]$
- 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

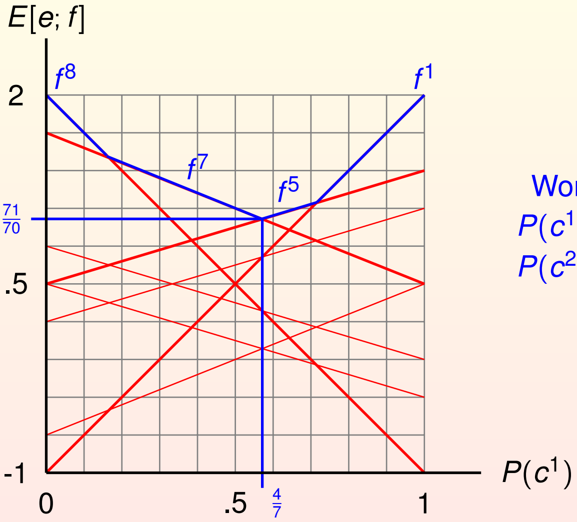
Let a_{jm} be a real numbers, $j = 1, \ldots, K; m = 1, \ldots, M$. Let $p_j \geq 0$ and $\sum_{j=1}^{K} p_j = 1$. Suppose

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

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

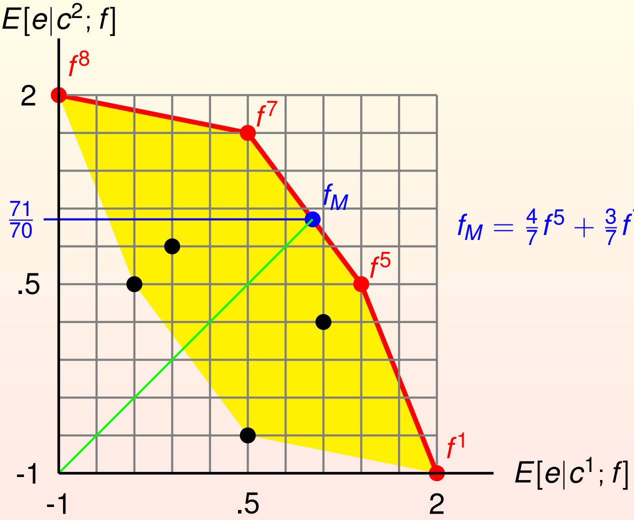


Worst Class Priors

$P(c^1) = 4/7 \approx .5714$

$P(c^2) = 3/7 \approx .4286$

Maximin Decision Rule



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

Find Worst Priors

- Find priors $p_1^0, \ldots, p_K^0$ satisfying

$$max_f E[e; f, p_1^0, \ldots, p_K^0] \leq max_g E[e; g, p_1, \ldots, p_K]$$

  - where f must be a Bayes Decision Rule
  - where g must be a Bayes Decision Rule
  - where $\sum_{k=1}^K p_k = 1$ and $p_k \geq 0$
- Choose priors $q_1^0, \ldots, q_K^0$
- Determine the Bayes Gain corresponding to $q_1^0, \ldots, q_K^0$
- Iterate to fixed point
  $(q_1^n, \ldots, q_K^n) = (q_1^{n-1}, \ldots, q_K^{n-1}) + (\pi_1, \ldots, \pi_K)$
  - where $\sum_{k=1}^K \pi_k = 0$
  - so that $\sum_{k=1}^K q_k^n = 1$ and $0 \leq q_k^n \leq 1$
  - so that over all Bayes decision rules f and g

$$max_f E[e; f, q_1^n, \ldots, q_K^n] \leq max_g E[e; f, q_1^{n-1}, \ldots, q_K^{n-1}]$$

- Let $p_1^0, \ldots, p_K^0$ be the worst priors
- Find all deterministic Bayes decision rules $f^1, \ldots, f^M$ such that for each $m = 1, \ldots, M$,

$$G_w = E[e; f^m, p_1^0, \ldots, p_K^0] = max_f E[e; f, p_1^0, \ldots, p_K^0]$$

- Find $\lambda_1, \ldots, \lambda_M$, $0 \leq \lambda_m \leq 1$ and $\sum_{m=1}^{M} \lambda_m = 1$ to satisfy

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

Maximin Decision Rules

- Find $\lambda_1, \ldots, \lambda_M$ to satisfy $G_w = \sum_{m=1}^{M} \lambda_m E[e \mid c^j; f^m]$
- Let $\alpha_{jm} = E[e \mid c^j; f^m]$
- Find $\lambda_1, \ldots, \lambda_M$ to satisfy

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1M} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2M} \\ & & \vdots & \\ \alpha_{M1} & \alpha_{M2} & \ldots & \alpha_{MM} \\ 1 & 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_M \end{pmatrix} = \begin{pmatrix} G_w \\ G_w \\ \vdots \\ G_w \\ 1 \end{pmatrix}$$

Maximin Decision Rule

- Let $\alpha_{jm} = E[e \mid c^j; f^m] - E[e \mid c^j; f^M]$
- Find $\lambda_1, \ldots, \lambda_M$ to satisfy

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1M} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2M} \\ & & \vdots & \\ \alpha_{M-11} & \alpha_{M-12} & \cdots & \alpha_{M-1M} \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_M \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

- The solution is valid only if $0 \leq \lambda_1, \ldots, \lambda_M \leq 1$