NIKON CORPORATION
Core Technology Center

# NIPS2007論文紹介:
# Optimal ROC Curve for a Combination of Classifiers

著者:Marco Barreno, Alvaro A. Cardenas, J. D. Tygar
読む人:中島伸一

January 18, 2008

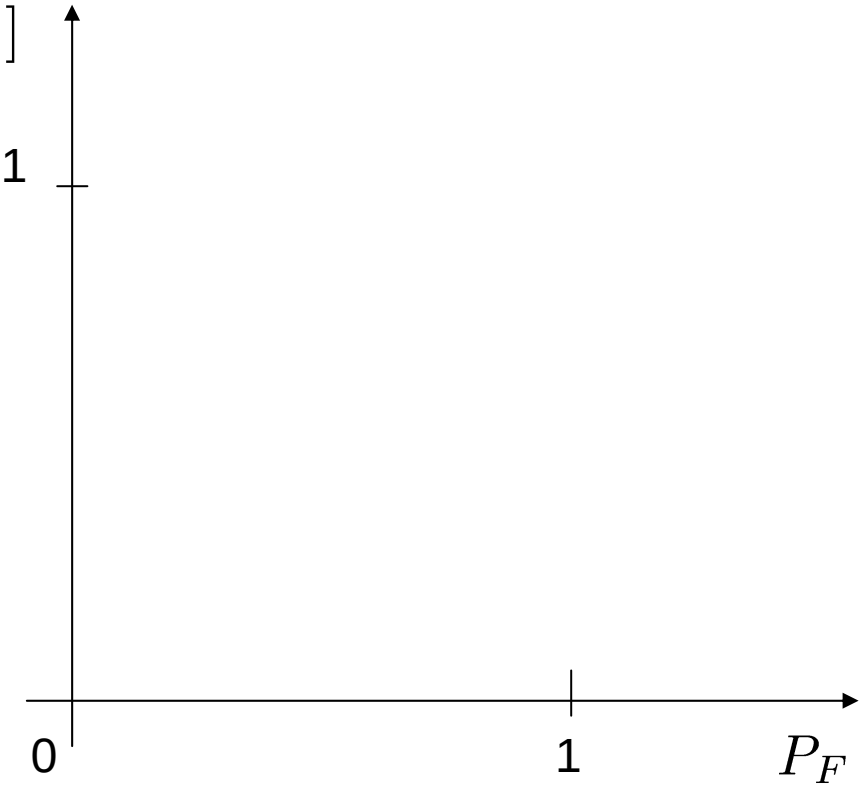
# 特徴

- binary classifierが複数あるときのNeyman-Peason的最強検定法
- 2^(2^n)通りの可能なmeta-classifierのうち，2^n個のoptimalを見つける.
- それでも2^n通りの計算が必要なので，しんどい.

# Receiver Operating Characteristic (ROC) curve

$P_D = P[Y = 1|\mathcal{H}_1]$

Decision rate,
True positive,
Power



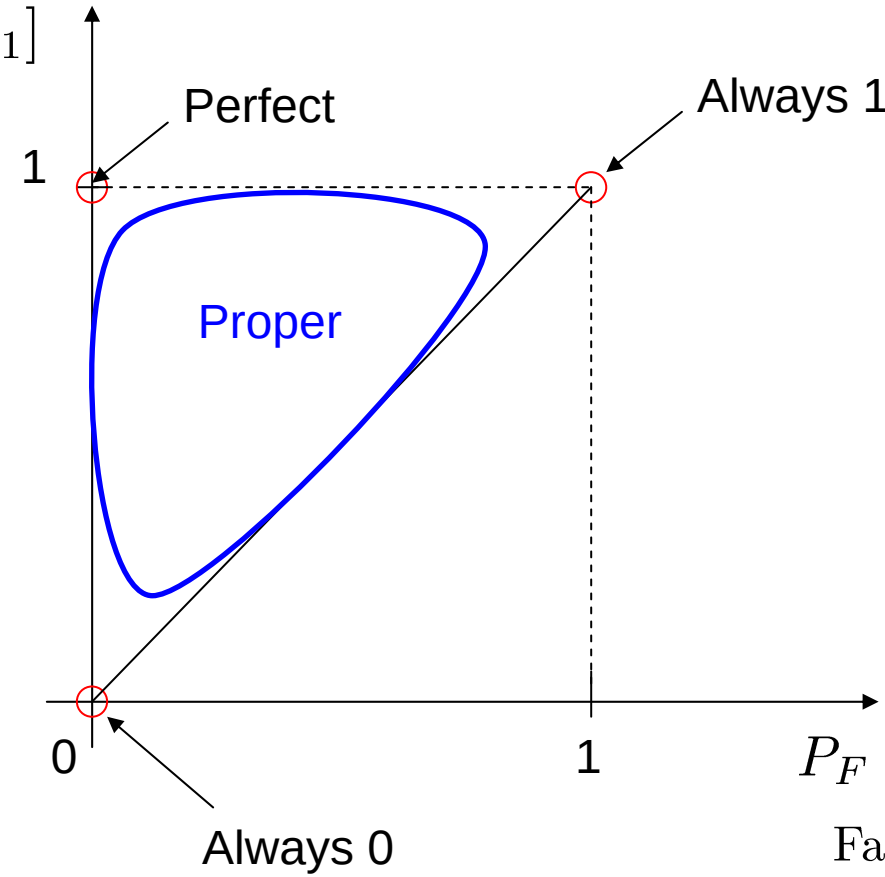

$P_F = P[Y = 1|\mathcal{H}_0]$

False alarm rate,
False positive,
Size

# Receiver Operating Characteristic (ROC) curve

$P_D = P[Y = 1|\mathcal{H}_1]$

Decision rate,
True positive,
Power




$P_F = P[Y = 1|\mathcal{H}_0]$

False alarm rate,
False positive,
Size

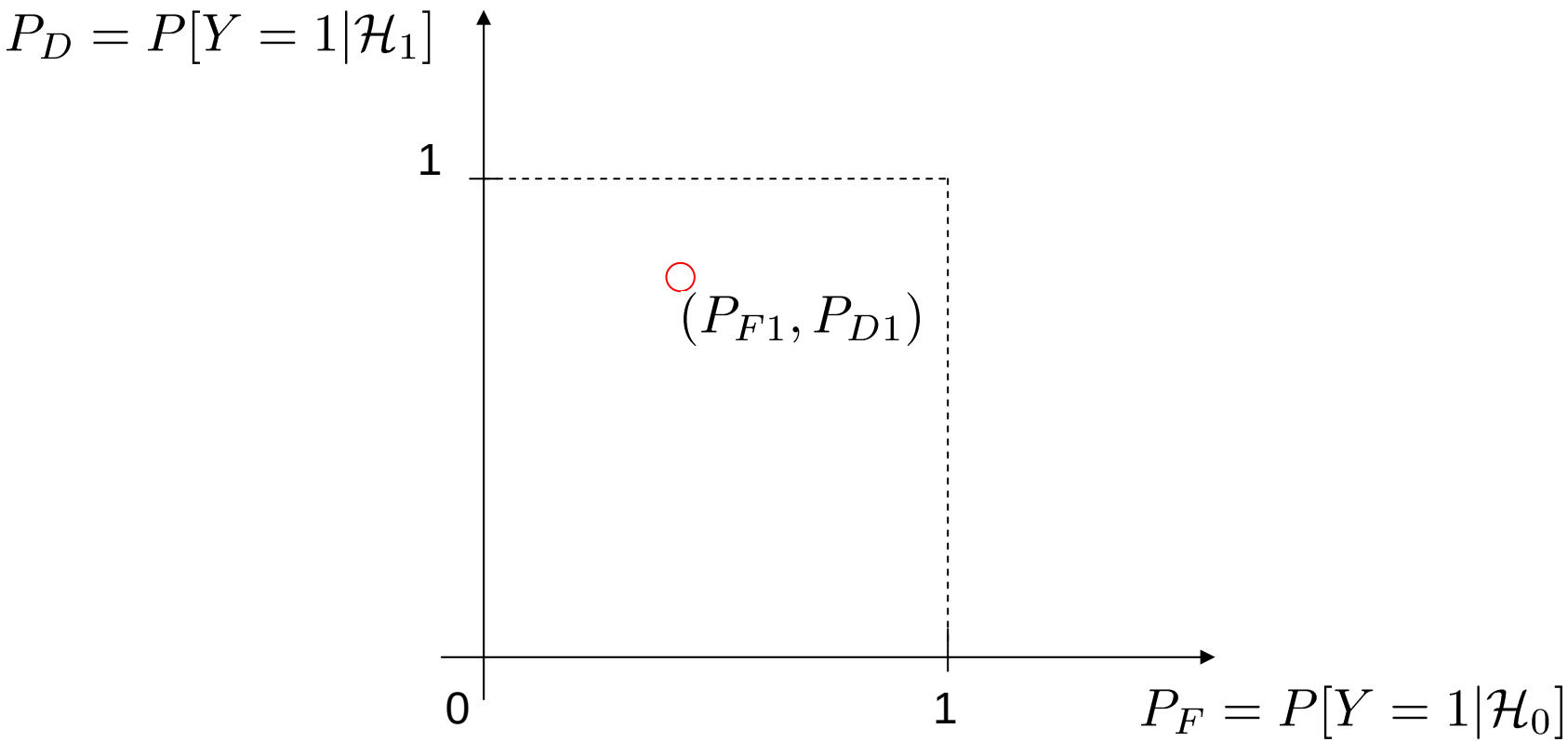
# Neyman Pearsonの補題

We wish to test a null hypothesis $H_0$ against an alternative $H_1$. Let the random variable $\mathbf{Y}$ have probability distributions $P(\mathbf{Y}|H_0)$ under $H_0$ and $P(\mathbf{Y}|H_1)$ under $H_1$, and define the *likelihood ratio* $\ell(\mathbf{Y}) = P(\mathbf{Y}|H_1)/P(\mathbf{Y}|H_0)$. The Neyman-Pearson lemma states that the likelihood ratio test

$$\mathcal{D}(\mathbf{Y}) = \begin{cases} 1 & \text{if } \ell(\mathbf{Y}) > \tau \\ \gamma & \text{if } \ell(\mathbf{Y}) = \tau \\ 0 & \text{if } \ell(\mathbf{Y}) < \tau \end{cases}, \tag{1}$$

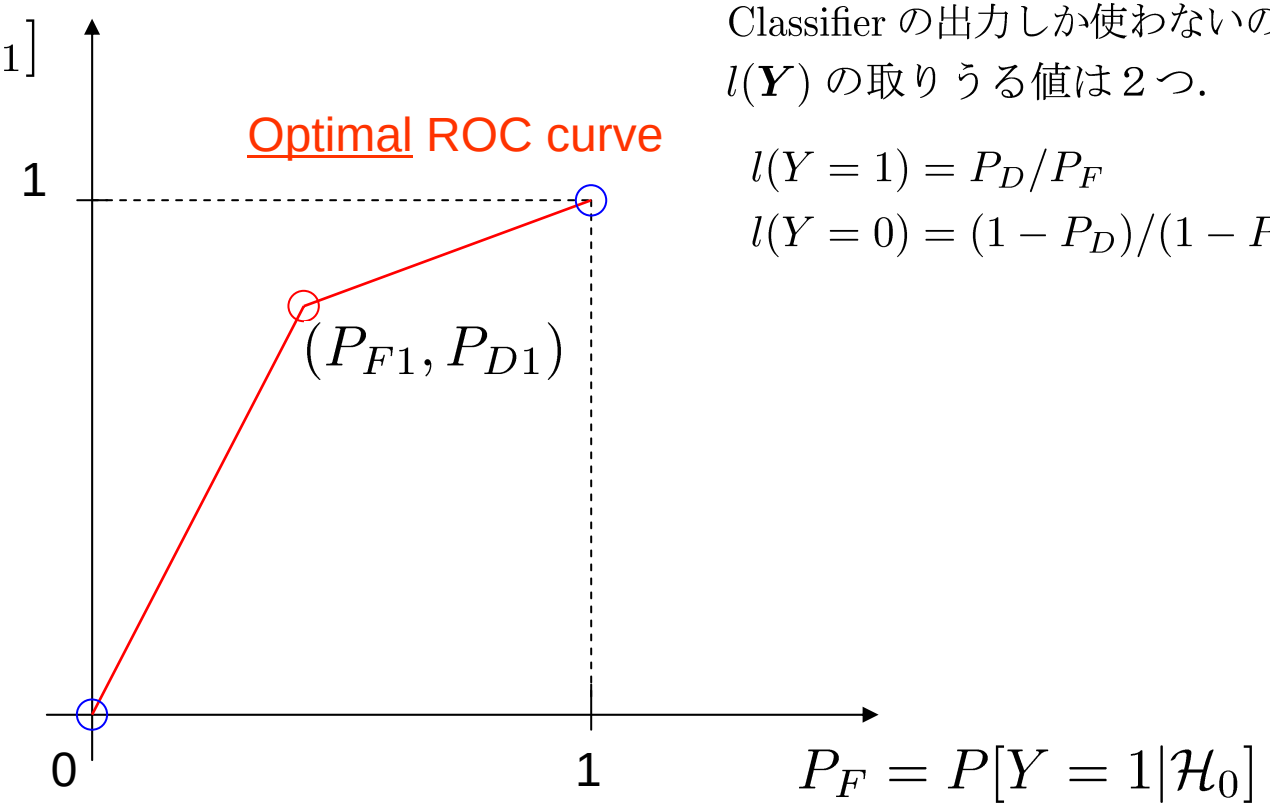
for some $\tau \in (0, \infty)$ and $\gamma \in [0, 1]$, is a most powerful test for its size: no other test has higher $P_D = \Pr[\mathcal{D}(\mathbf{Y}) = 1|H_1]$ for the same bound on $P_F = \Pr[\mathcal{D}(\mathbf{Y}) = 1|H_0]$. (When $\ell(\mathbf{Y}) = \tau$, $\mathcal{D} = 1$ with probability $\gamma$ and 0 otherwise.) Given a test size $\alpha$, we maximize $P_D$ subject to $P_F \leq \alpha$ by choosing $\tau$ and $\gamma$ as follows. First we find the smallest value $\tau^*$ such that $\Pr[\ell(\mathbf{Y}) > \tau^*|H_0] \leq \alpha$. To maximize $P_D$, which is monotonically nondecreasing with $P_F$, we choose the highest value $\gamma^*$ that satisfies $\Pr[\mathcal{D}(\mathbf{Y}) = 1|H_0] = \Pr[\ell(\mathbf{Y}) > \tau^*|H_0] + \gamma^* \Pr[\ell(\mathbf{Y}) = \tau^*|H_0] \leq \alpha$, finding $\gamma^* = (\alpha - \Pr[\ell(\mathbf{Y}) > \tau^*|H_0])/\Pr[\ell(\mathbf{Y}) = \tau^*|H_0]$.

# Classifierがひとつあるとき



# Classifierがひとつあるとき(Lemma1)

$$P_D = P[Y = 1|\mathcal{H}_1]$$




Classifier の出力しか使わないので，
$l(\boldsymbol{Y})$ の取りうる値は２つ．

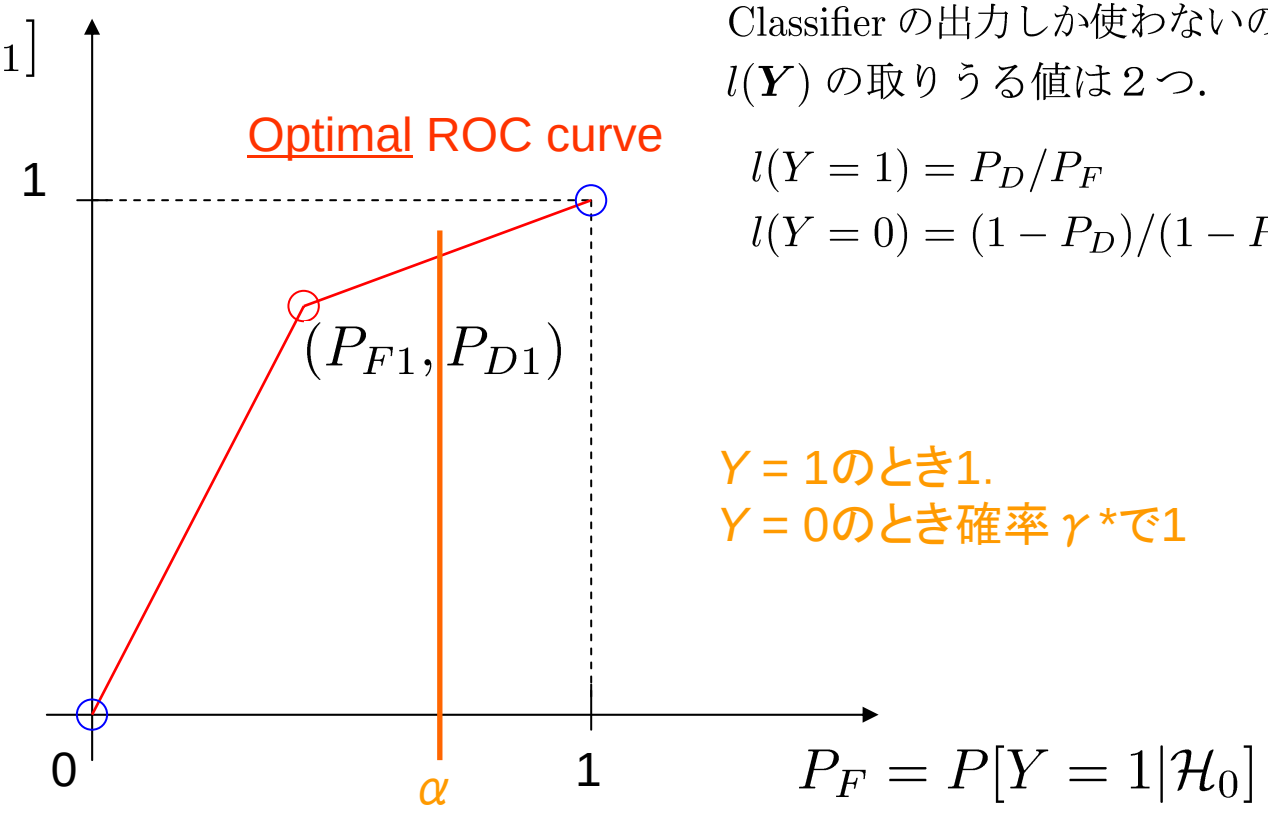
$$l(Y = 1) = P_D/P_F$$
$$l(Y = 0) = (1 - P_D)/(1 - P_F)$$

$$\mathcal{D}(\mathbf{Y}) = \begin{cases} 1 & \text{if } \ell(\mathbf{Y}) > \tau \\ \gamma & \text{if } \ell(\mathbf{Y}) = \tau \\ 0 & \text{if } \ell(\mathbf{Y}) < \tau \end{cases}$$

**Proof.** When $\alpha < P_{F1}$, we can obtain a likelihood ratio test by setting $\tau^* = \ell(1)$ and $\gamma^* = \alpha/P_{F1}$, and for $\alpha > P_{F1}$, we set $\tau^* = \ell(0)$ and $\gamma^* = (\alpha - P_{F1})/(1 - P_{F1})$. □

# Classifierがひとつあるとき(Lemma1)

$$P_D = P[Y = 1|\mathcal{H}_1]$$




Classifier の出力しか使わないので，
$l(\mathbf{Y})$ の取りうる値は２つ．

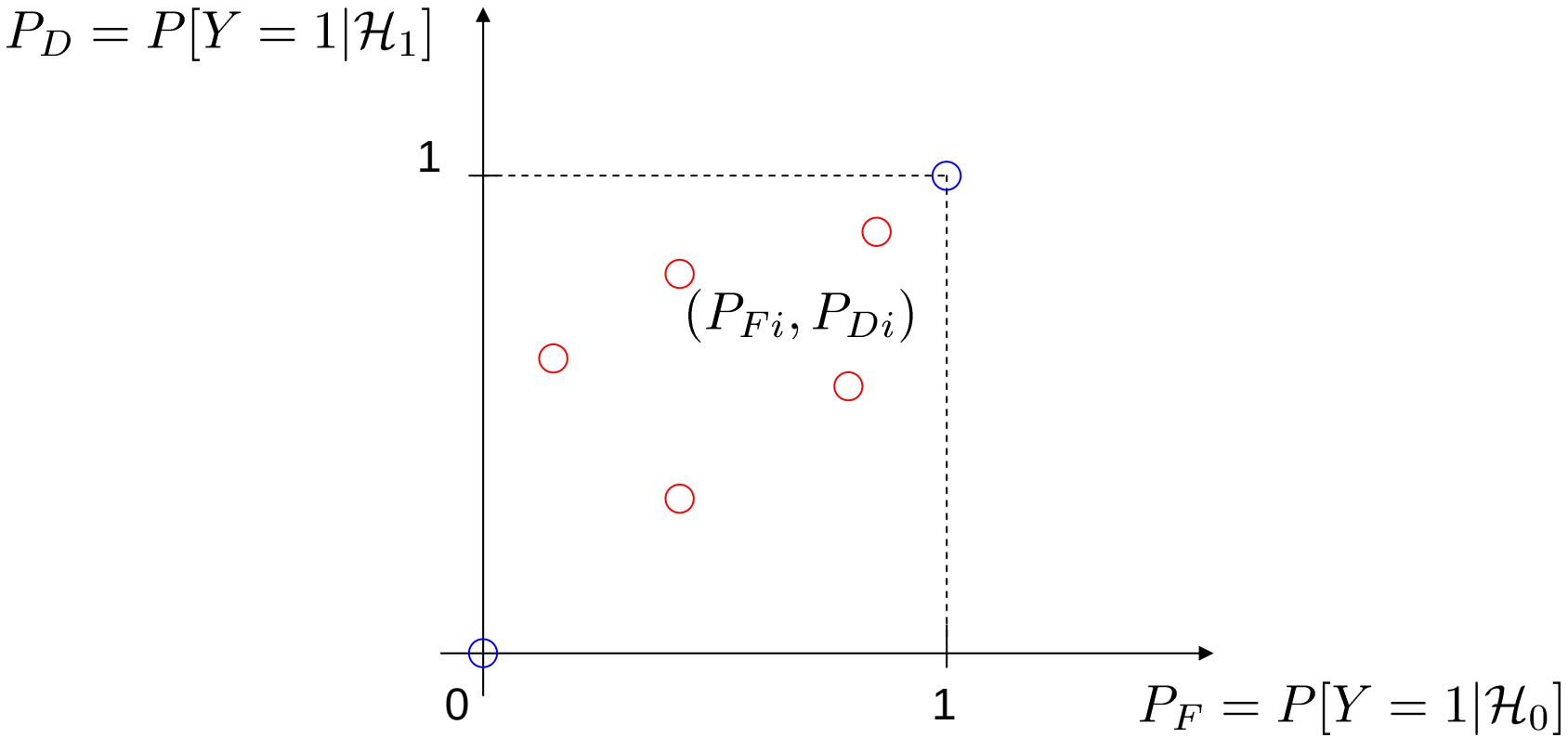
$$l(Y = 1) = P_D/P_F$$
$$l(Y = 0) = (1 - P_D)/(1 - P_F)$$

Y = 1のとき1.
Y = 0のとき確率 γ*で1

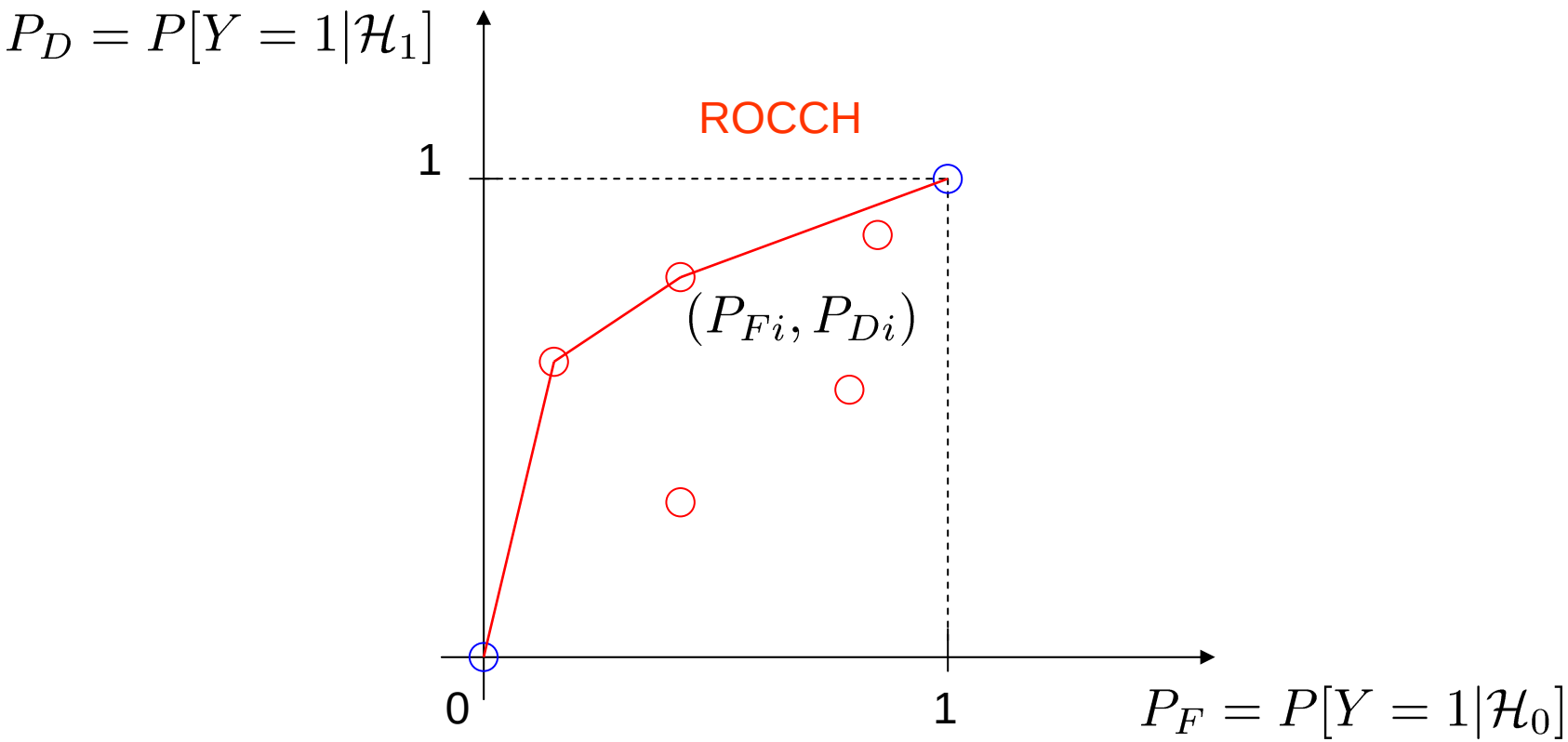
$$\mathcal{D}(\mathbf{Y}) = \begin{cases} 1 & \text{if } \ell(\mathbf{Y}) > \tau \\ \gamma & \text{if } \ell(\mathbf{Y}) = \tau \\ 0 & \text{if } \ell(\mathbf{Y}) < \tau \end{cases}$$

**Proof.** When $\alpha < P_{F1}$, we can obtain a likelihood ratio test by setting $\tau^* = \ell(1)$ and $\gamma^* = \alpha/P_{F1}$, and for $\alpha > P_{F1}$, we set $\tau^* = \ell(0)$ and $\gamma^* = (\alpha - P_{F1})/(1 - P_{F1})$. □

# Classifierが n 個あるとき

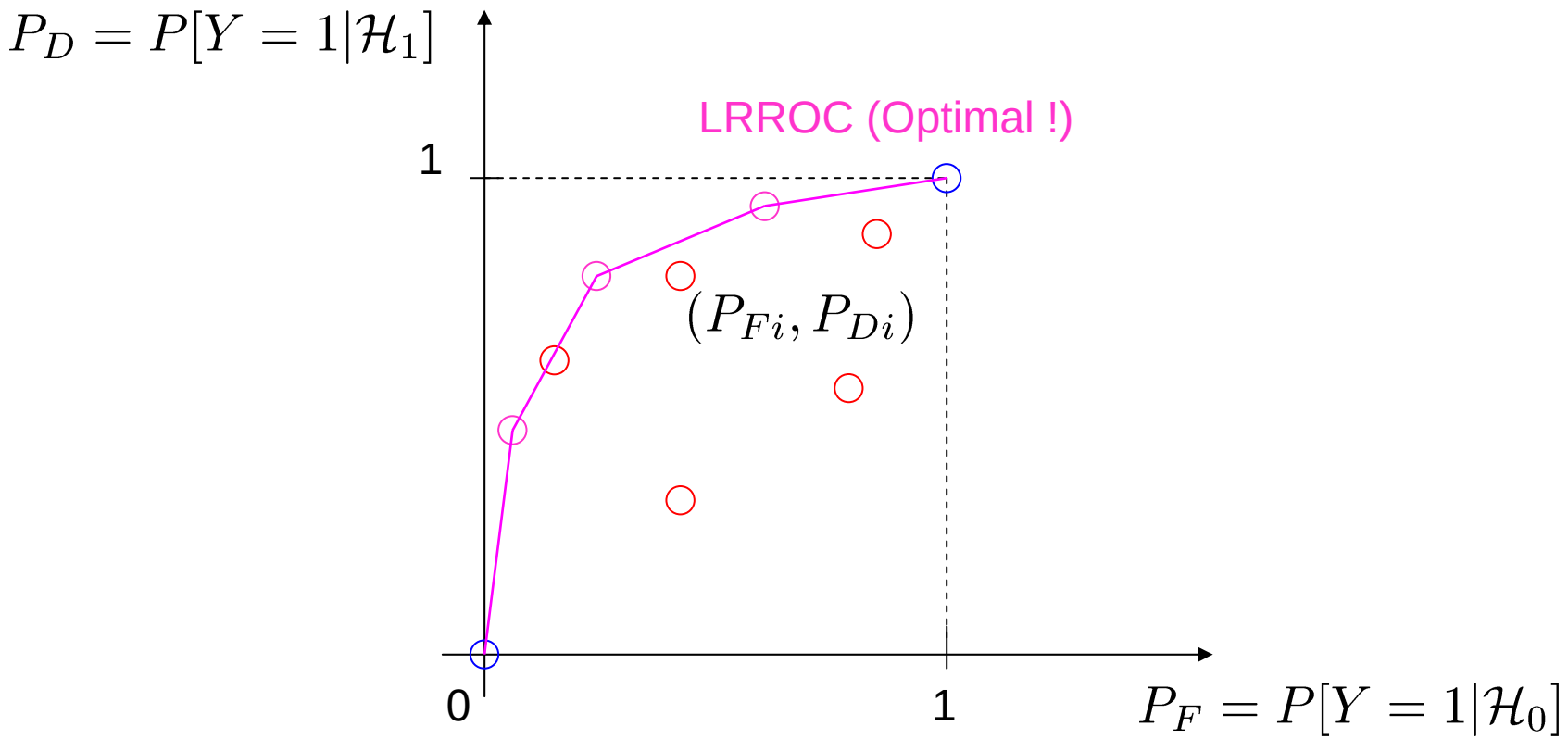


# ROCCH (ROC convex hull) [Provost&Fawcett2001]



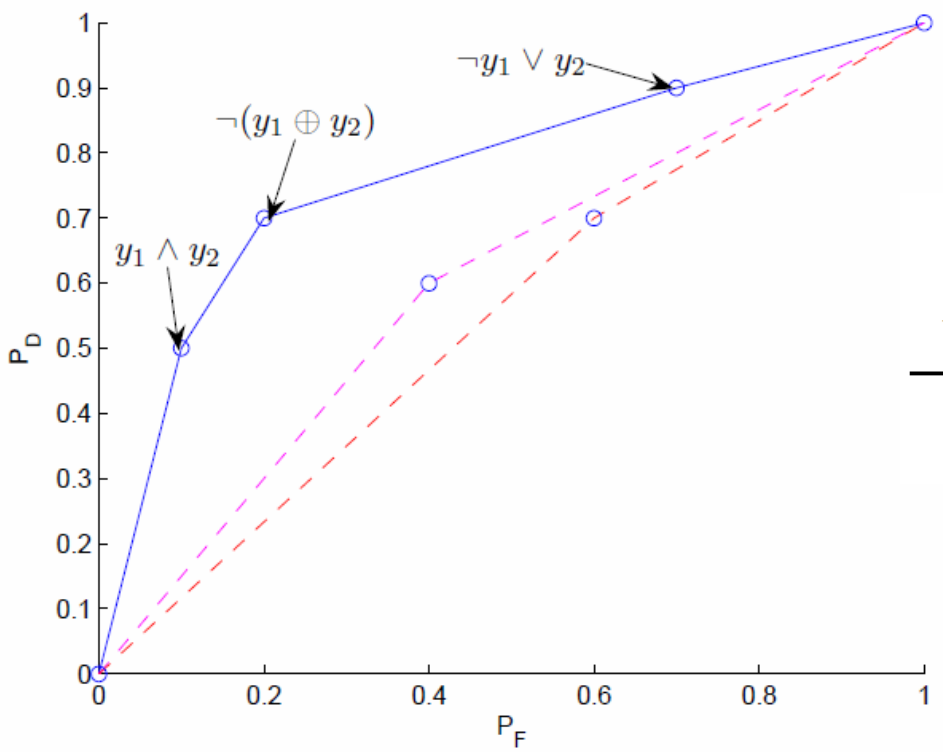

Optimal ではない. Optimalであるための条件は後述.
例えばANDルール, ORルールに負けることが多い. [Fawcett2003]

# 提案手法：LRROC (Likelihood Ratio ROC)




$\boldsymbol{Y} \in \{0,1\}^n$ のすべての実現値（$2^n$ 通り）の尤度比を調べ，尤度比の小さい順に繋いでいく．（尤度比検定になっているので任意の $\alpha(= P_F)$ において最強）

# LRROC構成法（n = 2の場合）




Class 1 ($H_1$)

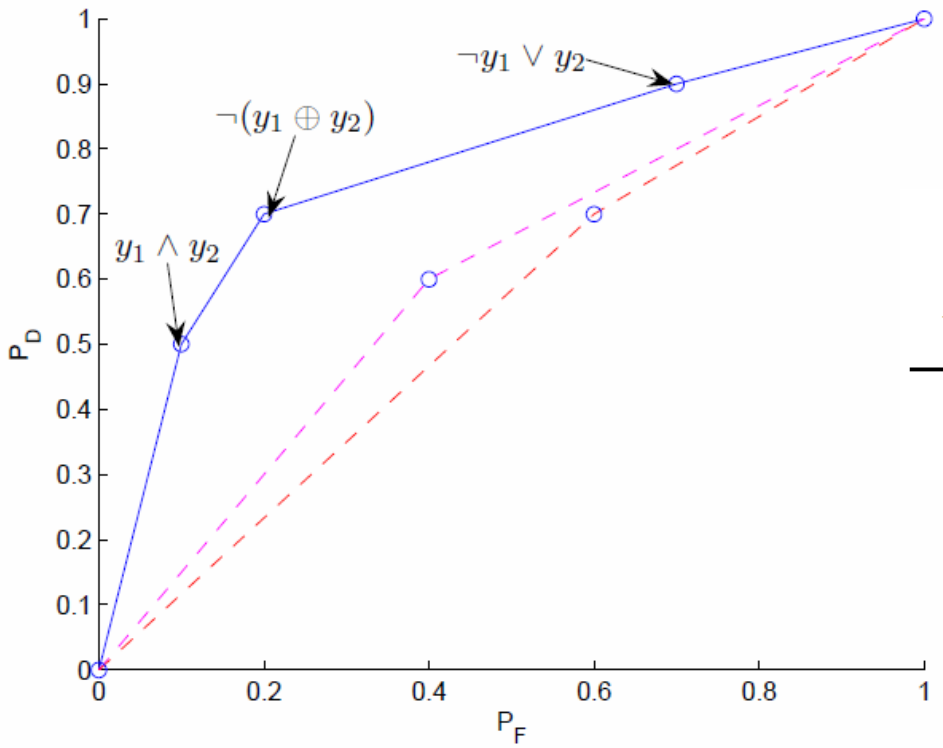
| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.2 | 0.1 |
| 1 | 0.2 | 0.5 |

Class 0 ($H_0$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.1 | 0.3 |
| 1 | 0.5 | 0.1 |

The distribution for the second example appears in Table 1b. The likelihood ratios of the possible outcomes are $\ell(00) = 2.0$, $\ell(10) = 1/3$, $\ell(01) = 0.4$, and $\ell(11) = 5$, so $\ell(10) < \ell(01) < \ell(00) < \ell(11)$ and the three points defining the optimal ROC curve are $\neg Y_1 \vee Y_2$, $\neg(Y_1 \oplus Y_2)$, and $Y_1 \wedge Y_2$ (see Figure 1b). In this case, an XOR rule emerges from the likelihood ratio analysis.

# LRROC構成法（n = 2の場合）




Class 1 ($H_1$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.2 | 0.1 |
| 1 | 0.2 | 0.5 |

Class 0 ($H_0$)

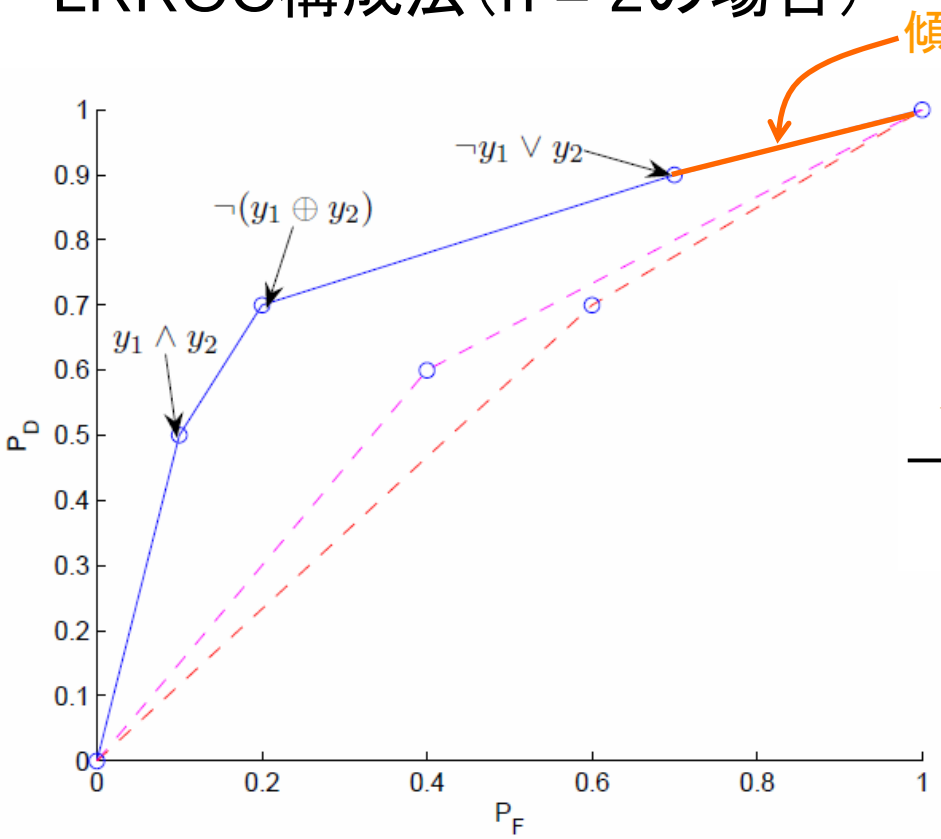
| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.1 | 0.3 |
| 1 | 0.5 | 0.1 |

尤度比

| | |
|---|---|
| 2.0 | 1/3 |
| 0.4 | 5.0 |

The distribution for the second example appears in Table 1b. The likelihood ratios of the possible outcomes are $\ell(00) = 2.0$, $\ell(10) = 1/3$, $\ell(01) = 0.4$, and $\ell(11) = 5$, so $\ell(10) < \ell(01) < \ell(00) < \ell(11)$ and the three points defining the optimal ROC curve are $\neg Y_1 \vee Y_2$, $\neg(Y_1 \oplus Y_2)$, and $Y_1 \wedge Y_2$ (see Figure 1b). In this case, an XOR rule emerges from the likelihood ratio analysis.

# LRROC構成法（n = 2の場合）




Class 1 ($H_1$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.2 | 0.1 |
| 1 | 0.2 | 0.5 |

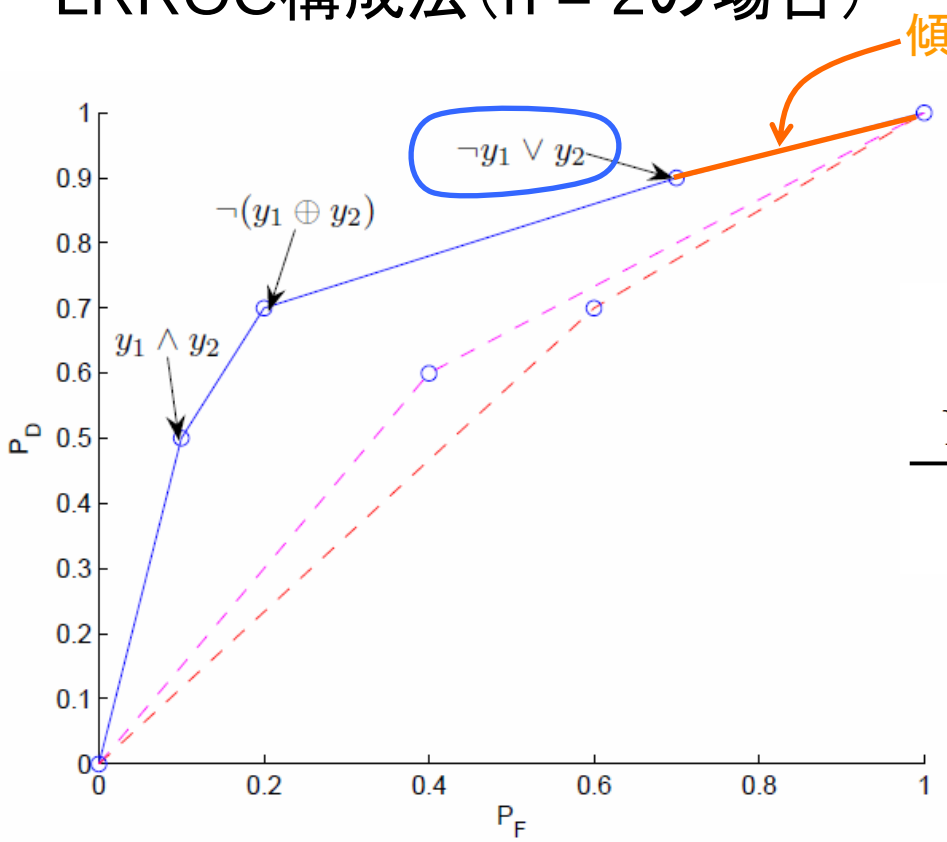
Class 0 ($H_0$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.1 | 0.3 |
| 1 | 0.5 | 0.1 |

| 尤度比 | |
|---|---|
| 2.0 | 1/3 |
| 0.4 | 5.0 |

The distribution for the second example appears in Table 1b. The likelihood ratios of the possible outcomes are $\ell(00) = 2.0$, $\ell(10) = 1/3$, $\ell(01) = 0.4$, and $\ell(11) = 5$, so $\ell(10) < \ell(01) < \ell(00) < \ell(11)$ and the three points defining the optimal ROC curve are $\neg Y_1 \vee Y_2$, $\neg(Y_1 \oplus Y_2)$, and $Y_1 \wedge Y_2$ (see Figure 1b). In this case, an XOR rule emerges from the likelihood ratio analysis.

# LRROC構成法（n = 2の場合）




Class 1 ($H_1$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.2 | 0.1 |
| 1 | 0.2 | 0.5 |

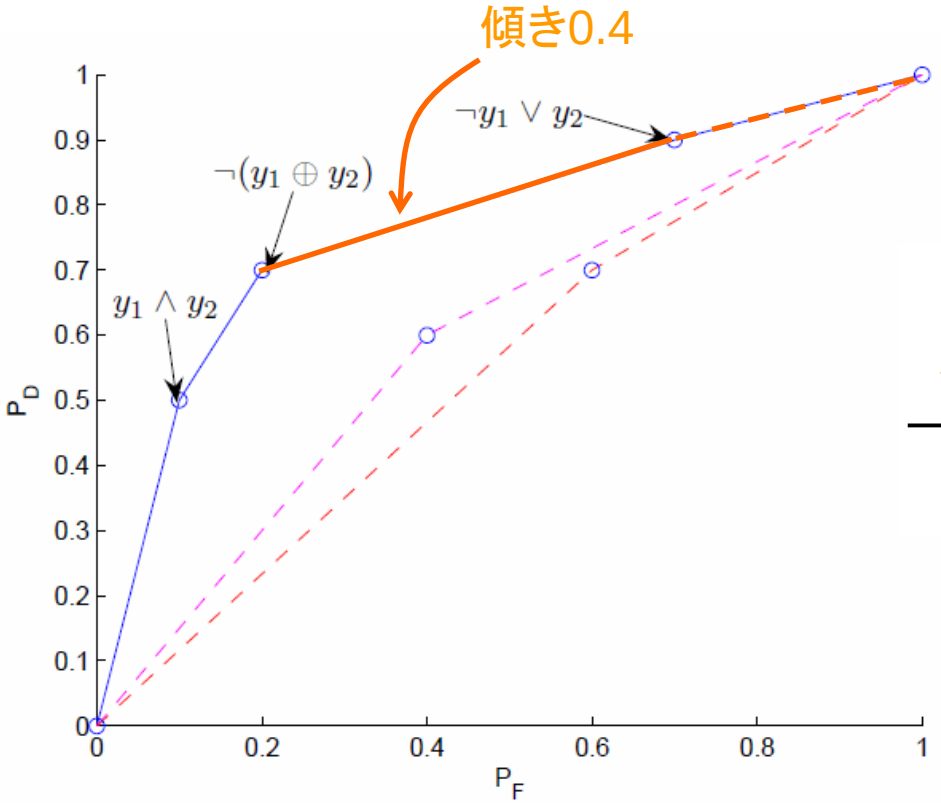
Class 0 ($H_0$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.1 | 0.3 |
| 1 | 0.5 | 0.1 |

| 尤度比 | |
|---|---|
| 2.0 | 1/3 |
| 0.4 | 5.0 |

The distribution for the second example appears in Table 1b. The likelihood ratios of the possible outcomes are $\ell(00) = 2.0$, $\ell(10) = 1/3$, $\ell(01) = 0.4$, and $\ell(11) = 5$, so $\ell(10) < \ell(01) < \ell(00) < \ell(11)$ and the three points defining the optimal ROC curve are $\neg Y_1 \vee Y_2$, $\neg(Y_1 \oplus Y_2)$, and $Y_1 \wedge Y_2$ (see Figure 1b). In this case, an XOR rule emerges from the likelihood ratio analysis.

# LRROC構成法（n = 2の場合）




Class 1 ($H_1$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.2 | 0.1 |
| 1 | 0.2 | 0.5 |

Class 0 ($H_0$)

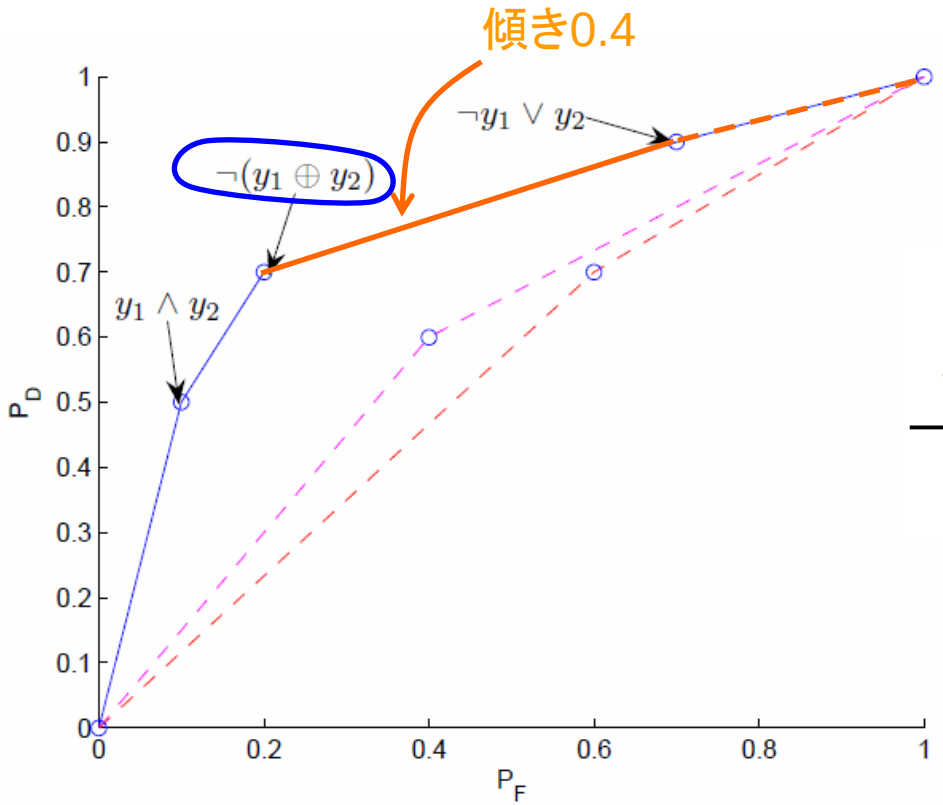
| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.1 | 0.3 |
| 1 | 0.5 | 0.1 |

尤度比

| | |
|---|---|
| 2.0 | 1/3 |
| 0.4 | 5.0 |

The distribution for the second example appears in Table 1b. The likelihood ratios of the possible outcomes are $\ell(00) = 2.0$, $\ell(10) = 1/3$, $\ell(01) = 0.4$, and $\ell(11) = 5$, so $\ell(10) < \ell(01) < \ell(00) < \ell(11)$ and the three points defining the optimal ROC curve are $\neg Y_1 \vee Y_2$, $\neg(Y_1 \oplus Y_2)$, and $Y_1 \wedge Y_2$ (see Figure 1b). In this case, an XOR rule emerges from the likelihood ratio analysis.

# LRROC構成法（n = 2の場合）




Class 1 ($H_1$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.2 | 0.1 |
| 1 | 0.2 | 0.5 |

Class 0 ($H_0$)

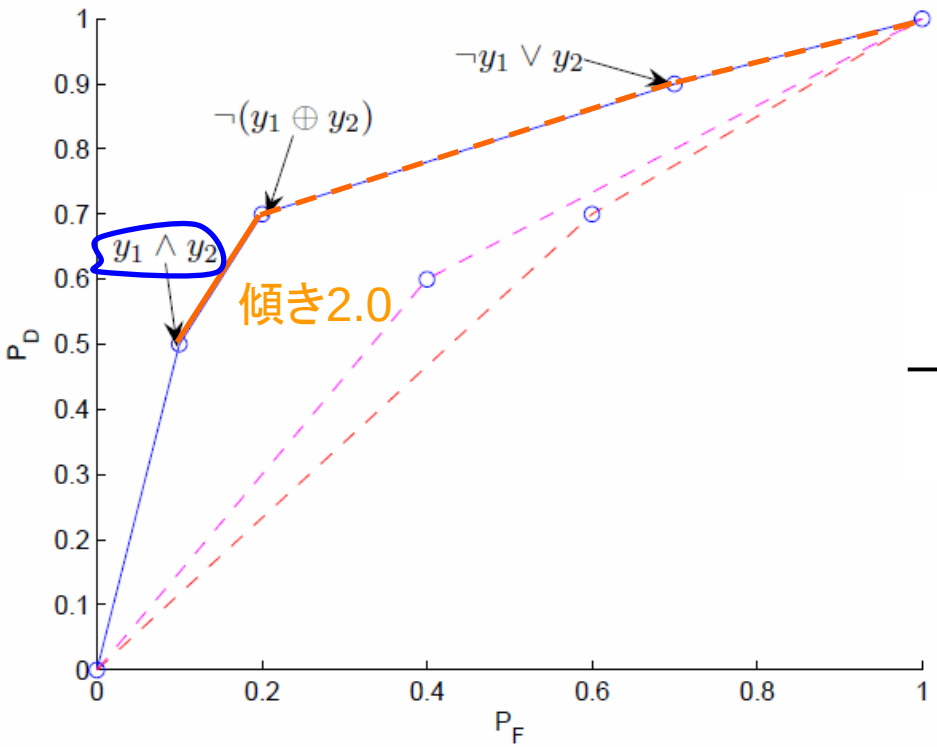
| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.1 | 0.3 |
| 1 | 0.5 | 0.1 |

尤度比

| | |
|---|---|
| 2.0 | 1/3 |
| 0.4 | 5.0 |

The distribution for the second example appears in Table 1b. The likelihood ratios of the possible outcomes are $\ell(00) = 2.0$, $\ell(10) = 1/3$, $\ell(01) = 0.4$, and $\ell(11) = 5$, so $\ell(10) < \ell(01) < \ell(00) < \ell(11)$ and the three points defining the optimal ROC curve are $\neg Y_1 \vee Y_2$, $\neg(Y_1 \oplus Y_2)$, and $Y_1 \wedge Y_2$ (see Figure 1b). In this case, an XOR rule emerges from the likelihood ratio analysis.

# LRROC構成法（n = 2の場合）




Class 1 ($H_1$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.2 | 0.1 |
| 1 | 0.2 | 0.5 |

Class 0 ($H_0$)

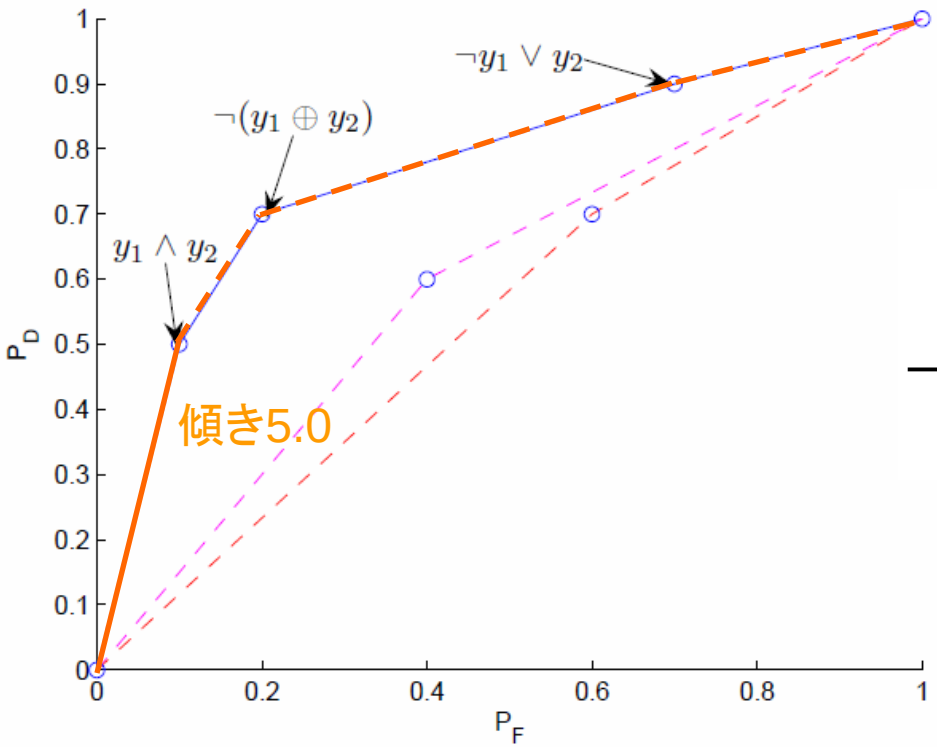
| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.1 | 0.3 |
| 1 | 0.5 | 0.1 |

尤度比

| | |
|---|---|
| 2.0 | 1/3 |
| 0.4 | 5.0 |

The distribution for the second example appears in Table 1b. The likelihood ratios of the possible outcomes are $\ell(00) = 2.0$, $\ell(10) = 1/3$, $\ell(01) = 0.4$, and $\ell(11) = 5$, so $\ell(10) < \ell(01) < \ell(00) < \ell(11)$ and the three points defining the optimal ROC curve are $\neg Y_1 \vee Y_2$, $\neg(Y_1 \oplus Y_2)$, and $Y_1 \wedge Y_2$ (see Figure 1b). In this case, an XOR rule emerges from the likelihood ratio analysis.

# LRROC構成法（n = 2の場合）




Class 1 ($H_1$)

| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.2 | 0.1 |
| 1 | 0.2 | 0.5 |

Class 0 ($H_0$)

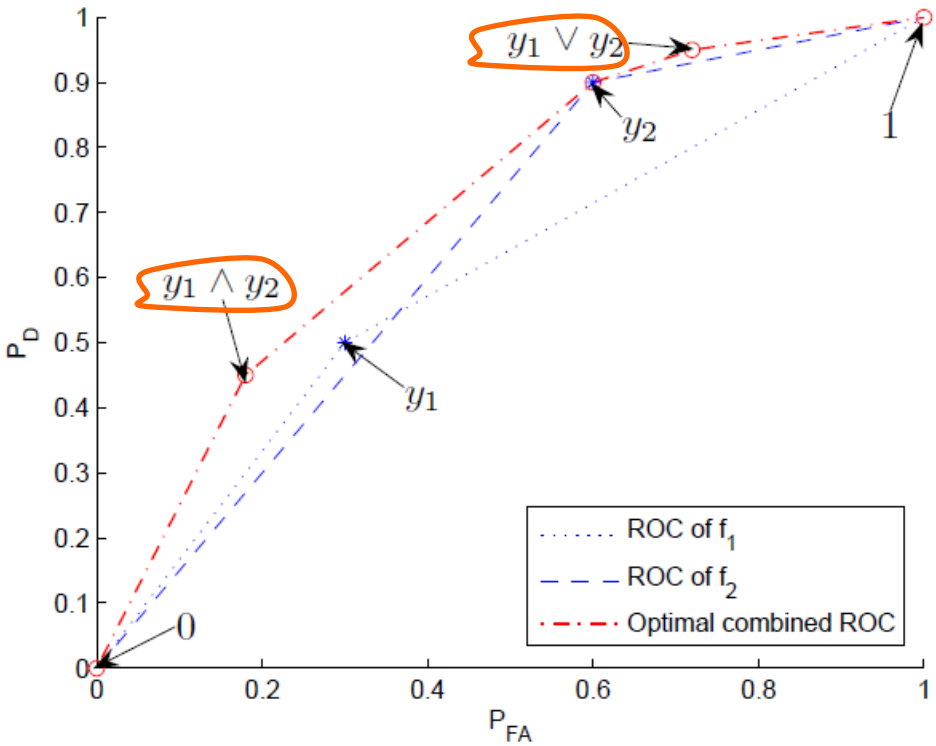
| $Y_2$ \ $Y_1$ | 0 | 1 |
|---|---|---|
| 0 | 0.1 | 0.3 |
| 1 | 0.5 | 0.1 |

尤度比

| | |
|---|---|
| 2.0 | 1/3 |
| 0.4 | 5.0 |

The distribution for the second example appears in Table 1b. The likelihood ratios of the possible outcomes are $\ell(00) = 2.0$, $\ell(10) = 1/3$, $\ell(01) = 0.4$, and $\ell(11) = 5$, so $\ell(10) < \ell(01) < \ell(00) < \ell(11)$ and the three points defining the optimal ROC curve are $\neg Y_1 \vee Y_2$, $\neg(Y_1 \oplus Y_2)$, and $Y_1 \wedge Y_2$ (see Figure 1b). In this case, an XOR rule emerges from the likelihood ratio analysis.

# 独立な場合




ANDとORが必ず含まれる.

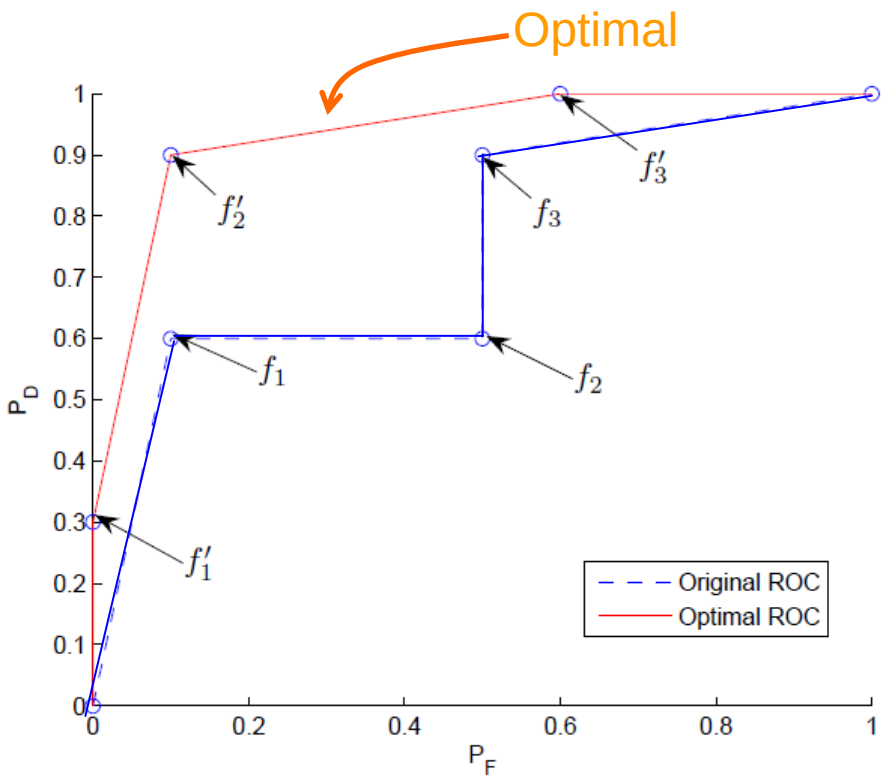
**Theorem 2** *If the distributions of the outputs of $n$ proper binary classifiers $Y_1, Y_2, \ldots, Y_n$ are conditionally independent given the instance class, then the points in ROC space for the rules AND $(Y_1 \wedge Y_2 \wedge \cdots \wedge Y_n)$ and OR $(Y_1 \vee Y_2 \vee \cdots \vee Y_n)$ are strictly above the convex hull of the ROC curves of the base classifiers $f_1, \ldots, f_n$. Furthermore, these Boolean rules belong to the LR-ROC.*

# ROCCHがOPTIMALであるための条件

**Theorem 3** *Consider $n$ classifiers $f_1, \ldots, f_n$. The convex hull of the points $(P_{Fi}, P_{Di})$ and $(0,0)$ and $(1,1)$ (the ROCCH) is an optimal ROC curve for the combination if $(Y_i = 1) \Rightarrow (Y_j = 1)$ for $i < j$ and the following ordering is satisfied: $\ell(00\cdots0) < \ell(00\cdots01) < \ell(00\cdots011) < \cdots < \ell(1\cdots1)$.*

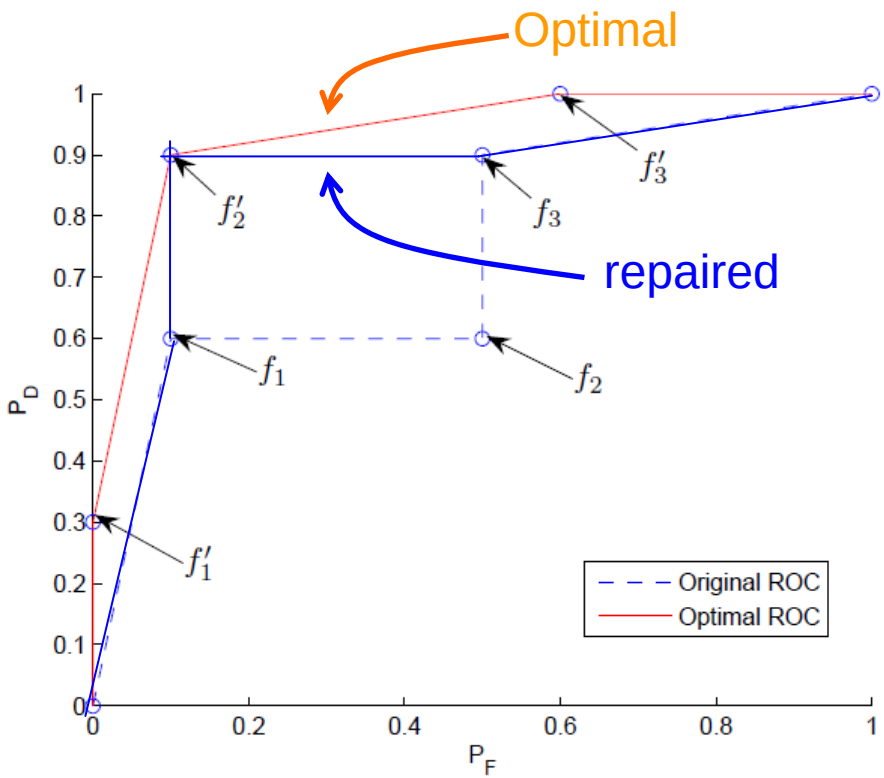
Classifierに順序がつけられる場合.
（尤度の取りうる値が n + 1 個）

# 従来法（repairing [Flach&Wu2002]との比較）




Here is an example comparing their method to ours. Consider the following probability distribution on a random variable $\mathbf{Y} \in \{0,1\}^2$: $P((00,10,01,11)|H_1) = (0.1,0.3,0.0,0.6)$, $P((00,10,01,11)|H_0) = (0.5,0.001,0.4,0.099)$.

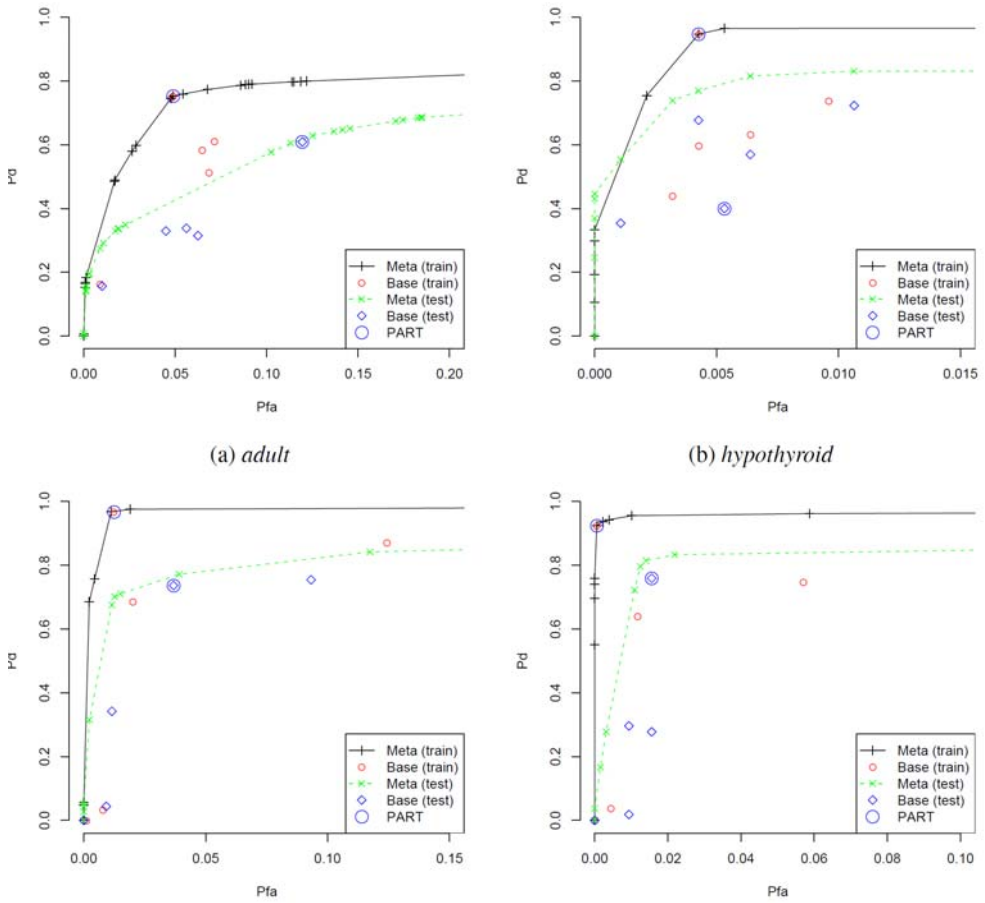
# 従来法（repairing [Flach&Wu2002]との比較）




Here is an example comparing their method to ours. Consider the following probability distribution on a random variable $\mathbf{Y} \in \{0,1\}^2$: $P((00,10,01,11)|H_1) = (0.1,0.3,0.0,0.6)$, $P((00,10,01,11)|H_0) = (0.5,0.001,0.4,0.099)$.

# 実験結果（UCI）

各Classifierを学習（ハイパーパラメータなどはデフォルト設定）してから学習データによって条件付尤度を推定.




(a) *adult* (b) *hypothyroid*

We chose five base classifiers from the YALE machine learning platform [7]: PART (a decision list algorithm), SMO (John Platt's Sequential Minimal Optimization), SimpleLogistic, VotedPerceptron, and Y-NaiveBayes. We used the default settings for all classifiers. The *adult* dataset has around

the data to hold out for testing.) We compute the likelihood ratios for all outcomes and order them. When there are outcomes with no positive or no negative training examples, we treat $\cdot/0$ as near-infinite and $0/\cdot$ as near-zero. For an outcome with no training examples at all, we define $0/0 = 1$.

# Future work

- 2^n通りの計算はしんどいので，近似でいいからもっと減らせないか？
- 条件つき尤度推定のロバスト化


NIKON CORPORATION