

Non-Bayesian Methods

lecturer: Ondřej Drbohlav, drbohlav@fel.cvut.cz

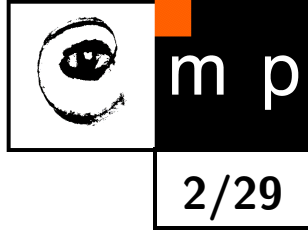
authors: Ondřej Drbohlav, Jiří Matas, Václav Hlaváč

Czech Technical University, Faculty of Electrical Engineering
Department of Cybernetics, Center for Machine Perception
121 35 Praha 2, Karlovo nám. 13, Czech Republic

<http://cmp.felk.cvut.cz>

September 2022, Ver. 1.1

Lecture Outline



1. Limitations of Bayesian Decision Theory
2. Neyman Pearson Task
3. Minimax Task
4. Wald Task

Bayesian Decision Theory



Recall:

X set of observations

K set of hidden states

D set of decisions

p_{XK} : $X \times K \rightarrow \mathbb{R}$: joint probability

W : $K \times D \rightarrow \mathbb{R}$: *loss function*,

q : $X \rightarrow D$: strategy

$R(q)$: risk:

$$R(q) = \sum_{x \in X} \sum_{k \in K} p_{XK}(x, k) W(k, q(x)) \quad (1)$$

Bayesian strategy q^* :

$$q^* = \operatorname{argmin}_{q \in X \rightarrow D} R(q) \quad (2)$$

Limitations of the Bayesian Decision Theory

The limitations follow from the very ingredients of the Bayesian Decision Theory — the necessity to know all the probabilities and the loss function.

- ◆ The loss function W must make sense, but in many tasks it wouldn't
 - medical diagnosis task (W : price of medicines, staff labor, etc. but what penalty in case of patient's death?) Uncomparable penalties on different axes of X .
 - nuclear plant
 - judicial error
- ◆ The prior probabilities $p_K(k)$: must exist and be known. But in some cases it does not make sense to talk about probabilities because the events are not random.
 - $K = \{1, 2\} \equiv \{\text{own army plane, enemy plane}\}$;
 $p(x|1)$, $p(x|2)$ do exist and can be estimated, but $p(1)$ and $p(2)$ don't.
- ◆ The conditionals may be subject to non-random intervention; $p(x | k, z)$ where $z \in Z = \{1, 2, 3\}$ are different interventions.
 - a system for handwriting recognition: The training set has been prepared by 3 different persons. But the test set has been constructed by one of the 3 persons only. This **cannot** be done:

$$(!) \quad p(x | k) = \sum_z p(z)p(x | k, z) \quad (3)$$

Neyman Pearson Task

- ◆ $K = \{D, N\}$ (dangerous state, normal state)
- ◆ X set of observations
- ◆ Conditionals $p(x | D)$, $p(x | N)$ are given
- ◆ The priors $p(D)$ and $p(N)$ are unknown or do not exist
- ◆ $q: X \rightarrow K$ strategy

The Neyman Person Task looks for the optimal strategy q^* for which

- i) the error of classification of the dangerous state is lower than a predefined threshold $\bar{\epsilon}_D$ ($0 < \bar{\epsilon}_D < 1$), while
- ii) the classification error for the normal state is as low as possible.

This is formulated as an optimization task with an inequality constraint:

$$q^* = \operatorname{argmin}_{q: X \rightarrow K} \sum_{x: q(x) \neq N} p(x | N) \quad (4)$$

$$\text{subject to: } \sum_{x: q(x) \neq D} p(x | D) \leq \bar{\epsilon}_D. \quad (5)$$

Neyman Pearson Task

(copied from the previous slide:)

$$q^* = \operatorname{argmin}_{q: X \rightarrow K} \sum_{x: q(x) \neq N} p(x | N) \quad (4)$$

$$\text{subject to: } \sum_{x: q(x) \neq D} p(x | D) \leq \bar{\epsilon}_D. \quad (5)$$

A strategy is characterized by the classification error values ϵ_N and ϵ_D :

$$\epsilon_N = \sum_{x: q(x) \neq N} p(x | N) \quad (\text{false alarm}) \quad (6)$$

$$\epsilon_D = \sum_{x: q(x) \neq D} p(x | D) \quad (\text{overlooked danger}) \quad (7)$$

Example: Male/Female Recognition (Neyman Pearson) (1)

An aging student at CTU wants to marry. He can't afford to miss recognizing a girl when he meets her, therefore he sets the threshold on female classification error to $\bar{\epsilon}_D = 0.2$. At the same time, he wants to minimize mis-classifying boys for girls.

- ◆ $K = \{D, N\} \equiv \{F, M\}$ (female, male)
- ◆ measurements $X = \{\text{short, normal, tall}\} \times \{\text{ultralight, light, avg, heavy}\}$
- ◆ Prior probabilities do not exist.
- ◆ Conditionals are given as follows:

$p(x|F)$

short	.197	.145	.094	.017
normal	.077	.299	.145	.017
tall	.001	.008	.000	.000
	u-light	light	avg	heavy

$p(x|M)$

short	.011	.005	.011	.011
normal	.005	.071	.408	.038
tall	.002	.014	.255	.169
	u-light	light	avg	heavy

(8)

Neyman Pearson : Solution

The optimal strategy q^* for a given $x \in X$ is constructed using the likelihood ratio $\frac{p(x | \text{N})}{p(x | \text{D})}$.

Let there be a constant $\mu \geq 0$. Given this μ , a strategy q is constructed as follows:

$$\frac{p(x | \text{N})}{p(x | \text{D})} > \mu \quad \Rightarrow \quad q(x) = \text{N}, \quad (9)$$

$$\frac{p(x | \text{N})}{p(x | \text{D})} \leq \mu \quad \Rightarrow \quad q(x) = \text{D}. \quad (10)$$

The optimal strategy q^* is obtained by selecting the minimal μ for which there still holds that $\epsilon_{\text{D}} \leq \bar{\epsilon}_{\text{D}}$.

Let us show this on an example.

Example: Male/Female Recognition (Neyman Pearson) (2)

$p(x|F)$

short	.197	.145	.094	.017
normal	.077	.299	.145	.017
tall	.001	.008	.000	.000
	u-light	light	avg	heavy

$p(x|M)$

short	.011	.005	.011	.011
normal	.005	.071	.408	.038
tall	.002	.014	.255	.169
	u-light	light	avg	heavy

$r(x) = p(x|M)/p(x|F)$

short	0.056	0.034	0.117	0.647
normal	0.065	0.237	2.814	2.235
tall	2.000	1.750	∞	∞
	u-light	light	avg	heavy

rank order of $p(x|M)/p(x|F)$

short	2	1	4	6
normal	3	5	10	9
tall	8	7	11	12
	u-light	light	avg	heavy

Here, different μ 's can produce 11 different strategies.

First, let us take $2.814 < \mu < \infty$, e.g. $\mu = 3$. This produces a strategy $q^*(x) = F$ everywhere except where $p(x|F) = 0$. Obviously, classification error ϵ_F for F is $\epsilon_F = 0$, and $\epsilon_M = 1 - .255 - .169 = .576$.

Example: Male/Female Recognition (Neyman Pearson) (3)

 $p(x|F)$

short	.197	.145	.094	.017
normal	.077	.299	.145	.017
tall	.001	.008	.000	.000
	u-light	light	avg	heavy

 $p(x|M)$

short	.011	.005	.011	.011
normal	.005	.071	.408	.038
tall	.002	.014	.255	.169
	u-light	light	avg	heavy

 $r(x) = p(x|M)/p(x|F)$

short	0.056	0.034	0.117	0.647
normal	0.065	0.237	2.814	2.235
tall	2.000	1.750	∞	∞
	u-light	light	avg	heavy

 rank, and $q^*(x) = \{F, M\}$ for $\mu = 2.5$

short	2	1	4	6
normal	3	5	10	9
tall	8	7	11	12
	u-light	light	avg	heavy

Next, take μ which satisfies

$$r_9 < \mu < r_{10} \quad (\text{e.g. } \mu = 2.5) \tag{11}$$

(where r_i is the likelihood ratios indexed by its rank.)

Here, $\epsilon_F = .145$, and $\epsilon_M = 1 - .255 - .169 - .408 = .168$.

Example: Male/Female Recognition (Neyman Pearson) (4)

 $p(x|F)$

short	.197	.145	.094	.017
normal	.077	.299	.145	.017
tall	.001	.008	.000	.000
	u-light	light	avg	heavy

 $p(x|M)$

short	.011	.005	.011	.011
normal	.005	.071	.408	.038
tall	.002	.014	.255	.169
	u-light	light	avg	heavy

 $r(x) = p(x|M)/p(x|F)$

short	0.056	0.034	0.117	0.647
normal	0.065	0.237	2.814	2.235
tall	2.000	1.750	∞	∞
	u-light	light	avg	heavy

 rank, and $q^*(x) = \{F, M\}$ for $\mu = 2.1$

short	2	1	4	6
normal	3	5	10	9
tall	8	7	11	12
	u-light	light	avg	heavy

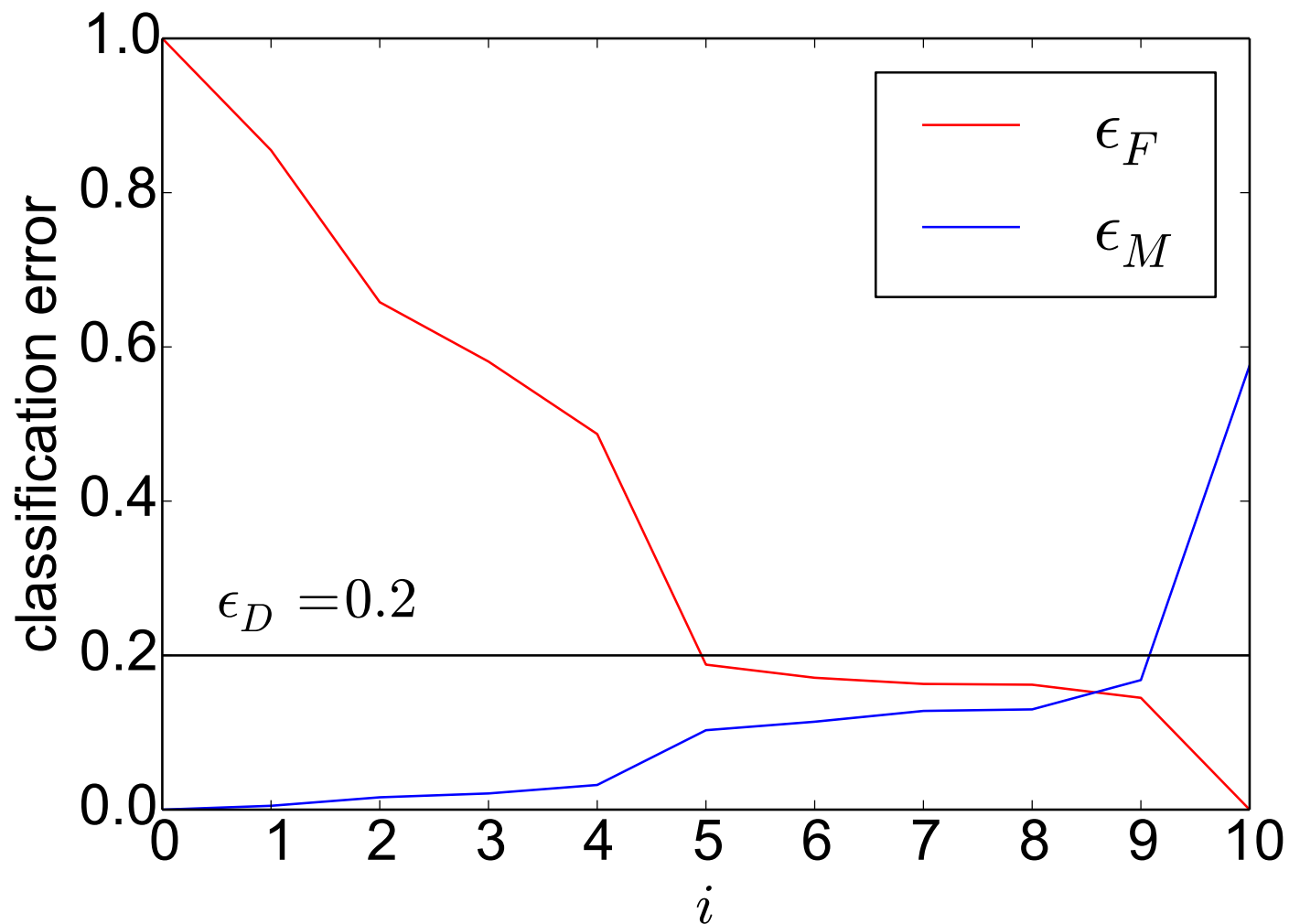
Do the same for μ satisfying

$$r_8 < \mu < r_9 \quad (\text{e.g. } \mu = 2.1) \quad (12)$$

$\Rightarrow \epsilon_F = .162$, and $\epsilon_M = 0.13$.

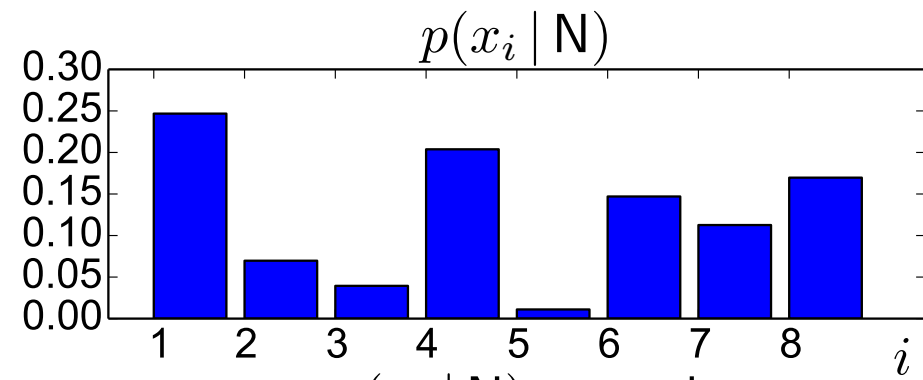
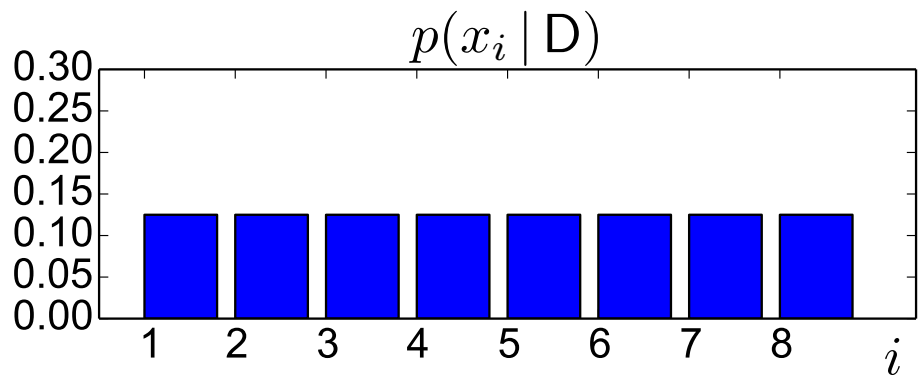
Example: Male/Female Recognition (Neyman Pearson) (5)

Classification errors for F and M, for $\mu_i = \frac{r_i+r_{i+1}}{2}$ and $\mu_0 = 0$.

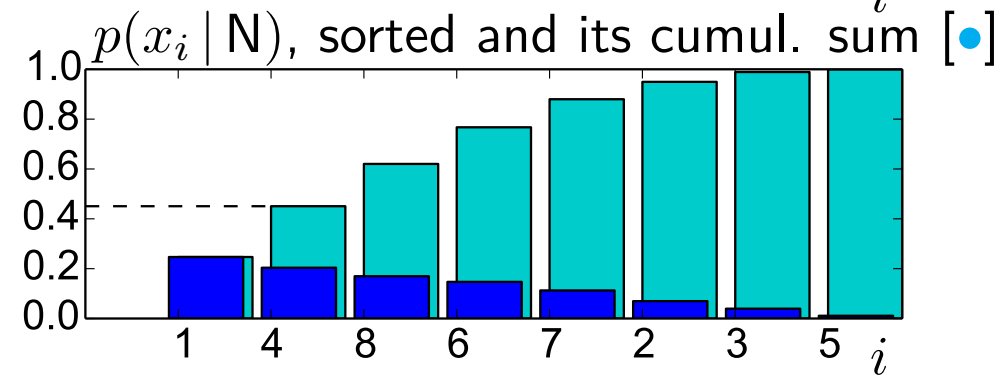
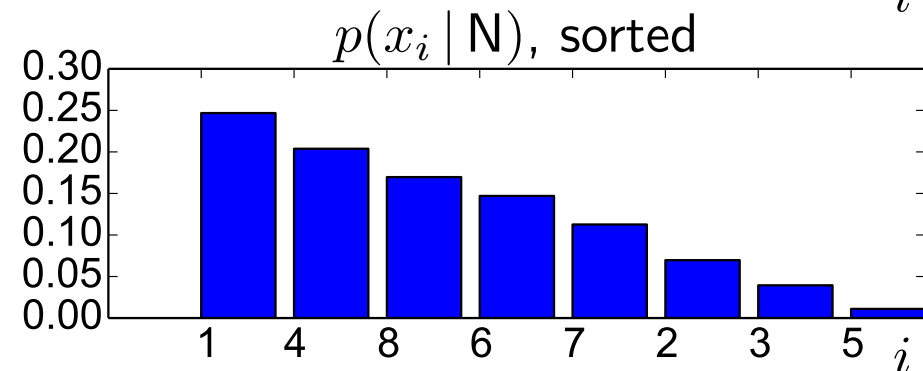


The optimum is reached for $r_5 < \mu < r_6$; $\epsilon_F = .188$, $\epsilon_M = .103$

Neyman Pearson : Solution (1, special case)



Consider first a special case when $p(x_i | D) = \text{const} = \frac{1}{8}$. Possible values for ϵ_D are $0, \frac{1}{8}, \frac{2}{8}, \dots, 1$. If a strategy q classifies P observations as normal then $\epsilon_D = \frac{P}{8}$. Let $P = 1$ and thus $\epsilon_D = \frac{1}{8}$. It is clear that ϵ_N will attain minimum if the (one) observation which is classified as normal is the one with the highest $p(x_i | N)$. Similarly, if $P = 2$ then the two observations to be classified as normal are the one with the first two highest $p(x_i | N)$. Etc.



↑ cumulative sum of sorted $p(x_i | N)$ shows the classification success rate for N, that is, $1 - \epsilon_N$, for $\epsilon_D = \frac{1}{8}, \frac{2}{8}, \dots, 1$. For example, for $\epsilon_D = \frac{2}{8}$ ($P = 2$), $\epsilon_N = 1 - 0.45 = 0.55$ (as shown, dashed.)

Neyman Pearson : Solution (2, general case)

In general, $p(x_i | D) \neq \text{const.}$ Consider the following example:

x_1	x_2	x_3
0.5	0.25	0.25

x_1	x_2	x_3
0.6	0.35	0.05

But this can easily be converted to the previous special case by (only formally) splitting x_1 to two observations x'_1 and x''_1 :

x'_1	x''_1	x_2	x_3
0.25	0.25	0.25	0.25

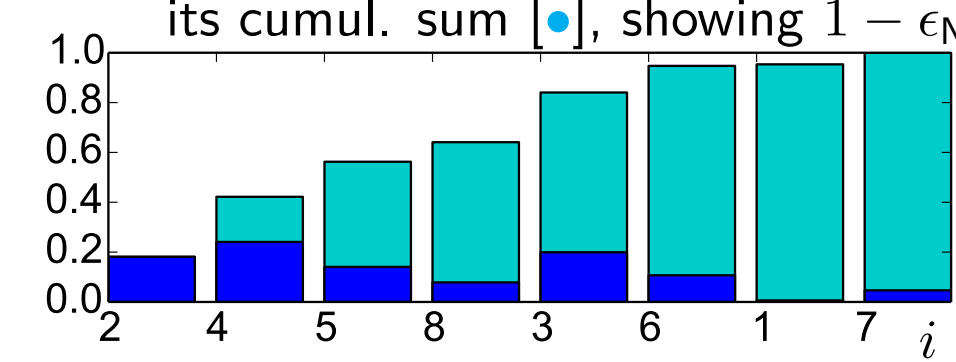
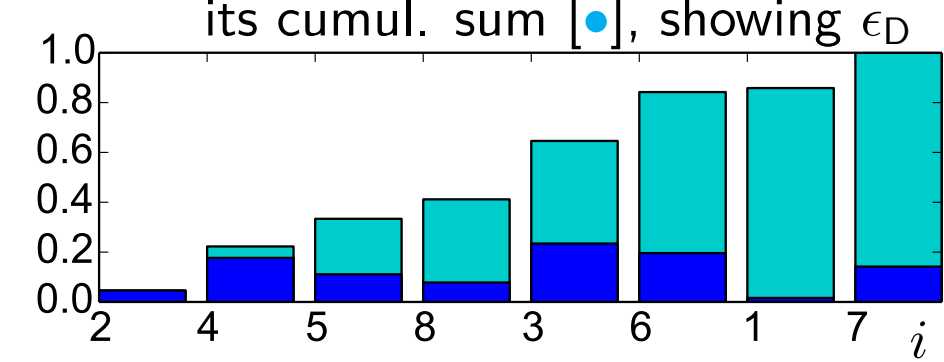
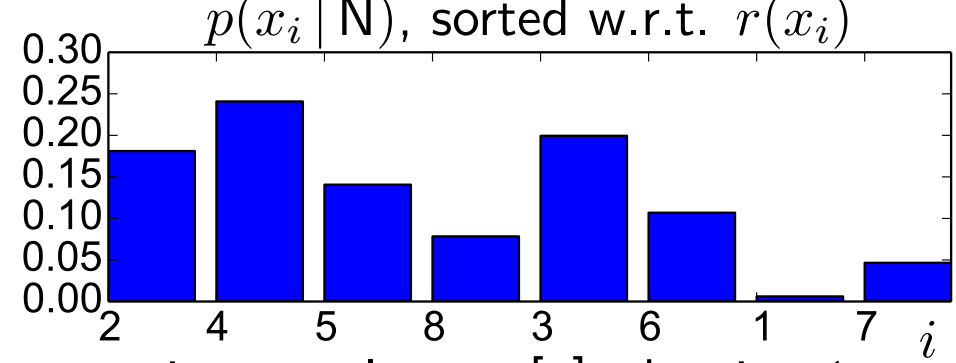
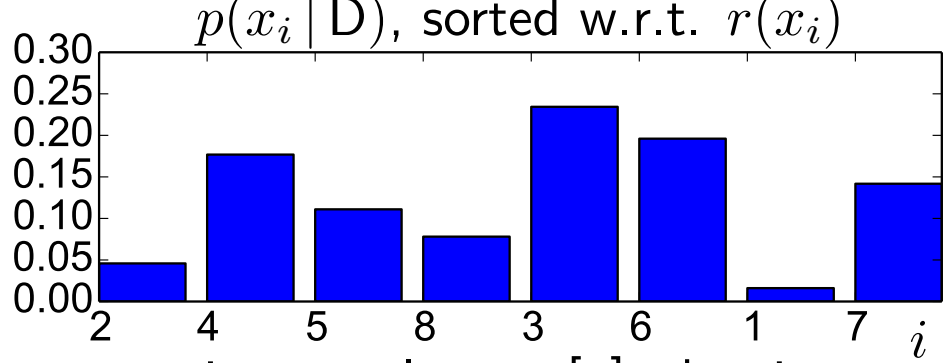
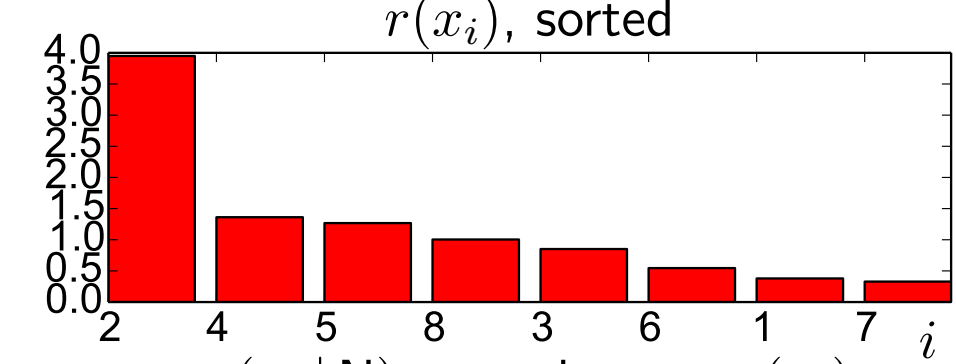
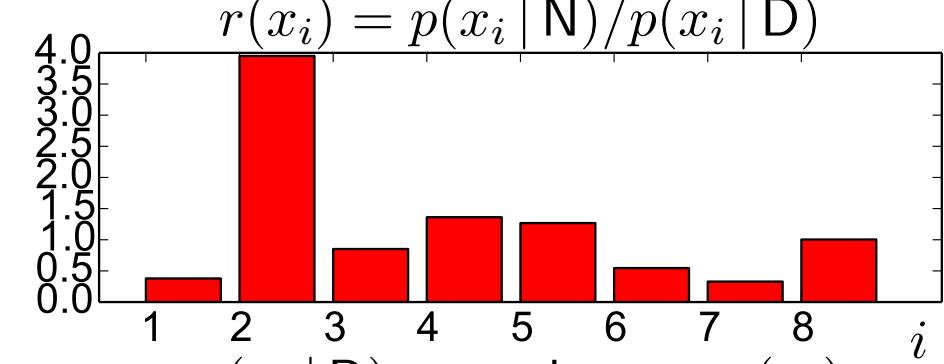
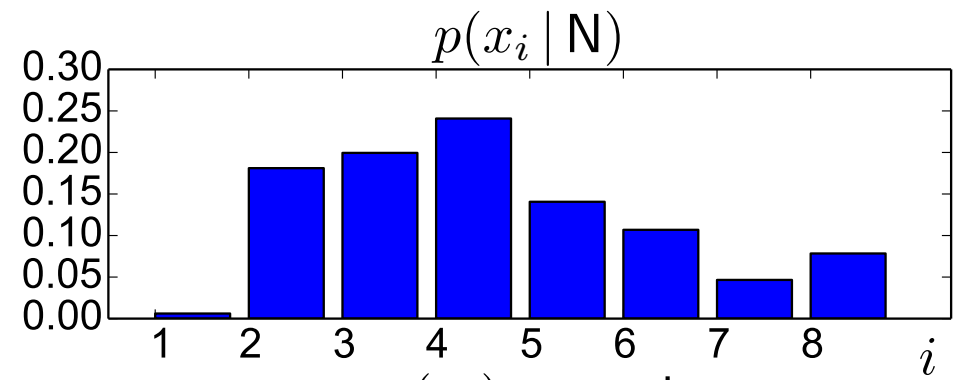
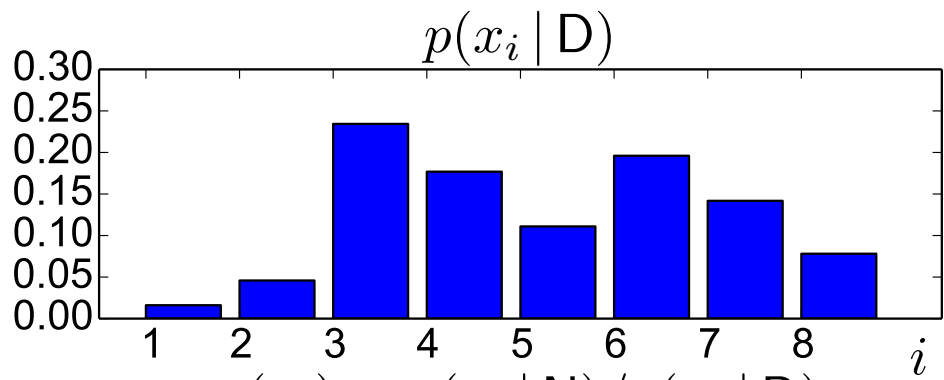
x'_1	x''_1	x_2	x_3
0.3	0.3	0.35	0.05

which would result in ordering the observations by decreasing $p(x_i | N)$ as: x_2, x_1, x_3 .

Obviously, the same ordering is obtained when $p(x_i | N)$ is 'normalized' by $p(x_i | D)$, that is, using the likelihood ratio

$$r(x_i) = \frac{p(x_i | N)}{p(x_i | D)}. \tag{13}$$

Neyman Pearson : Solution (3, general case, example)



Neyman Pearson Solution : Illustration of Principle

Lagrangian of the Neyman Pearson Task is

$$L(q) = \underbrace{\sum_{x: q(x)=D} p(x | N)}_{=} + \mu \left(\sum_{x: q(x)=N} p(x | D) - \bar{\epsilon}_D \right) \quad (14)$$

$$= 1 - \sum_{x: q(x)=N} p(x | N) + \mu \left(\sum_{x: q(x)=N} p(x | D) \right) - \mu \bar{\epsilon}_D \quad (15)$$

$$= 1 - \mu \bar{\epsilon}_D + \sum_{x: q(x)=N} \underbrace{\{\mu p(x | D) - p(x | N)\}}_{T(x)} \quad (16)$$

If $T(x)$ is negative for an x then it will decrease the objective function and the optimal strategy q^* will decide $q^*(x) = N$. This illustrates why the solution to the Neyman Pearson Task has the form

$$\frac{p(x | N)}{p(x | D)} > \mu \quad \Rightarrow \quad q(x) = N, \quad (9)$$

$$\frac{p(x | N)}{p(x | D)} \leq \mu \quad \Rightarrow \quad q(x) = D. \quad (10)$$

Neyman Pearson : Derivation (1)

$$q^* = \min_{q: X \rightarrow K} \sum_{x: q(x) \neq N} p(x | N) \quad \text{subject to:} \quad \sum_{x: q(x) \neq D} p(x | D) \leq \bar{\epsilon}_D. \quad (17)$$

Let us rewrite this as

$$q^* = \min_{q: X \rightarrow K} \sum_{x \in X} \alpha(x) p(x | N) \quad \text{subject to:} \quad \sum_{x \in X} [1 - \alpha(x)] p(x | D) \leq \bar{\epsilon}_D. \quad (18)$$

$$\text{and:} \quad \alpha(x) \in \{0, 1\} \quad \forall x \in X \quad (19)$$

This is a combinatorial optimization problem. If the relaxation is done from $\alpha(x) \in \{0, 1\}$ to $0 \leq \alpha(x) \leq 1$, this can be solved by **linear programming** (LP). The Lagrangian of this problem with inequality constraints is:

$$L(\alpha(x_1), \alpha(x_2), \dots, \alpha(x_N)) = \sum_{x \in X} \alpha(x) p(x | N) + \mu \left(\sum_{x \in X} [1 - \alpha(x)] p(x | D) - \bar{\epsilon}_D \right) \quad (20)$$

$$- \sum_{x \in X} \mu_0(x) \alpha(x) + \sum_{x \in X} \mu_1(x) (\alpha(x) - 1) \quad (21)$$

Neyman Pearson : Derivation (2)

$$L(\alpha(x_1), \alpha(x_2), \dots, \alpha(x_N)) = \sum_{x \in X} \alpha(x)p(x | \mathbf{N}) + \mu \left(\sum_{x \in X} [1 - \alpha(x)]p(x | \mathbf{D}) - \bar{\epsilon}_D \right) \quad (20)$$

$$- \sum_{x \in X} \mu_0(x)\alpha(x) + \sum_{x \in X} \mu_1(x)(\alpha(x) - 1) \quad (21)$$

The conditions for optimality are ($\forall x \in X$):

$$\frac{\partial L}{\partial \alpha(x)} = p(x | \mathbf{N}) - \mu p(x | \mathbf{D}) - \mu_0(x) + \mu_1(x) = 0, \quad (22)$$

$$\mu \geq 0, \mu_0(x) \geq 0, \mu_1(x) \geq 0, \quad 0 \leq \alpha(x) \leq 1, \quad (23)$$

$$\mu_0(x)\alpha(x) = 0, \mu_1(x)(\alpha(x) - 1) = 0, \mu \left(\sum_{x \in X} [1 - \alpha(x)]p(x | \mathbf{D}) - \bar{\epsilon}_D \right) = 0. \quad (24)$$

Case-by-case analysis:

case	implications
$\mu = 0$	L minimized by $\alpha(x) = 0 \quad \forall x$
$\mu \neq 0, \alpha(x) = 0$	$\mu_1(x) = 0 \Rightarrow \mu_0(x) = p(x \mathbf{N}) - \mu p(x \mathbf{D}) \Rightarrow p(x \mathbf{N})/p(x \mathbf{D}) \leq \mu$
$\mu \neq 0, \alpha(x) = 1$	$\mu_0(x) = 0 \Rightarrow \mu_1(x) = -[p(x \mathbf{N}) - \mu p(x \mathbf{D})] \Rightarrow p(x \mathbf{N})/p(x \mathbf{D}) \geq \mu$
$\mu \neq 0,$ $0 < \alpha(x) < 1$	$\mu_0(x) = \mu_1(x) = 0 \Rightarrow p(x \mathbf{N})/p(x \mathbf{D}) = \mu$

Neyman Pearson : Derivation (3)

Case-by-case analysis:

case	implications
$\mu = 0$	L minimized by $\alpha(x) = 0 \quad \forall x$
$\mu \neq 0, \alpha(x) = 0$	$\mu_1(x) = 0 \Rightarrow \mu_0(x) = p(x \text{N}) - \mu p(x \text{D}) \Rightarrow p(x \text{N})/p(x \text{D}) \leq \mu$
$\mu \neq 0, \alpha(x) = 1$	$\mu_0(x) = 0 \Rightarrow \mu_1(x) = -[p(x \text{N}) - \mu p(x \text{D})] \Rightarrow p(x \text{N})/p(x \text{D}) \geq \mu$
$\mu \neq 0,$ $0 < \alpha(x) < 1$	$\mu_0(x) = \mu_1(x) = 0 \Rightarrow p(x \text{N})/p(x \text{D}) = \mu$

Optimal Strategy for a given $\mu \geq 0$ and particular $x \in X$:

$$\frac{p(x | \text{N})}{p(x | \text{D})} \begin{cases} < \mu & \Rightarrow q(x) = \text{D (as } \alpha(x) = 0) \\ > \mu & \Rightarrow q(x) = \text{N (as } \alpha(x) = 1) \\ = \mu & \Rightarrow \text{LP relaxation does not give the desired solution, as } \alpha \notin \{0, 1\} \end{cases} \quad (25)$$

Neyman Pearson : Note on Randomized Strategies (1)

Consider:

$p(x D)$		
x_1	x_2	x_3
0.9	0.09	0.01

$p(x N)$		
x_1	x_2	x_3
0.09	0.9	0.01

$r(x) = p(x N)/p(x D)$		
x_1	x_2	x_3
0.1	10	1

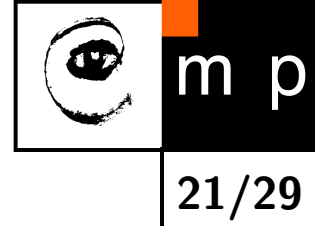
and $\bar{\epsilon}_D = 0.03$.

- ◆ $q_1 : (x_1, x_2, x_3) \rightarrow (D, D, D) \Rightarrow \epsilon_D = 0.00, \epsilon_N = 1.00$
- ◆ $q_2 : (x_1, x_2, x_3) \rightarrow (D, D, N) \Rightarrow \epsilon_D = 0.01, \epsilon_N = 0.99$
- ◆ no other deterministic strategy q is feasible, that is all other ones have $\epsilon_D > \bar{\epsilon}_D$
- ◆ q_2 is the best deterministic strategy but it does not comply with the previous basic result of constructing the optimal strategy because it decides for N for likelihood ratio 1 but decides for D for likelihood ratios 0.01 and 10. Why is that?
- ◆ we can construct a randomized strategy which attains $\bar{\epsilon}_D$ and reaches lower ϵ_N :

$$q(x_1) = q(x_3) = D, \quad q(x_2) = \begin{cases} N & 1/3 \text{ of the time} \\ D & 2/3 \text{ of the time} \end{cases} \quad (26)$$

For such strategy, $\epsilon_D = 0.03, \epsilon_N = 0.7$.

Neyman Pearson : Note on Randomized Strategies (2)



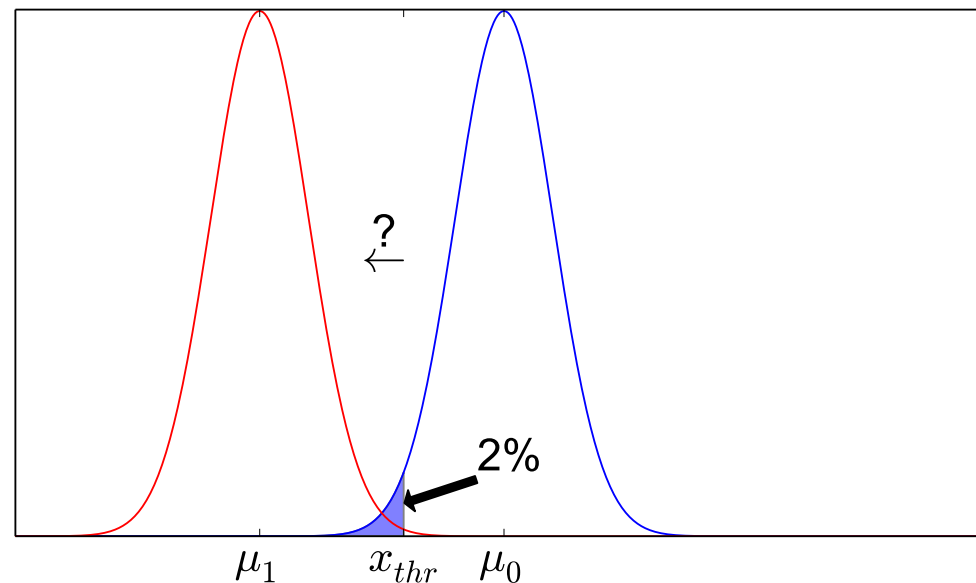
- ◆ This is not a problem but a feature which is caused by discrete nature of X (does not happen when X is continuous).
- ◆ This is exactly what the case of $\mu = p(x | N)/p(x | D)$ is on slide 18.

Neyman Pearson : Notes (1)

- ◆ The task can be generalized to 3 hidden states, of which 2 are dangerous, $K = \{N, D_1, D_2\}$. It is formulated as an analogous problem with two inequality constraints and minimization of classification error for N.
- ◆ Neyman's and Pearson's work dates to 1928 and 1933.
- ◆ A particular strength of the approach lies in that the likelihood ratio $r(x)$ or even $p(x | N)$ need not be known. For the task to be solved, it is enough to know the $p(x | D)$ and the **rank order** of the likelihood ratio (to be demonstrated on the next page)

Neyman Pearson : Notes (2)

- Consider a medicine for reducing weight which is taken for one month. The control group taking placebos has known distribution of weight change $p(x | D)$ as shown in blue. Let it be normal, $p(x | D) = \mathcal{N}(x | \mu_0, \sigma)$. The distribution of weight change after 1 month of taking the real pill is assumed to be normal as well, with the same variance but unknown shift of mean to the left, $p(x | N) = \mathcal{N}(x | \mu_1, \sigma)$, with $\mu_1 < \mu_0$ but otherwise unknown (shown in red). The likelihood ratio is $r(x) = \exp \frac{1}{2\sigma^2} (-(x - \mu_1)^2 + (x - \mu_0)^2) = \exp \left(\frac{1}{\sigma^2} (\mu_1 - \mu_0)x + \text{const} \right)$. It is thus decreasing (monotone) with x (irrespective of μ_1 , $\mu_1 < \mu_0$).
- Setting $\bar{\epsilon}_D = 0.02$, we go along the decreasing $r(x)$ and find the point x_{thr} for which $\int_{-\infty}^{x_{thr}} p(x | D) = \bar{\epsilon}_D = 0.02$ (0.02-quantile). Note that the threshold μ on $r(x)$ is still unknown as $p(x | N)$ is unknown.



Minimax Task

- ◆ $K = \{1, 2, \dots, N\}$
- ◆ X set of observations
- ◆ Conditionals $p(x | k)$ are known $\forall k \in K$
- ◆ The priors $p(k)$ are unknown or do not exist
- ◆ $q: X \rightarrow K$ strategy

The Minimax Task looks for the optimum strategy q^* which minimizes the classification error of the worst classified class:

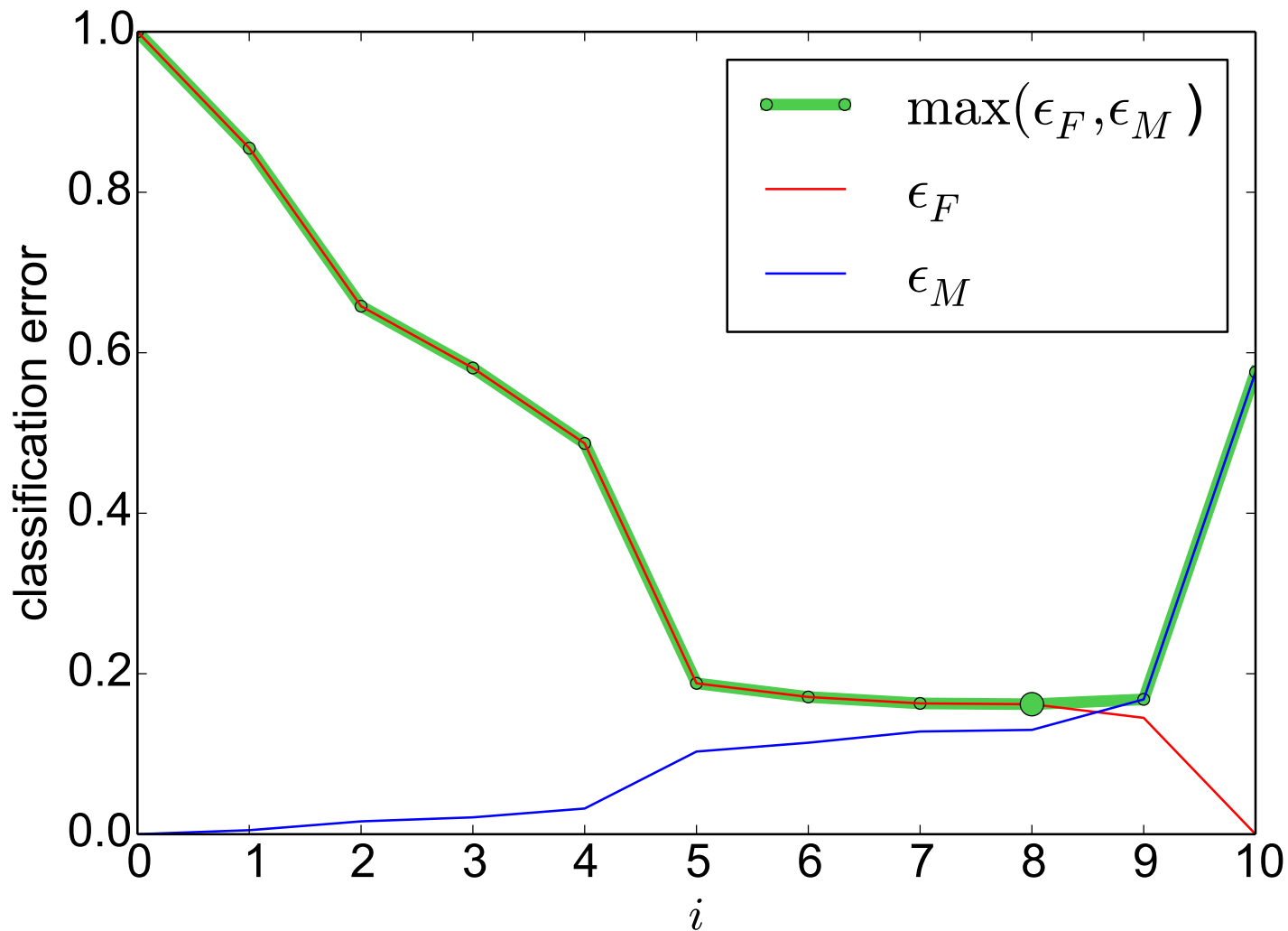
$$q^* = \operatorname{argmin}_{q: X \rightarrow K} \max_{k \in K} \epsilon(k), \quad \text{where} \quad (27)$$

$$\epsilon(k) = \sum_{x: q(x) \neq k} p(x | k) \quad (28)$$

- ◆ Example: A recognition algorithm qualifies for a competition using preliminary tests. During the final competition, only objects from the hardest-to-classify class are used.
- ◆ For a 2-class problem, the strategy is again constructed using the likelihood ratio.
- ◆ In the case of continuous observations space X , equality of classification errors is attained: $\epsilon_1 = \epsilon_2$
- ◆ The derivation can again be done using Linear Programming.

Example: Male/Female Recognition (Minimax)

Classification errors for F and M, for $\mu_i = \frac{r_i+r_{i+1}}{2}$ and $\mu_0 = 0$.



The optimum is attained for $i = 8$, $\epsilon_F = .162$, $\epsilon_M = .13$. The corresponding strategy is as shown on slide [11](#).

Minimax: Comparison with Bayesian Decision with Unknown Priors

- ◆ Consider the same setting as in the Minimax task, but let the priors $p(k)$ exist but be unknown.
- ◆ The Bayesian error ϵ for strategy q is

$$\epsilon = \sum_k \sum_{x: q(x) \neq k} p(x, k) = \sum_k p(k) \underbrace{\sum_{x: q(x) \neq k} p(x | k)}_{\epsilon(k)} \quad (29)$$

- ◆ We want to minimize ϵ but we do not know $p(k)$'s. What is the maximum it can attain? Obviously, the $p(k)$'s do the convex combination of the class errors $\epsilon(k)$; the maximum Bayesian error will be attained when $p(k) = 1$ for the class k with the highest class error $\epsilon(k)$.
- ◆ Thus, to minimize the Bayesian error ϵ under this setting, the solution is to minimize the error of the hardest-to-classify class.
- ◆ Therefore, Minimax formulation and the Bayesian formulation with Unknown Priors lead to the same solution.

Wald Task (1)

- ◆ Let us consider classification with two states, $K = \{1, 2\}$.
- ◆ We want to set a threshold ϵ on the classification error of both of the classes: $\epsilon_1 \leq \epsilon$, $\epsilon_2 \leq \epsilon$.
- ◆ As the previous analysis shows (Neyman Pearson, Minimax), there may be **no** feasible solution if ϵ is set too low.
- ◆ That is why the possibility of decision “do not know” is introduced. Thus $D = K \cup \{?\}$
- ◆ A strategy $q : X \rightarrow D$ is characterized by:

$$\epsilon_1 = \sum_{x: q(x)=2} p(x | 1) \quad (\text{classification error for 1}) \quad (30)$$

$$\epsilon_2 = \sum_{x: q(x)=1} p(x | 2) \quad (\text{classification error for 2}) \quad (31)$$

$$\kappa_1 = \sum_{x: q(x)=?} p(x | 1) \quad (\text{undecided rate for 1}) \quad (32)$$

$$\kappa_2 = \sum_{x: q(x)=?} p(x | 2) \quad (\text{undecided rate for 2}) \quad (33)$$

Wald Task (2)

- ◆ The optimal strategy q^* :

$$q^* = \operatorname{argmin}_{q: X \rightarrow D} \max_{i=\{1,2\}} \kappa_i \quad (34)$$

$$\text{subject to: } \epsilon_1 \leq \epsilon, \epsilon_2 \leq \epsilon \quad (35)$$

- ◆ The task is again solvable using LP (even for more than 2 classes)
- ◆ The optimal solution is again based on the likelihood ratio

$$r(x) = \frac{p(x | 1)}{p(x | 2)} \quad (36)$$

- ◆ The optimal strategy is constructed using suitably chosen thresholds μ_l and μ_h such that:

$$q(x) = \begin{cases} 2 & \text{for } r(x) < \mu_l \\ 1 & \text{for } r(x) > \mu_h \\ ? & \text{for } \mu_l \leq r(x) \leq \mu_h \end{cases} \quad (37)$$

Example: Male/Female Recognition (Wald)

Solve the Wald task for $\epsilon = 0.05$.

$p(x|F)$

	short	.197	.145	.094	.017
	normal	.077	.299	.145	.017
	tall	.001	.008	.000	.000
	u-light				
	light				
	avg				
	heavy				

$p(x|M)$

	short	.011	.005	.011	.011
	normal	.005	.071	.408	.038
	tall	.002	.014	.255	.169
	u-light				
	light				
	avg				
	heavy				

$r(x) = p(x|M)/p(x|F)$

	short	0.056	0.034	0.117	0.647
	normal	0.065	0.237	2.814	2.235
	tall	2.000	1.750	∞	∞
	u-light				
	light				
	avg				
	heavy				

rank, and $q^*(x) = \{F, M, ?\}$

	short	2	1	4	6
	normal	3	5	10	9
	tall	8	7	11	12
	u-light				
	light				
	avg				
	heavy				

Result: $\epsilon_M = 0.032$, $\epsilon_F = 0$, $\kappa_M = 0.544$, $\kappa_F = 0.487$

$(r_4 < \mu_l < r_5, r_{10} < \mu_h < \infty)$