1. $\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i=j \quad i,j=1,\dots,c \\ \lambda_r & i=c+1 \quad \text{(rejection)} \\ \lambda_s & \text{otherwise} \end{cases}$

a) Minimum risk with:

$$\text{Decide} \begin{cases} \omega_i & \text{if } P(\omega_i|x) \geq P(\omega_j|x) \text{ for all } j, \text{ and if } P(\omega_i|x) \geq 1-\frac{\lambda_r}{\lambda_s} \\ \text{reject otherwise} \end{cases}$$

For $i=1,\dots,c$ the loss for $i=j$ is zero and substitution error does not vary across values of i and j. This means the function can be represented as a

$$\text{Decide} \begin{cases} 0 & i=j \quad i,j=1,\dots,c \\ 1 & i \neq j \end{cases} \quad \underline{\lambda_s \text{ the same across } 1,\dots,c}$$

and has a Minimum risk when $P(\omega_i|x)$ is maximized or

$\boxed{P(\omega_i|x) \geq P(\omega_j|x)}$

for $i=c+1$, values are simply rejected so $P(\omega_j|x) = 1$

$$\lambda_r P(\omega_j|x) + \lambda_s P(\omega_i|x) \geq \lambda_s P(\omega_j|x)$$

$$P(\omega_i|x) \geq \frac{\lambda_s - \lambda_r}{\lambda_s} \cdot 1$$

$$\boxed{P(\omega_i|x) \geq 1 - \frac{\lambda_r}{\lambda_s}}$$

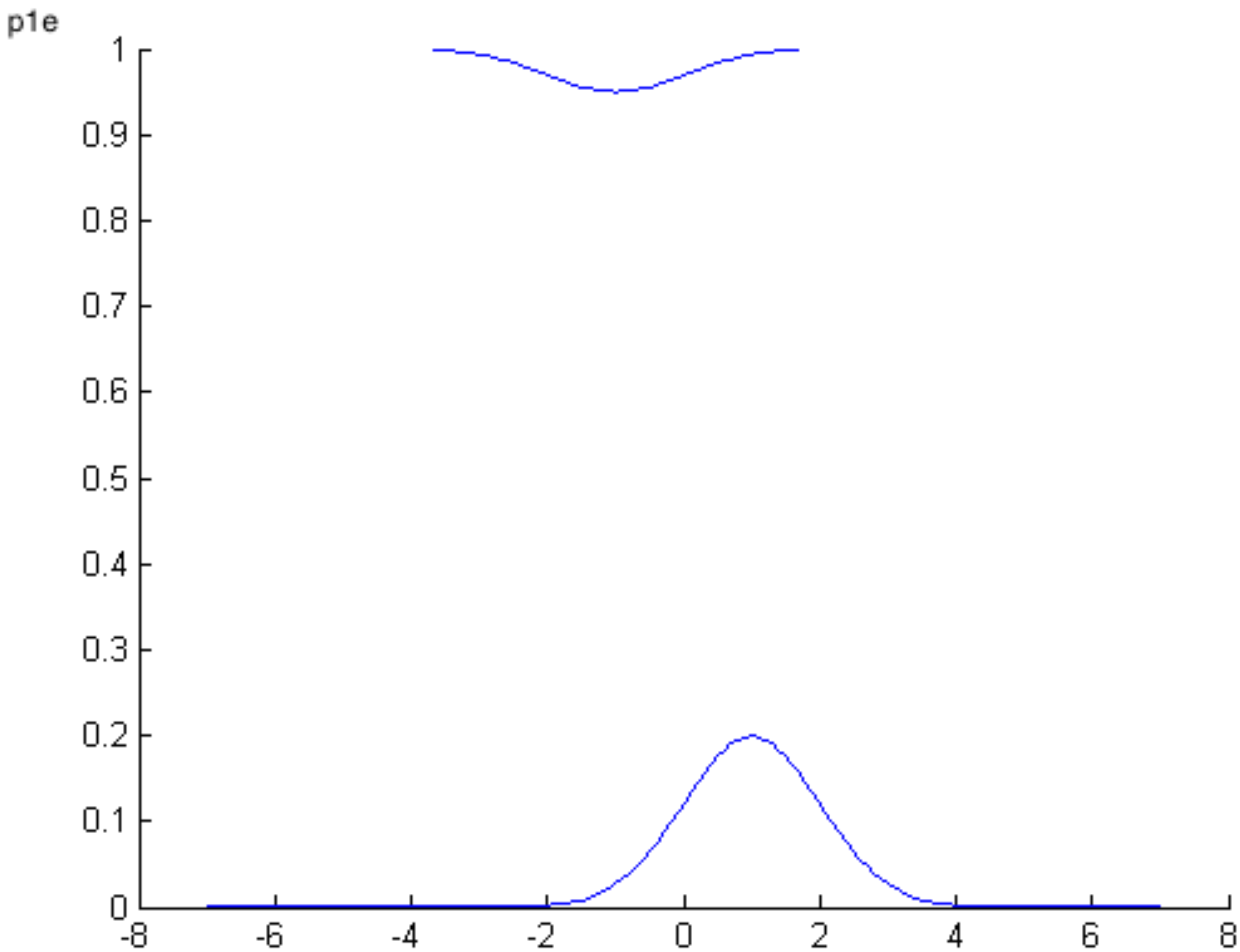
b) when $\lambda_r = 0$ there is no penalty for rejection so things are only classified when $P(\omega_i|x) \geq 1$

c) if the penalty for rejection is too high, $P(\omega_i|x) \geq P(\omega_j|x)$ determines all classification

1. d) Since values under c have no possibility for rejection so $p(x|\omega_i)P(\omega_i)$ governs classification.
for value of $i = c+1$, rejection of values can occur.
Since ω_i is being rejected, all likelihood functions to that "c" point are summed and weighed against the difference between substitution cost and rejection cost over rejection cost

ie

$$\frac{\lambda_s - \lambda_r}{\lambda_s} \sum_{j=1}^{c} p(x|\omega_j)P(\omega_j)$$



p1e
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-8
-6
-4
-2
0
2
4
6
8

2. a) $p_{ij} = P(x_i = 1 | \omega_j)$ for $i = 1, \ldots, d$ and $j = 1, \ldots, c$

p_{ij} is the likelihood of a component of a vector (x_i) in class ω_j of being one. Since this is a binary vector, the likelihood of $x_i = 0$ in ω_j is $1 - p_{ij}$.

b) Minimum error rate is at

$$g_j(x) = \sum_{i=1}^{d} x_i \ln\left(\frac{p_{ij}}{1-p_{ij}}\right) + \sum_{i=1}^{d} \ln(1-p_{ij}) + \ln(P(\omega_j))$$

Satisfying minimum error rate with maximum posterior:

$$g_j(x_i) = P(\omega_j | x_i) = p(x_i | \omega_j) * P(\omega_j)$$

$$= \ln p(x_i|\omega_j) + \ln(P(\omega_j))$$

Looks like MER $\longrightarrow$ $= \sum_{i=1}^{d} x_i \ln(p_{ij}) + \ln(P(\omega_j))$

but x_i is binary so a way to add weight to "0" values that would otherwise not show in the function

$$= \sum_{i=1}^{d} x_i \frac{p_{ij}}{(1-p_{ij})} \cdot (1-p_{ij}) + \ln(P(\omega_j))$$

$$= \sum_{i=1}^{d} x_i \frac{p_{ij}}{(1-p_{ij})} + \sum_{i=1}^{d} \ln(1-p_{ij}) + \ln(P(\omega_j))$$

By multiplying the "0" likelihood $(1-p_{ij})$, the minimum error conditions aren't lost and additional errors are reduced by representing additional components of the vector. In an extreme case where all component of x_i are 0 and it belongs to class ω_j, $\sum_{i=1}^{d} x_i \ln(p_{ij})$ will go to zero. $\sum_{i=1}^{d} \ln(1-p_{ij})$ assures zero values are represented.

$\sigma^2_{ML} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \mu_{ML})^2 \quad \mu_{ML} = \frac{1}{n}\sum_{k=1}^{n} x_k$

3. Assume samples drawn independently from a univariate Gaussian dist.

derriving μ_{ML}, σ^2_{ML}

ML condition: $\nabla_\theta \ln p(x_k|\theta) = \sum_{k=1}^{n} \nabla_\theta (\ln p(x_k|\theta)) = 0$

$$\ln p(x_k|\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}(x_k - \mu)^2$$

$$\nabla_\theta \ln p(x_k|\theta) = \frac{\delta}{\delta\mu}\ln p(x_k|\theta), \frac{\delta}{\delta\sigma^2}\ln p(x_k|\theta)$$

$$= \frac{1}{\sigma^2}(x_k - \mu), \; -\frac{1}{2\sigma^2} + \frac{(x_k - \mu)^2}{2(\sigma^2)^2}$$

ML condition: $0 = \sum_{k=1}^{n}\frac{1}{\sigma^2}(x_k - \mu_{ML}), \; 0 = -\sum_{k=1}^{n}\frac{1}{2\sigma} + \sum_{k=1}^{n}\frac{(x_k - \mu_{ML})^2}{2(\sigma^2_{ML})^2}$

$$\begin{cases} -\sum_{k=1}^{n}\frac{1}{2\sigma_{ML}} + \sum_{k=1}^{n}\frac{(x_k - \mu_{ML})^2}{2(\sigma^2_{ML})^2} = 0 \\ \sum_{k=1}^{n}\frac{1}{\sigma^2}(x_k - \mu_{ML}) = 0 \end{cases}$$

$$0 = -\frac{n}{2\sigma^2_{ML}} + \sum_{k=1}^{n}\frac{(x_k - \mu_{ML})^2}{2(\sigma^2_{ML})^2}$$

$$0 = \frac{1}{2\sigma^2_{ML}}\left(-n + \sum_{k=1}^{n}\frac{(x_k - \mu_{ML})^2}{(\sigma^2_{ML})}\right)$$

$$n = \frac{1}{\sigma^2_{ML}}\sum_{k=1}^{n}(x_k - \mu_{ML})^2$$

$$\boxed{\sigma^2_{ML} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \mu_{ML})^2}$$ ... see back

$$0 = \frac{-n}{2\sigma^2} + \sum_{k=1}^{n} \frac{(x_k - \mu_{ML})^2}{2(\sigma^2)^2} \quad (\text{eq } 1)$$

$$0 = \sum_{k=1}^{n} \frac{1}{\sigma^2}(x_k - \mu) \quad (\text{eq } 2)$$

$$\frac{1}{\sigma^2} \sum (x_k - \mu_{ML}) = \frac{-n}{2\sigma^2} + \frac{n}{2(\sigma^2)^2} \cdot \frac{1}{n} \sum_{k=1}^{n} (x_k - \mu_{ML})^2$$

$$\frac{1}{\cancel{\sigma^2}} \sum_{k=1}^{n} (x_k - \mu_{ML}) = \frac{-n}{2\cancel{\sigma^2}} + \frac{n}{2(\cancel{\sigma^2})^2} \cdot \cancel{\sigma^2}$$

$$\sum_{k=1}^{n} x_k - n \cdot \mu_{ML} = \cancel{\frac{-n}{2} + \frac{n}{2}}$$

$$\sum_{k=1}^{n} x_k = n \cdot \mu_{ML}$$

$$\boxed{\frac{1}{n} \sum_{k=1}^{n} x_k = \mu_{ML}}$$

Alternatively:

$$0 = \sum_{k=1}^{n} \frac{1}{\sigma^2}(x_k - \mu_{ML})$$

$$\sigma^2 \cdot 0 = \frac{1}{\sigma^2} \cdot \sigma^2 \cdot \sum_{k=1}^{n} x_k - \mu_{ML}$$

$$0 = \sum_{k=1}^{n} x_k - \sum_{k=1}^{n} \mu_{ML}$$

$$0 = \sum_{k=1}^{n} x_k - n \cdot \mu_{ML}$$

$$n \mu_{ML} = \sum_{k=1}^{n} x_k$$

$$\boxed{\mu_{ML} = \frac{1}{n} \sum_{k=1}^{n} x_k}$$