4.4

(a) The minimum probability of error rule is

\[ \hat{w} = \arg \max_i \ p(x \mid w_i) \ p(w_i) \]

The PNN rule is

\[ \hat{w} = \arg \max_i \ \sum_{j=1}^n a_{ji} e^{-\frac{(x_j^T x - 1)^2}{2\sigma^2}} \]

\[ |x_j| = 1 \quad \text{and} \quad |x| = 1 \] by design, so

\[ e^{-\frac{|x_j - x|^2}{2\sigma^2}} = e^{-\frac{|x_j - x|^2}{2\sigma^2}} \]

\[ a_{ji} = \begin{cases} 1 & \text{if } w(x_j) = w_i \\ 0 & \text{else} \end{cases} \]

so \( \hat{w} = \arg \max_i \ \sum_{j=1}^n a_{ji} e^{-\frac{|x_j - x|^2}{2\sigma^2}} \)

where \( n_i = \# \text{samples in class } w_i \)

\[ n_i \leq n \]

so \( \hat{w} = \arg \max_i \left( \frac{n_i}{n} \right) \left( \sum_{j=1}^{n_i} \frac{1}{n_i} e^{-\frac{|x_j - x|^2}{2\sigma^2}} \right) \)

\[ = \arg \max_i \left( \frac{n_i}{n} \right) \left( \frac{1}{n_i} \sum_{j=1}^{n_i} e^{-\frac{|x_j - x|^2}{2\sigma^2}} \right) \]

\[ = \arg \max_i \hat{p}(w_i) \hat{p}(x \mid w_i) \]

\( \hat{p}(w_i) \) is the ML estimate, \( \hat{p}(x \mid w_i) \) is the Parzen window estimate w/Gaussian Q
The minimum Bayes risk rule is

\[ \hat{\omega} = \arg \min_i \sum_{j=1}^c \lambda_{i,j} \cdot p(x|w_j) \cdot p(w_j) \]

\[ = \arg \max_i \sum_{j=1}^c (\lambda_{\max} - \lambda_{i,j}) \cdot p(x|w_j) \cdot p(w_j) \]

**MODEL:**

\[ p(w_j) \rightarrow \hat{p}_{ML}(w_j) = \frac{n_j}{n} \]

\[ p(x|w_j) \rightarrow \hat{p}_{PARZ}(x|w_j) = \frac{1}{n_j} \sum_{k=1}^{n_j} e^{-\frac{1}{2} \frac{(x-x_k)^2}{2\sigma^2}} \]

\[ \hat{\omega} = \arg \max_i \left( \sum_{j=1}^c (\lambda_{\max} - \lambda_{i,j}) \left( \frac{n_j}{n} \right) \sum_{k=1}^{n_j} e^{-\frac{1}{2} \frac{(x-x_k)^2}{2\sigma^2}} \right) \]

\[ = \arg \max_i \sum_{k=1}^n a_{i,k} e^{-\frac{1}{2} \frac{(x-x_k)^2}{2\sigma^2}} \]

**FOR**

\[ a_{i,k} = \lambda_{\max} - \lambda_{i,j} \quad \Rightarrow \quad w_j = \omega(x_k) \]

**EQUALLY, WE COULD SET**

\[ \hat{\omega} = \arg \min_i \sum_{k=1}^n a_{i,k} e^{-\frac{1}{2} \frac{(x-x_k)^2}{2\sigma^2}} \]

\[ a_{i,k} = \lambda_{i,j} \quad \Rightarrow \quad w_j = \omega(x_k) \]

**ONE WAY TO DO THIS:** USE \( w_j \) TOKENS FROM CLASS \( j \) TO TRAIN OUTPUT NODE \( \omega_j \) WHERE \( w_j = (\lambda_{\max} - \lambda_{i,j}) n_j \). FOR EACH SUCH TOKEN, \( a_{i,k} = \frac{w_j}{n_j} \). THEN:

\[ \hat{\omega} = \arg \max_i \sum_{k=1}^n a_{i,k} \cdot \phi \left( \frac{x-x_k}{\sigma} \right) = \sum_{j=1}^c \left( \frac{n_j}{n} \right) \sum_{k=1}^{n_j} \phi \left( \frac{x-x_k}{\sigma} \right) \]
4.27
\[ D(S_1, S_2) = \frac{n_1 + n_2 - 2n_{12}}{n_1 + n_2 - n_{12}} \]

\[ D(S_1, S_2) \geq 0 \quad \text{iff} \quad n_{12} < n_1 \quad n_{12} < n_2 \]

**Non-Negative:**
\[ D(S_1, S_2) > 0 \]

**Reflexivity:**
\[ D(S_1, S_2) = 0 \quad \text{iff} \quad n_{12} = n_1 = n_2 \]

which happens iff \( S_1, S_2 \) have same members

**Symmetry:**
\[ D(S_2, S_1) = \frac{n_2 + n_1 - 2n_{12}}{n_2 + n_1 - n_{12}} = D(S_1, S_2) \]

**Triangle Inequality:**
\[ D(S_1, S_2) + D(S_2, S_3) \geq D(S_1, S_3) \quad \text{iff} \]
\[ \frac{m_{12} - n_{12}}{m_{12}} + \frac{m_{23} - n_{23}}{m_{23}} - \frac{m_{13} - n_{13}}{m_{13}} \geq 0 \]

for \( m_{ij} = n_i + n_j - n_{ij} \)

**Notice:**
\[ n_{13} \geq n_{12} + n_{23} - n_2 \]
\[ m_{13} \leq m_{12} + m_{23} - n_2 \]

so \( D(S_1, S_2) + D(S_2, S_3) - D(S_1, S_3) \geq \frac{m_{12} - n_{12}}{m_{12}} + \frac{m_{23} - n_{23}}{m_{23}} - \frac{m_{13} - n_{13}}{m_{13}} \)
so \( D(S_1, S_2) + D(S_2, S_3) - D(S_1, S_3) \geq 0 \)

iff \( \frac{m_{12} - n_{12}}{m_{12}} + \frac{m_{23} - n_{23}}{m_{23}} - \frac{m_{12} + m_{23} - n_{12} - n_{23}}{m_{12} + m_{23} - n_{2}} \geq 0 \)

iff \( (2m_{12}m_{23} - m_{23}n_{12} - m_{12}n_{23})(m_{12} + m_{23} - n_{2}) \)
\[ - m_{12}m_{23}(m_{12} + m_{23} - n_{12} - n_{23}) \geq 0 \]

iff \( m_{12}m_{23}(m_{12} + m_{23}) - m_{23}^{2}n_{12} - m_{12}^{2}n_{23} \)
\[ - n_{2}(2m_{12}m_{23} - m_{23}n_{12} - m_{12}n_{23}) \geq 0 \]

iff \( m_{12}^{2}(m_{23} - n_{23}) + m_{23}^{2}(m_{12} - n_{12}) \)
\[ - n_{2}((m_{12} - n_{12})m_{23} + (m_{23} - n_{23})m_{12}) \geq 0 \]

iff \( m_{12}(m_{23} - n_{23})(m_{12} - n_{2}) + m_{23}(m_{12} - n_{12})(m_{23} - n_{2}) \geq 0 \)

but \( m_{23} - n_{23} = n_{2} + n_{3} - 2n_{23} \geq 0 \)
\( m_{12} - n_{12} = n_{1} + n_{2} - 2n_{12} \geq 0 \)
\( m_{23} - n_{2} = n_{3} - n_{23} \geq 0 \)
\( m_{12} - n_{2} = n_{1} - n_{12} \geq 0 \)

\[ \therefore D(S_1, S_2) + D(S_2, S_3) - D(S_1, S_3) \geq 0 \]

\( \Rightarrow \) All 4 criteria satisfied. \( D_{\text{d\text{-}ani\text{-}moto}} \) is a metric.
### Table

<table>
<thead>
<tr>
<th></th>
<th>Pattern</th>
<th>Pat</th>
<th>Pots</th>
<th>Stop</th>
<th>Taxonomy</th>
<th>Elementary</th>
</tr>
</thead>
<tbody>
<tr>
<td>pattern</td>
<td>0</td>
<td>4/7</td>
<td>7/9</td>
<td>7/9</td>
<td>9/12</td>
<td>7/12</td>
</tr>
<tr>
<td>pat</td>
<td>4/7</td>
<td>0</td>
<td>3/5</td>
<td>3/5</td>
<td>7/9</td>
<td>9/11</td>
</tr>
<tr>
<td>pots</td>
<td>7/9</td>
<td>3/5</td>
<td>0</td>
<td>0</td>
<td>8/10</td>
<td>12/13</td>
</tr>
<tr>
<td>stop</td>
<td>7/9</td>
<td>3/5</td>
<td>0</td>
<td>0</td>
<td>8/10</td>
<td>12/13</td>
</tr>
<tr>
<td>taxonomy</td>
<td>9/12</td>
<td>7/9</td>
<td>8/10</td>
<td>8/10</td>
<td>8/13</td>
<td>0</td>
</tr>
<tr>
<td>elementary</td>
<td>7/12</td>
<td>9/11</td>
<td>12/13</td>
<td>12/13</td>
<td>8/13</td>
<td>0</td>
</tr>
</tbody>
</table>

### Rank ordering of positive-distance pairs

1. (elementary, pattern) 7/12 0.583
2. (pattern, pattern) 4/7 0.5
3. (pat, pots), (pat, stop) (pots, taxonomic) (pattern, taxonomic) 3/5 0.6
4. (pat, elementary) 9/12 0.75
5. (taxonomy, pat), (pattern, pots), (patter, stop) (taxonomy, pots), (taxonomy, stop) 7/9 0.777
6. (pat, elementary) 3/11 0.817
7. (pots, elementary), (stop, elementary) 12/13 0.817

**Yes, D\text{\texttt{A\texttt{R\texttt{A}\texttt{N\texttt{I\texttt{M\texttt{E}}}}}}} always obeys Δ inequality. In this case, all non-zero distances are between 1/2 and 1, so it's trivially satisfied.
A PWL machine implements

\[ R_i = \{ x : z = \max_i g_i(x) \} \]

\[ g_i(x) = \max_j g_{ij}(x) \]

\[ g_{ij}(x) = w_{ij}^T x + w_{ij0} \]

Consider the linear machine

\[ R_{ij} = \{ x : y_{ij} = \max_{i,j} g_{ij}(x) \} \]

\( R_{ij} \) are convex, mutually exclusive, collectively exhaustive, with hyperplane boundaries.

\[ R_i = R_{i1} \cup R_{i2} \cup \ldots \]

\( R_{1n} \) is the union of arbitrary such regions, thus it can be multiply connected.

**Proof by example:**

\[ \mathbf{w}_{11} = [1, 0] \]

\[ w_{12} = [-1, 0] \]

\[ w_{21} = [0, 1] \]

\[ w_{22} = [0, -1] \]

Shaded = \( R_1 \)

\[ g_{11}(x) = x - 2 \]

\[ g_{12}(x) = -x \]

\[ g_{21}(x) = 0.3 \]
where $J_q(a(k)) = (a^T y_1 - b)^2$,

$$\nabla_q\ J_q = 2 (a^T y_1 - b) y_1$$

**Proof:**

$$\frac{\partial J_q}{\partial a_i} = \frac{\partial}{\partial a_i} \left( \sum_j a_j y_j - b \right)^2 = 2 \left( \sum_j a_j y_j - b \right) y_i$$

$$\frac{\partial^2 J_q}{\partial a_i \partial a_j} = 2 y_j y_i \quad \text{so} \quad H_a J_q = 2 y_i y_i^T$$

Eq. 12:

$$a(k+1) = a(k) - \eta(k) \nabla J(a(k))$$

$$\eta(k) = \frac{\| \nabla J \|^2}{\nabla^T J \nabla J} = \frac{4 (a^T y_1 - b)^2}{4 (a^T y_1 - b)^2 y_1^T (2 y_1 y_1^T) y_1} = \frac{1}{2 |y_1|^2} = \frac{1}{2 |y_1|^2}$$

so

$$a(k+1) = a(k) - \frac{(a^T y_1 - b) y_1}{|y_1|^2}$$