Ex 8.1

I ran the code (modified from Jason Rebello's R in Mathworks file exchange) with $\lambda = 0.001, 0.01, 0.1, 1, 10$.

I did not see big dependence of $\lambda$. Still if I followed the best $\lambda$ from 5 fold cross validation, these are the numbers I got for error rates

<table>
<thead>
<tr>
<th>Normalized</th>
<th>CV training</th>
<th>Test</th>
<th>$\lambda_{opt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.077</td>
<td>0.087</td>
<td>10</td>
</tr>
<tr>
<td>Log</td>
<td>0.056</td>
<td>0.057</td>
<td>10</td>
</tr>
<tr>
<td>Binary</td>
<td>0.070</td>
<td>0.072</td>
<td>1</td>
</tr>
</tbody>
</table>

In different runs, the shallow min error $\lambda$ came out at a different value. The error rates were roughly the same.
Ex 14.1

a) \( \phi(x_1) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \), \( \phi(x_2) = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \)

The vector separating them is \( \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \)

\[ \omega = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \]

\( \omega^T \phi(x_1) = 0 \)

\( \omega^T \phi(x_2) = 4 \times \)

Depending on \( \omega_0 \), \( P + 1 \),

is at \( \frac{1(\omega_0)}{11\omega_0^2} \) and \( P + 2 \) is at \( \frac{1(\omega_2 + \omega_0)}{11\omega_1^2} \)

Minimum of the two

allowed region

Best \( \omega_0 = -2\lambda \)

b)

C) If we insist that \( y(\omega^T \phi(x_2) + \omega_0) \geq 1 \)

Then \( 2\lambda \geq 1 \) or \( \lambda \geq \frac{1}{2} \)
Since margin $\frac{1}{\|w\|} = \frac{1}{\sqrt{2}}$, we have margin $= \sqrt{2}$.

The two vectors are:

$$\phi(x_1) = (1, 1, 0, 0)^T, \quad y_1 = -1$$
$$\phi(x_2) = (1, 2, 1, 2)^T, \quad y_2 = +1$$

$$\min \|w\|^2 \quad \text{s.t.} \quad y_1 (w^T \phi(x_1) + w_0) \geq 1$$
$$y_2 (w^T \phi(x_2) + w_0) \geq 1$$

So, minimize $w_1^2 + w_2^2 + w_3^2$

subject to

$$w_1 + w_0 \leq -1$$

$$w_1 + 2(w_2 + w_3) + w_0 \geq 1$$

First $w_2$ variation gives $\lambda_1 = -\lambda_2 = \lambda$

$$\max \min \left[ w_1^2 + w_2^2 + w_3^2 - 2\lambda (w_2 + w_3 - 1) \right]$$

$$\lambda \geq 0 \quad \nabla w = 0 \quad \implies \quad w_1 = 0$$

$$\frac{\lambda}{w_2} = 1$$

$$\frac{\lambda}{w_3} = 1$$
With only two points, one positive and the other negative, both are support vectors.

So,

\[ y_1 \, w^T \phi(x_1) + w_0 = 1 \]
\[ y_2 \, w^T \phi(x_2) + w_0 = 1 \]

\[ \Rightarrow \quad (-1) \left[ (0, \lambda, 1) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + w_0 \right] = 1 \quad \Rightarrow \quad \overline{w_0} = -1 \]

and

\[ (+1) \left[ (0, \lambda, 1) \begin{pmatrix} 1 \\ 2 \end{pmatrix} + w_0 \right] = 1 \quad \Rightarrow \quad 4 \lambda + w_0 = 0 \]

\[ \lambda = \frac{1}{2} \]

\[ w_0 = -1 \]
\[ \left( \begin{array}{c} w_1 \\ w_2 \\ w_3 \end{array} \right) = \frac{1}{2} \left( \begin{array}{c} 0 \\ 1 \\ 1 \end{array} \right) \]

\[ e) \quad \text{Discriminant function} \]

\[ f(x) = w_0 + w^T \phi(x) \]
\[ = -1 + \frac{x}{\sqrt{2}} + \frac{x^2}{2} \]