Problem 1 Decision Trees

1.1 Classification error for T_1 left leaf: Since class A is the majority with 150 against 50, A will be predicted, leaving 50 B's as error, so $\frac{50}{150+50} = 0.25$

Classification error for T_1 right leaf: Since class B is the majority with 150 against 50, B will be predicted, leaving 50 A's as error, so $\frac{50}{150+50} = 0.25$

Classification error for T_2 left leaf: Since class B is the majority with 100 against 0, B will be predicted, leaving 0 A's as error, so $\frac{0}{0+100} = 0$

Classification error for T_2 right leaf: Since class A is the majority with 200 against 100, A will be predicted, leaving 100 A's as error, so $\frac{100}{100+200} \approx 0.3333$

Entropy for T_1 left leaf:

$$\begin{aligned}
&= -(\frac{150}{150+50} * log\frac{150}{150+50} + \frac{50}{150+50} * log\frac{50}{150+50}) \\
&= -(0.75 * log(0.75) + 0.25 * log(0.25)) \\
&\approx 0.5623
\end{aligned}$$

Entropy for T_1 right leaf:

$$\begin{aligned}
&= -(\frac{200-150}{400-150-50} * log\frac{200-150}{400-150-50} + \frac{200-50}{400-150-50} * log\frac{200-50}{400-150-50}) \\
&= -(0.25 * log(0.25) + 0.75 * log(0.75)) \\
&\approx 0.5623
\end{aligned}$$

Entropy for T_2 left leaf:

$$\begin{aligned}
&= -(\frac{0}{0+100} * log\frac{0}{0+100} + \frac{100}{0+100} * log\frac{100}{0+100}) \\
&= -(0 * log(0) + 1 * log(1)) \\
&= 0
\end{aligned}$$

Entropy for T_2 right leaf:

$$\begin{aligned}
&= -(\frac{200-0}{400-0-100} * log\frac{200-0}{400-0-100} + \frac{200-100}{400-0-100} * log\frac{200-100}{400-0-100}) \\
&\approx -(0.6667 * log(0.6667) + 0.3333 * log(0.3333)) \\
&\approx 0.6365
\end{aligned}$$

Gini impurity for T_1 left leaf:

$$\begin{aligned}
&= \frac{150}{150+50} * (1 - \frac{150}{150+50}) + \frac{50}{150+50} * (1 - \frac{50}{150+50}) \\
&= 0.75 * 0.25 + 0.25 * 0.75 \\
&= 0.375
\end{aligned}$$

Gini impurity for T_1 right leaf:

$$\begin{aligned}
&= \frac{200-150}{400-150-50} * (1 - \frac{200-150}{400-150-50}) + \frac{200-50}{400-150-50} * (1 - \frac{200-50}{400-150-50}) \\
&= 0.25 * 0.75 + 0.75 * 0.25 \\
&= 0.375
\end{aligned}$$

Gini impurity for T_2 left leaf:

$$
\begin{aligned}
&= \frac{0}{0+100} * (1 - \frac{0}{0+100}) + \frac{100}{0+100} * (1 - \frac{100}{0+100}) \\
&= 0 * 1 + 1 * 0 \\
&= 0
\end{aligned}
$$

Gini impurity for T_2 right leaf:

$$
\begin{aligned}
&= \frac{200-0}{400-0-100} * (1 - \frac{200-0}{400-0-100}) + \frac{200-100}{400-0-100} * (1 - \frac{200-100}{400-0-100}) \\
&\approx 0.6667 * 0.3333 + 0.3333 * 0.6667 \\
&\approx 0.4444
\end{aligned}
$$

1.2 Based on classification error:
T_1 has 0.25 error on both leaves, and the average error weighed by branch probabilities is $0.5 * 0.25 + 0.5 * 0.25 = 0.25$
T_2 has 0 error on left leaf and 0.3333 error on right leaf, and the average error weighed by branch probabilities is $0.25 * 0 + 0.75 * 0.3333 = 0.25$
Thus T_1 and T_2 are on draw according to this measure

Based on entropy:
T_1 has 0.5623 entropy on both leaves, and the average entropy weighed by branch probabilities is $0.5 * 0.5623 + 0.5 * 0.5623 = 0.5623$
T_2 has 0 entropy on left leaf and 0.6365 entropy on right leaf, and the average entropy weighed by branch probabilities is $0.25 * 0 + 0.75 * 0.6365 \approx 0.4774$
Thus T_2 is better than T_1 according to this measure

Based on Gini impurity:
T_1 has 0.375 impurity on both leaves, and the average impurity weighed by branch probabilities is $0.5 * 0.375 + 0.5 * 0.375 = 0.375$
T_2 has 0 impurity on left leaf and 0.4444 impurity on right leaf, and the average impurity weighed by branch probabilities is $0.25 * 0 + 0.75 * 0.4444 \approx 0.3333$
Thus T_2 is better than T_1 according to this measure

Problem 2 Boosting

2.1

$$\begin{aligned}
\frac{d}{d(\beta_t)}\epsilon_t(e^{\beta_t}-e^{-\beta_t})+e^{-\beta_t} &= \epsilon_t(e^{\beta_t}+e^{-\beta_t})-e^{-\beta_t}\\
\epsilon_t(e^{\beta_t^*}+e^{-\beta_t^*})-e^{-\beta_t^*} &= 0\\
\frac{\epsilon_t(e^{2\beta_t^*}+1)}{e^{\beta_t^*}} &= e^{-\beta_t^*}\\
e^{2\beta_t^*}+1 &= \frac{1}{\epsilon_t}\\
e^{2\beta_t^*} &= \frac{1-\epsilon_t}{\epsilon_t}\\
\beta_t^* &= \frac{1}{2}ln(\frac{1-\epsilon_t}{\epsilon_t})
\end{aligned}$$

2.2

$$\begin{aligned}
\Sigma_{n:h_t(\mathbf{x}_n)\neq y_n}D_{t+1}(n) &= \Sigma_{n:h_t(\mathbf{x}_n)\neq y_n}D_t(n)e^{-\beta_t y_n h_t(\mathbf{x}_n)}\\
&= \Sigma_{n:h_t(\mathbf{x}_n)\neq y_n}D_t(n)e^{\beta_t}\\
&= e^{\frac{1}{2}ln(\frac{1-\epsilon}{\epsilon})}\Sigma_{n:h_t(\mathbf{x}_n)\neq y_n}D_t(n)\\
&= \sqrt{\frac{1-\epsilon}{\epsilon}}\epsilon\\
&= \sqrt{(1-\epsilon)\epsilon}
\end{aligned}$$

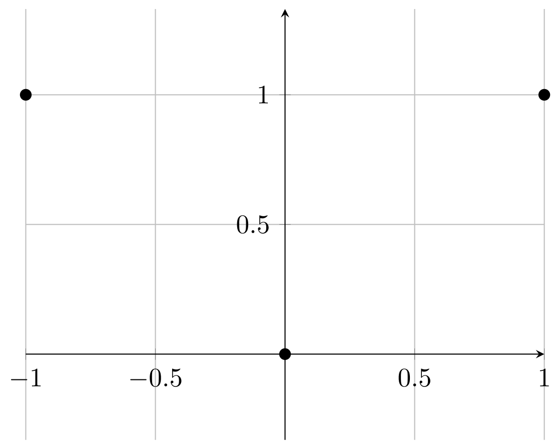
$$\begin{aligned}
\Sigma_{n:h_t(\mathbf{x}_n)=y_n}D_{t+1}(n) &= \Sigma_{n:h_t(\mathbf{x}_n)=y_n}D_t(n)e^{-\beta_t y_n h_t(\mathbf{x}_n)}\\
&= \Sigma_{n:h_t(\mathbf{x}_n)=y_n}D_t(n)e^{-\beta_t}\\
&= e^{-\frac{1}{2}ln(\frac{1-\epsilon}{\epsilon})}\Sigma_{n:h_t(\mathbf{x}_n)=y_n}D_t(n)\\
&= \sqrt{\frac{\epsilon}{1-\epsilon}}(1-\epsilon)\\
&= \sqrt{(1-\epsilon)\epsilon}
\end{aligned}$$

Since $\Sigma_{n:h_t(\mathbf{x}_n)\neq y_n}D_{t+1}(n)+\Sigma_{n:h_t(\mathbf{x}_n)=y_n}D_{t+1}(n)=1$, and $\Sigma_{n:h_t(\mathbf{x}_n)\neq y_n}D_{t+1}(n)=\Sigma_{n:h_t(\mathbf{x}_n)=y_n}D_{t+1}(n)=\sqrt{(1-\epsilon)\epsilon}$, $\Sigma_{n:h_t(\mathbf{x}_n)\neq y_n}D_{t+1}(n)=\frac{1}{2}$.

Problem 3 Support Vector Machines

3.1 No; drawing the linear separator anywhere would at least misclassify 1 point.

3.2 $\phi(x_1)=[-1,1]^T,\quad \phi(x_2)=[1,1]^T,\quad \phi(x_3)=[0,0]^T$; yes; any horizontal line between 0 and 1 can separate the points.



3.3 $k(\mathbf{x},\mathbf{x'}) = \phi(\mathbf{x})^T\phi(\mathbf{x'}) = [x, x^2][x', x'^2]^T = xx' + x^2x'^2 = (\mathbf{x}^T\mathbf{x'} + 1)^2 - \mathbf{x}^T\mathbf{x'} - 1$

$$\mathbf{K} = \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & k(x_1,x_3) \\ k(x_2,x_1) & k(x_2,x_2) & k(x_2,x_3) \\ k(x_3,x_1) & k(x_3,x_2) & k(x_3,x_3) \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$\forall \mathbf{u} \in R^3, \quad \mathbf{u}^T K \mathbf{u} = [2u_1, 2u_2, 0][u_1, u_2, u_3]^T = 2u_1^2 + 2u_2^2 \geq 0$, so $\mathbf{K}$ is PSD.

3.4 $max_{\{\alpha_n\}} \quad \Sigma_n \alpha_n - \frac{1}{2}\Sigma_{m,n} y_m y_n \alpha_m \alpha_n k(x_m, x_n) =$
$max_{\{\alpha_n\}} \quad \alpha_1 + \alpha_2 + \alpha_3 - \alpha_1^2 - \alpha_2^2$
s.t. $-\alpha_1 - \alpha_2 + \alpha_3 = 0; \quad \forall n \quad \alpha_n \geq 0$

3.5

$$\begin{aligned} -\alpha_1 - \alpha_2 + \alpha_3 &= 0 \\ \alpha_3 &= \alpha_1 + \alpha_2 \\ \alpha_1 + \alpha_2 + \alpha_3 - \alpha_1^2 - \alpha_2^2 &= 2\alpha_1 + 2\alpha_2 - \alpha_1^2 - \alpha_2^2 \end{aligned}$$

$$\begin{aligned} \frac{d}{d(\alpha_1)}\alpha_1 + \alpha_2 + \alpha_3 - \alpha_1^2 - \alpha_2^2 = \frac{d}{d(\alpha_1)}2\alpha_1 + 2\alpha_2 - \alpha_1^2 - \alpha_2^2 &= 0 \\ 2 - 2\alpha_1 &= 0 \\ \alpha_1 &= 1 \end{aligned}$$

$$\begin{aligned} \frac{d}{d(\alpha_2)}\alpha_1 + \alpha_2 + \alpha_3 - \alpha_1^2 - \alpha_2^2 = \frac{d}{d(\alpha_2)}2\alpha_1 + 2\alpha_2 - \alpha_1^2 - \alpha_2^2 &= 0 \\ 2 - 2\alpha_2 &= 0 \\ \alpha_2 &= 1 \end{aligned}$$

$$\begin{aligned} \alpha_3 &= \alpha_1 + \alpha_2 \\ &= 1 + 1 \\ &= 2 \end{aligned}$$

$$\begin{aligned} \mathbf{w}^* &= \Sigma_{n:\alpha_n>0}\alpha_n^* y_n \phi(\mathbf{x}_n) \\ &= 1 * (-1) * [-1, 1]^T + 1 * (-1) * [1, 1]^T + 2 * 1 * [0, 0]^T \\ &= [1, -1]^T + [-1, -1]^T \\ &= [0, -2]^T \end{aligned}$$

$$
\begin{aligned}
b^* &= y_n - \Sigma_m y_m \alpha_m^* k(\mathbf{x}_m, \mathbf{x}_n) \\
&= -1 - ((-1) * 1 * 2 + 0 + 0) = 1 \quad \textbf{or} && \text{(taking n = 1)} \\
&= -1 - (0 + (-1) * 1 * 2 + 0) = 1 \quad \textbf{or} && \text{(taking n = 2)} \\
&= 1 - (0 + 0 + 0) = 1 && \text{(taking n = 3)} \\
\frac{1+1+1}{3} &= 1
\end{aligned}
$$

3.6

$$
\begin{aligned}
[0, -2][\phi_1, \phi_2]^T + 1 &= 0 \\
-2\phi_2 + 1 &= 0 \\
\phi_2 &= \frac{1}{2}
\end{aligned}
$$