SET A

Q1

$$\begin{aligned}
\text{Squarred loss: } E_i(w) &= \left(y(x_i) - y_{true}\right)^2 \\
&= \left(w_0 + w_1 x_i + w_2 x_i^2 + w_3 x_i^3 - y_{true}\right)^2 \\
\text{Let: } e_i(w) &= \left(w_0 + w_1 x_i + w_2 x_i^2 + w_3 x_i^3 - y_{true}\right) \\
\text{Computing gradient: } \frac{\partial E}{\partial w_2} &= 2 e_i(w) x_i^2 \\
\text{Weight update: } w_2^{\tau+1} &= w_2^{\tau} - \eta \frac{\partial E}{\partial w_2} \\
&= w_2^{\tau+1} - \eta e_i(w) x_i^2
\end{aligned}$$

Q2

(i)

An epoch is the forward pass and weight update over all the training examples of the data once.

(ii)

Momentum weights both the averaged gradient over the past updates along with the current gradient for performing the current update. This smooths out the steps of gradient descent and helps in achieving faster convergence with lesser oscillations.

Q3

While learning a Decision Tree classifier, the validation dataset can be used to prevent overfitting in the following two ways-
1. Parameter tuning - One could use the validation data to tune the parameters that control the configuration of the tree like the maximum depth, number of leaf nodes etc, minimum number of samples required to be at a leaf node etc.
2. Pruning - In pruning, the branches of the tree are trimmed off, i.e., remove the decision nodes starting from the leaf node such that the performance on the validation dataset is optimized. It can be done while growing the tree(pre-pruning) or by post-pruning the overfitted tree.

SET B

Q1

$$\begin{aligned}
\text{Squarred loss: } E_i(w) &= \left(y(x_i) - y_{true}\right)^2 \\
&= \left(w_0 + w_1 x_i + w_2 x_i^2 + w_3 x_i^3 - y_{true}\right)^2 \\
\text{Let: } e_i(w) &= \left(w_0 + w_1 x_i + w_2 x_i^2 + w_3 x_i^3 - y_{true}\right) \\
\text{Computing gradient: } \frac{\partial E}{\partial w_3} &= 2 e_i(w) x_i^3 \\
\text{Weight update: } w_3^{\tau+1} &= w_3^{\tau} - \eta \frac{\partial E}{\partial w_3} \\
&= w_3^{\tau+1} - \eta e_i(w) x_i^3
\end{aligned}$$

Q2

(i)

Backpropogation is using chain rule to obtain the gradients of all the weights w.r.t to the loss. The gradient is first calculated at the last layer and this is then propagated backwards towards the first layer.

(ii)

Momentum weights both the averaged gradient over the past updates along with the current gradient for performing the current update. This smooths out the steps of gradient descent and helps in achieving faster convergence with lesser oscillations.

Q3

While learning a Decision Tree classifier, the validation dataset can be used to prevent overfitting in the following two ways-
1. Parameter tuning - One could use the validation data to tune the parameters that control the configuration of the tree like the maximum depth, number of leaf nodes etc, minimum number of samples required to be at a leaf node etc.
2. Pruning - In pruning, the branches of the tree are trimmed off, i.e., remove the decision nodes starting from the leaf node such that the performance on the validation dataset is optimized. It can be done while growing the tree(pre-pruning) or by post-pruning the overfitted tree.