Appendices

A) First-Order Approximation Gradient Descent Boosting

Derivation of the partial derivative of the risk function with respect to ϵ:

$$\begin{aligned}
-\frac{\partial R[f^t+\epsilon g+\delta h]}{\partial \epsilon} &= -\frac{\partial}{\partial \epsilon}\sum_{i=1}^{n} L_M[y_i, f^t(x_i)+\epsilon g(x_i)+\delta h(x_i)] \\
&= -\sum_{i=1}^{n}\frac{\partial}{\partial \epsilon}\sum_{k=1}^{M} e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i)+\delta h(x_i), y_i-y^k>} \\
&= -\sum_{i=1}^{n}\sum_{k=1}^{M}\Big(\frac{\partial}{\partial \epsilon} e^{-\frac{1}{2}\epsilon<g(x_i),y_i-y^k>}\Big) e^{-\frac{1}{2}<f^t(x_i)+\delta h(x_i),y_i-y^k>} \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i), y_i-y^k > e^{-\frac{1}{2}\epsilon<g(x_i),y_i-y^k>} e^{-\frac{1}{2}<f^t(x_i)+\delta h(x_i),y_i-y^k>} \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i), y_i-y^k > e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i)+\delta h(x_i),y_i-y^k>}.
\end{aligned}$$

Its value at $\epsilon=\delta=0$:

$$\begin{aligned}
-\frac{\partial R[f^t+\epsilon g+\delta h]}{\partial \epsilon}\Big|_{\substack{\epsilon=0\\ \delta=0}} &= \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i), y_i-y^k > e^{-\frac{1}{2}<f^t(x_i),y_i-y^k>} \\
&= \frac{1}{2}\sum_{i=1}^{n} < g(x_i), \sum_{k=1}^{M}(y_i-y^k)e^{-\frac{1}{2}<f^t(x_i),y_i-y^t>} > \\
&= \sum_{i=1}^{n} < g(x_i), w_i >,
\end{aligned}$$

where:

$$\begin{aligned}
w_i &= \frac{1}{2}\sum_{k=1}^{M}(y_i-y^k)e^{-\frac{1}{2}<f^t(x_i),y_i-y^k>} \\
&= \frac{1}{2}e^{-\frac{1}{2}<f^t(x_i),y_i>}\sum_{k=1}^{M}(y_i-y^k)e^{\frac{1}{2}<f^t(x_i),y^k>}.
\end{aligned}$$

Similarly we can derive with respect to δ:

$$-\frac{\partial R[f^t + \epsilon g + \delta h]}{\partial \delta}\Big|_{\substack{\epsilon=0\\ \delta=0}} = \sum_{i=1}^{n} < h(x_i), w_i >,$$

where w_i is defined as before.

B) Second-Order Approximation Gradient Descent Boosting

Derivation of the second partial derivative of the risk function with respect to ϵ:

$$\begin{aligned}
\frac{\partial^2 R[f^t + \epsilon g + \delta h]}{\partial \epsilon^2} &= \frac{\partial}{\partial \epsilon}\Bigg[-\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i), y_i - y^k > e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i)+\delta h(x_i), y_i - y^k>} \Bigg] \\
&= \frac{\partial}{\partial \epsilon}\Bigg[-\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i), y_i - y^k > e^{-\frac{1}{2}\epsilon<g(x_i), y_i - y^k>} e^{-\frac{1}{2}<f^t(x_i)+\delta h(x_i), y_i - y^k>} \Bigg] \\
&= -\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i), y_i - y^k > e^{-\frac{1}{2}<f^t(x_i)+\delta h(x_i), y_i - y^k>} \frac{\partial}{\partial \epsilon}\Big[e^{-\frac{1}{2}\epsilon<g(x_i), y_i - y^k>} \Big] \\
&= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} (< g(x_i), y_i - y^k >)^2 e^{-\frac{1}{2}<f^t(x_i)+\delta h(x_i), y_i - y^k>} e^{-\frac{1}{2}\epsilon<g(x_i), y_i - y^k>} \\
&= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} (< g(x_i), y_i - y^k >)^2 \; e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i)+\delta h(x_i), y_i - y^k>}.
\end{aligned}$$

Its value at $\epsilon = \delta = 0$:

$$\begin{aligned}
\frac{\partial^2 R[f^t + \epsilon g + \delta h]}{\partial \epsilon^2}\Big|_{\substack{\epsilon=0\\ \delta=0}} &= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} (< g(x_i), y_i - y^k >)^2 \; e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>} \\
&= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} \Big\langle g(x_i), y_i - y^k \Big\rangle \Big\langle g(x_i), y_i - y^k \Big\rangle * \\
&\qquad (e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}} (e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}} \\
&= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} \Big\langle g(x_i), (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}} \Big\rangle * \\
&\qquad \Big\langle g(x_i), (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}} \Big\rangle \\
&= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} \Big\langle g(x_i), (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}} \Big\rangle^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M}\Bigg[\Big\langle g(x_i), g(x_i)\Big\rangle + 2\Big\langle g(x_i), (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}}\Big\rangle + \\
&\qquad \Big\langle (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}}, (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}}\Big\rangle\Bigg] \\
&= \frac{1}{4}\sum_{i=1}^{n}\Bigg[\Big\langle g(x_i), g(x_i)\Big\rangle + 2\Big\langle g(x_i), \sum_{k=1}^{M}\Big[(y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}}\Big]\Big\rangle \\
&\qquad + \sum_{k=1}^{M}\Big\langle (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}}, (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}}\Big\rangle\Bigg] \\
&= \frac{1}{4}\sum_{i=1}^{n}\Big[\Big\langle g(x_i), g(x_i)\Big\rangle + 2\Big\langle g(x_i), \tilde{w}_i\Big\rangle + \hat{w}_i\Big],
\end{aligned}
$$

where:

$$
\tilde{w}_i = \sum_{k=1}^{M}\Big[(y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}}\Big]
$$

and

$$
\begin{aligned}
\hat{w}_i &= \sum_{k=1}^{M}\Big\langle (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}}, (y_i - y^k)(e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>})^{\frac{1}{2}}\Big\rangle \\
&= \sum_{k=1}^{M} e^{-\frac{1}{2}<f^t(x_i), y_i - y^k>}\Big\langle y_i - y^k, y_i - y^k\Big\rangle \\
&= e^{-\frac{1}{2}<f^t(x_i), y_i>}\sum_{k=1}^{M}\Big\langle y_i - y^k, y_i - y^k\Big\rangle e^{\frac{1}{2}<f^t(x_i), y^k>}.
\end{aligned}
$$

Similarly we can derive with respect to δ:

$$
\frac{\partial^2 R[f^t + \epsilon g + \delta h]}{\partial \delta^2}\Big|_{\substack{\epsilon=0\\ \delta=0}} = \frac{1}{4}\sum_{i=1}^{n}\Big[\Big\langle h(x_i), h(x_i)\Big\rangle + 2\Big\langle h(x_i), \tilde{w}_i\Big\rangle + \hat{w}i\Big],
$$

where: $\tilde{w}_i$ and $\hat{w}_i$ are defined as before.

Derivation of the mixed partial derivative of the risk function with respect to ϵ and δ:

$$
\begin{aligned}
\frac{\partial^2 R[f^t+\epsilon g+\delta h]}{\partial\delta\partial\epsilon} &= \frac{\partial^2 R[f^t,g,h]}{\partial\delta\partial\epsilon} \\
&= \frac{\partial}{\partial\delta}\left[\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i), y_i - y^k > e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i)+\delta h(x_i), y_i-y^k>}\right] \\
&= \frac{\partial}{\partial\delta}\left[-\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i), y_i - y^k > e^{-\frac{1}{2}\delta<h(x_i),y_i-y^k>} e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i), y_i-y^k>}\right] \\
&= -\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i), y_i - y^k > e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i), y_i-y^k>} \frac{\partial}{\partial\delta}\left[e^{-\frac{1}{2}\delta<h(x_i),y_i-y^k>}\right] \\
&= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} \Big\langle g(x_i), y_i - y^k \Big\rangle \Big\langle h(x_i), y_i - y^k \Big\rangle e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i), y_i-y^k>} e^{-\frac{1}{2}\delta<h(x_i),y_i-y^k>} \\
&= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} \Big\langle g(x_i), y_i - y^k \Big\rangle \Big\langle h(x_i), y_i - y^k \Big\rangle\ e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i)+\delta h(x_i), y_i-y^k>} \\
&= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} \Big\langle g(x_i) + h(x_i), y_i - y^k \Big\rangle\ \ e^{-\frac{1}{2}<f^t(x_i)+\epsilon g(x_i)+\delta h(x_i), y_i-y^k>}.
\end{aligned}
$$

Its value at $\epsilon = \delta = 0$:

$$
\begin{aligned}
\frac{\partial^2 R[f^t+\epsilon g+\delta h]}{\partial\delta\partial\epsilon}\Big|_{\substack{\epsilon=0\\ \delta=0}} &= \frac{1}{4}\sum_{i=1}^{n}\sum_{k=1}^{M} < g(x_i) + h(x_i), y_i - y^k >\ \ e^{-\frac{1}{2}<f^t(x_i), y_i-y^k>} \\
&= \frac{1}{4}\sum_{i=1}^{n} < g(x_i) + h(x_i), \sum_{k=1}^{M}(y_i - y^k)\ e^{-\frac{1}{2}<f^t(x_i), y_i-y^k>} > \\
&= \sum_{i=1}^{n} < g(x_i) + h(x_i), \frac{1}{2}w_i >,
\end{aligned}
$$

where w_i is defined as before (in the first order derivatives).